A TOPOGRAPHY STRUCTURE USED IN AUDIO STEGANOGRAPHY

Xuejie Ding Weiqing Huang Meng Zhang Jianlin Zhao
{*dingxuejie, huangweiqing, zhangmeng, zhaojianlin*}*@iie.ac.cn*
Institute of information engineering, Chinese academy of sciences

ABSTRACT

There are inter-frame and inter-element correlations in audio signal which we called correlated topography structure in this paper. The structure will be perturbed when a message is hided into the audio. Inspired by this principle, we propose a novel adaptive steganography scheme on audio by means of minimizing the perturbation to improve the undetectability and imperceptibility of stego audio. In the novel framework, the embedding algorithm is formulated into two phases of designing: 1) finding an optimal embedding path 2) designing an optimal modification strategy. The two phases focus on the stability of topography structure from global and local points respectively. Practical merits of this approach are validated by testing adaptive embedding scheme for audio in wavelet domain.

Index Terms—audio steganography, correlated topography structure, minimizing perturbation, embedding path, embedding strategy.

1. INTRODUCTION

As a technology of covert communication, steganography can embed secret to the open cover in public communication mode. It is not only hiding the information contents but also concealing the existence of this behavior. Therefore, the steganography techniques have to satisfy two requirements, one is the imperceptibility and the other is high hiding rate. Comparing with other steganography techniques， audio steganography is particular important resulted from the prevailing presence of audio or speech signals in our human society. Meanwhile, it is also challenging because of the sensitivity of human auditory system (HAS) [1-2].

Since imperceptibility is one of the important factors, the psycho acoustic masking is employed to audio steganography commonly. And the stego can be attained by modifying the amplitude or phase at the masked frequencies [3-4]. However, this kind of technique is complex and has a low hiding rate. As known, the algorithm of the least significant bit substitution (LSBs) is one of most simple and popular methods with a high rate of embedding. Meanwhile, it also faces different steganalysis algorithms to challenge its security [7]. To solve this problem, there are two ways, one is that LSBs technique is employed in transform domain, such as Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) and so on. It has been shown that modifying the LSB of transform coefficients can increase both the embedding capacity and minimize the distortion of cover. And the other one is to combine the LSBs with other rules or methods to improve the security, such as setting the hearing threshold as embedding threshold, only modifying high frequency components, employing sparse representation to address the undetectability concern and so on. The purpose of these methods is to make the stego indistinguishable from cover [5-8].

Recently, there is a novel principle called distortion minimizing (DM) becomes popular owning to the fact that both optimal embedding simulator and practical coding schemes [9]. In the DM framework, usually an additive distortion function is defined to express as a summation of the embedding impact. Then some stego-coding is applied to make the cover communicate larger payloads at the same embedding distortion or decrease the distortion for a given payload. Some excellent steganography systems using the DM can resist various detections have been designed for spatial and JEPG domains of image, such as HUGO, WOW, S-UNIWARD, MG and so on. In the paper [10], the authors gave some rules for cost assignment in spatial image steganography, and the methods above can be integrated into their scheme.

In this paper, we propose a novel audio stegangoraphy based on LSBs in discrete wavelet domain. Different from previous steganographic methods, the novel embedding processing is divided into two phases as choosing an embedding path and deciding a modifying style. Inspired by DM framework, the two processes are dependent on two distortion functions. The distortion functions are designed from the hypothesis that hiding data breaks the correlated topography structure existing in audio, which is a novel principle contrasted with DM framework, we call it perturbation minimizing (PM).In detail, inter-frame correlation is considered to compute and assign the distortion values for the elements by an unsupervised learning algorithm automatically. It needn't design an assignment scheme subjected to the rules proposed in [10]. Then syndrome-trellis codes (STCs) is applied to select the an optimal embedding path by minimizing the additive

This work is supported by the"Strategic Priority Research Program" of the Chinese Academy of Science, Grant No. Y2W0012102.

value from the hierarchical perspective. At last, for the elements on the embedding path, the inter-element correlation is considered to decide the optimal modification strategy for the LSB of wavelet coefficients.

2. PRELIMINARIES

2.1 Correlated topography structure (CTS)

In the audio signal, there are inter-frame and inter-element correlation structures which we called correlated topography structure in this paper. It represents the temporal correlation and consecution due to the short-time stationary of audio. Topography structure is observed after the audio is framed, and the principle is shown in Fig.1.

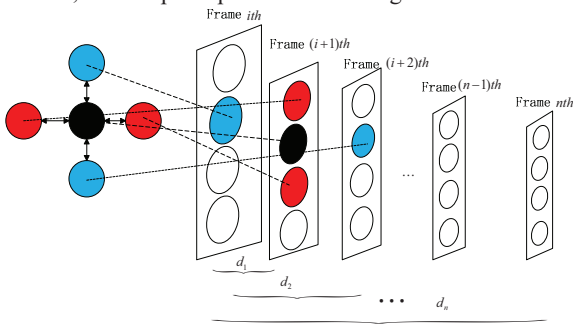

Fig.1 Correlated topography structure in audio

As shown in Fig.1, the adjacent frames are strong correlated, and the correlation becomes weaker when the intervals between the reference frame and others are increased. Therefore, we assume that the adjacent frames are strong correlated and far frames are independent. At the same time, the correlations between the elements of inter-frame and intra-frame are also focused. The centered black pad is a reference element, and the right and left red pads represent the correlations between reference element and others in the same frame. Correspondently, the up and down blue pads mean the correlations between the different frame elements and reference element in the same position. Therefore, there are two kinds of correlated topography structures exiting in audio file. One is the relationship of inter-frames and can be seen as a global concept. And the other is local topography structure which builds the relationship of inter-elements. In fact, the philosophy behind these two structures may not be contradicted. The inter-element topography structure is the basis of inter-frame. And we define the correlation relationship as linear correlation and energy correlation in this paper.

2.2 Framework of perturbation minimizing

The correlated topography structure maintains the basic property of audio, and it will be changed when the secret data is embedded. In order to improve the imperceptibility and undetectability of stego, a framework called perturbation minimization (PM) is proposed to guide the embedding process adaptively. The principle is shown as follows.

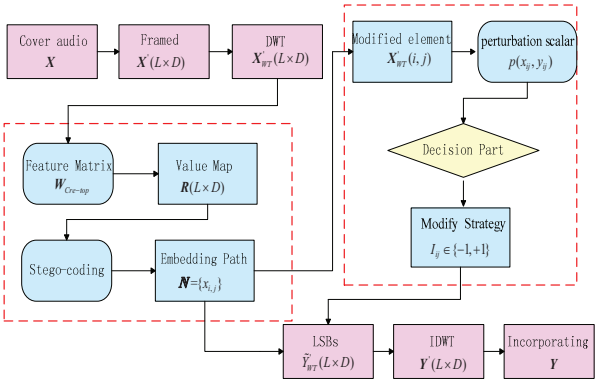

Fig.2. Novel framework of perturbation minimizing

In Fig.2, the audio cover is pre-processed with framing and DWT, which change a audio $\boldsymbol{X}$ into a coefficients matrix $\boldsymbol{X}'_{WT}$ .Then an optimal embedding path is found using stego-coding, which is on the base of a reasonable perturbation value map $\boldsymbol{R}$. The embedding path is designed under the global CTS and the modification elements' index $\boldsymbol{N}$ is selected at some payload. At last, a modification strategy on LSB is generated with considering the local CTS of elements. Then the stego signal $\boldsymbol{Y}$ is reconstructed with the inverse process of pre-processing.

Compared with embedding process, extracting the secret is easy. The coefficients matrix $\boldsymbol{Y}'_{WT}$ is attained after pre-processing of stego $\boldsymbol{Y}$ .Then the LSB of $\boldsymbol{Y}'_{WT}$ is extracted and reshape to a one-dimensional style as $\tilde{\boldsymbol{Y}}$. At last the secret information is got by $\boldsymbol{m} = \boldsymbol{H} \cdot \tilde{\boldsymbol{Y}}$, here $\boldsymbol{H} \in \{0,1\}^{m \times n}$ is the parity check which is shared between the sender and receiver.

In the next section, we will introduce the two key steps as selecting the optimal embedding path and optimal modification strategy.

3. EMBEDDING SCHEM

3.1 Finding an optimal embedding path

The goal of this section is to find an optimal path by means of minimizing the perturbation from global perspective. First, we assign a perturbation value $\boldsymbol{R}$ for the coefficients matrix $\boldsymbol{X}'_{WT}$. The value of $\boldsymbol{R}$ is computed using an unsurprised method called correlated topography analysis, and the process is given in Fig.3.

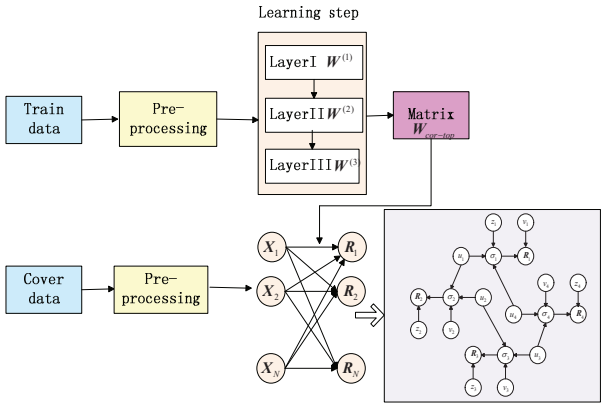

Fig.3 The process of generating $\boldsymbol{R}$ in topography structure

In a cortical interpretation, $\boldsymbol{R}_i$ model the responses of (signed) simple cells which reflect the perceptions of human brain from the external stimuli. The value of response is to represent the conspicuity of every neuron in the same perception level for a stimulus. The neuron response that the greatest value is important for cover, on the contrary, the neuron that has the smallest value can be omitted. Therefore, in this section, the neuron response is used to measure the intensity of perturbation for embedding secret into cover.

At the bottom right of Fig.3, the topography correlation is generated on the base of several random variables are used to represent as $\boldsymbol{R}=\boldsymbol{\sigma}\odot\boldsymbol{z}$. Where $\odot$ denotes element-wise multiplication, and $\boldsymbol{\sigma}$ and $\boldsymbol{z}$ are statistically independent. While the nearby elements in $\boldsymbol{z}$ and the squares of nearby elements are not statistically independent. u_i and v_i are hypothesized as positive random variables and statistically independent from each other. And σ can be represented as the formula $\sigma_i=(u_{i-1}+u_i+v_i)^{-1/2}$. The value map $\boldsymbol{R}$ can be computed according to the matrix $\boldsymbol{W}_{cor-top}$, which is attained in training phase with four steps as follows.

Step 1: White and central for $(\boldsymbol{X}'_{WT})_{train}$.

Step 2: Maximize $J_{\log-likelihood}(\boldsymbol{W})$ to obtain $\boldsymbol{W}^{(1)}$.

$$\boldsymbol{W}^{(1)}=\arg\max_{\boldsymbol{W}} J_{\log-likelihood}(\boldsymbol{W}) \tag{1}$$

Step3: Obtain $\boldsymbol{W}^{(2)}=(c_1^*\boldsymbol{w}_{k_1^*}^{(1)},\cdots,c_d^*\boldsymbol{w}_{k_d^*}^{(1)})^T$ by optimizing the order and sign vector $\boldsymbol{k}$ and $\boldsymbol{c}$, and $\boldsymbol{R}^{(1)}=\boldsymbol{W}^{(1)}(\boldsymbol{X}'_{WT})_{train}$.

$$\boldsymbol{k}^*,\boldsymbol{c}^*=\arg\max_{k,c}\left[-\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{d}G(c_i\boldsymbol{R}_{k_i}^{(1)}-c_{i+1}\boldsymbol{R}_{k_{i+1}}^{(1)})\right] \tag{2}$$

Step4: $\boldsymbol{W}^{(3)}$ is the output and called as $\boldsymbol{W}_{cor-top}$,

$$\boldsymbol{W}^{(3)}=\arg\max_{\boldsymbol{W}} J_{\log-likelihood}(\boldsymbol{W}^{(2)})+J_{topography}(\boldsymbol{W}^{(2)}) \tag{3}$$

The special forms of $J_{\log-likelihood}(\boldsymbol{W})$ and $J_{topography}(\boldsymbol{W})$ are applied in the paper [9].Then we can assign the value to the elements of cover as

$$\hat{\boldsymbol{R}}=\boldsymbol{W}_{cor-top}\boldsymbol{X}'_{WT} \tag{4}$$

The value map is attained by $\hat{\boldsymbol{R}}$ and it is quantified to $0\sim1$ as $\boldsymbol{R}=Q(\hat{\boldsymbol{R}})$.Then the STCs is used to find the optimal path under certain payload. The elements to be modified are stored as

$$\mathbb{N}=\{x_{i,j}|\arg\min_{n}\sum_{i,j}R_{i,j}\} \tag{5}$$

3.2 Designing an optimal modification strategy

In steganography system, the sender communicates secret to the receiver by introducing modifications to cover. We employed the operations of ±1 embedding at LSB which can be represented by $\mathrm{I}=\{x_{i,j}-1,x_{i,j},x_{i,j}+1\}$. In this section, we design an embedding strategy to determine $+1$ or -1 to maintain the stability of local correlated topography structure better.

When the elements are modified, the correlation or consistency between neighbor elements is broken, so for the element on the embedding path $x_{i,j}$, the consistency of $x_{i,j}$ and x_{kl} is computed as follows,

$$\gamma_{x_{i,j},x_{k,l}}=|\,x_{i,j}^T\cdot x_{k,l}\,|\,/\sqrt{x_{i,j}^T\cdot x_{i,j}}\sqrt{x_{k,l}^T\cdot x_{k,l}} \tag{6}$$

Then the sum consistency of $x_{i,j}$ in local CST is given as

$$\gamma(x_{i,j})=\alpha\cdot\sum_{l=j-1}^{j+1}\gamma_{x_{i,j},x_{i,l}}+(1-\alpha)\cdot\sum_{k=i-1}^{i+1}\gamma_{x_{i,j},x_{k,j}},(l\neq j,k\neq i) \tag{7}$$

α is the direction weight, which is used to adjust the contributions of two directions on perturbation scalar. In this paper, $\alpha=0.6$ is set because of the stronger correlation on same frame. Then the perturbation scalar is attained as

$$p(x_{i,j},y_{i,j})=|\,\gamma(x_{i,j})-\gamma(y_{i,j})\,|\,/\,|\,\gamma(x_{i,j})\,|+\varepsilon \tag{8}$$

Here $y_{i,j}$ is the result of +1 or –1 modification operated on $x_{i,j}$. $\varepsilon>0$ is a stabilizing constant to avoid dividing by zero. We choose the modification strategy on $x_{i,j}$ as follows,

$$I_{i,j}=\arg\min p(x_{i,j},y_{i,j}) \tag{9}$$

4. EXPERIMENTAL SETUP

4.1 Experimental Setup

In this section, we evaluate the proposed method based on the performance of imperceptibility and undetectability. In addition, the methods of wavelet domain LSB substitution in [5] and DWT-FFT reported in [6] and DWT-

SD in [7] are implemented for performance comparisons. Our method and the method of DWT-SD are used the optimum frame length value as 128 reported in [7].

The experimental data is come from TIMIT database, which combines 630 people come from eight dialect areas and together with 6300 pronounce sentences. The audio signal is monophonic waveform and sampled at 16 kHz, where each sample is quantized by 16bits. We use 500 speech data (100speakers) as the train data to learn the matrix $\boldsymbol{W}_{cor-top}$ and 500 speech data as the test data to prove the performance of proposed method randomly.

4.2 Audio quality evaluation

To prove the imperceptibility of stego, two objective performance measures are employed to evaluate the quality of audio. The first objective measure is average SegSNRs. The second is perceptual evaluation of speech quality (PESQ) which is one of technologies to use for evaluating the sound quality. The method of PESQ is with the score between $-0.5 \sim 4.5$, where 4.5 is the best quality and 3.8 is the threshold to be acceptable. The results of average SegSNRs are shown in Fig.4. The percents of scores lower than 3.8 at different payloads (bit per frame) are displayed in Table I.

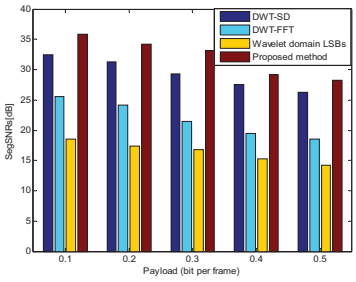

Fig.4 The results of SegSNRs

Table 1 PESQ Evaluation Results

Payload (bpf)	algorithms			
	DWT-SD	DWT-FFT	Wavelet domain LSBs	Proposed method
0.1	0.6%	16%	12%	0.2%
0.2	1.2%	58%	74%	0.25%
0.3	1.8%	64%	82%	0.52%
0.4	7.6%	72%	88%	6.2%
0.5	8.2%	84%	35%	7.4%

As can be seen, our method has attained the highest average SegSNRs at different payloads compared with other algorithms. It is observed that for our method, the percent of distortion is lowest evaluated by PESQ. Both of results have shown that the stego hidden data using our method has better quality.

4.3 Steganalysis

Steganalysis is an effective tool to evaluate the undetectability performance of a steganography scheme. We use two successful audio steganalysis methods, the first one is the second-order derivative-based Mel-cepstrum(2D-Mel)audio steganalysis method and its results demonstrated that it improved the state of the art method significantly[12]. The other one is CC-PEV report in paper [13], which is an efficient image steganalysis method. When the CC-PEV is employed, the audio matrix $\boldsymbol{Y}'_{L\times D}$ of stego needs to be provided, which is generated after being framed. And the CC-PEV is introduced to extract the coefficients and Markov features of stego matrix. The LibSVM toolbox is employed for classification, and the Gaussian Kernel is utilized in our experiment. The receiver operating characteristic (ROC) curves are depicted in Fig.5.

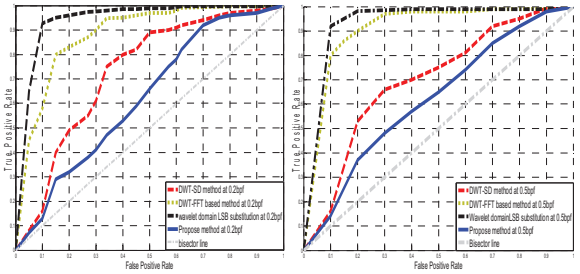

Fig5. ROC curves of 2D-Mel steganalysis method [12] on the stego audio files generated by four different steganography methods at two different payloads.

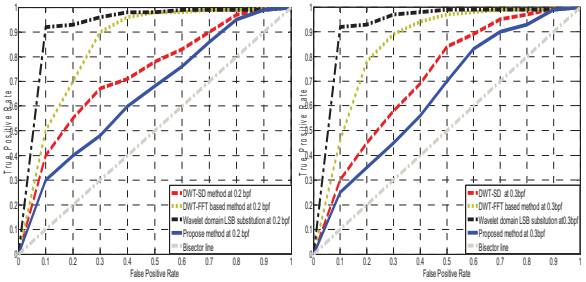

Fig6. ROC curves of CC-PEV steganalysis method [13] on the stego audio files generated by four different steganography methods at two different payloads.

Compared with previous three methods, the ROC curves show that our method reduces the probability of detection significantly, which implies that our method has a higher level of steganographic undetectability.

5. CONCLUSION

In this paper, we propose a novel method called perturbation-minimizing (PM) to design security stegangraphy scheme on audio. The novel scheme is on the purpose of maintaining the correlated topography structure before and after embedding by means of choosing an optimal embedding path and a modification strategy adaptively. It suggests that the structure existing in cover can be used in covert communication, stegangraphy and maybe steganalysis. The experimental results show that, satisfactory performances of imperceptible and undetectable on audio stego are achieved.

6. REFERENCES

[1] R, Tanwar, M. Bisla, "Audio Steganography", in *Proc. ICROIT,* India, 2014, pp. 322-325.

[2]M. Nosrati, R. Karimi, M, Hariri, "Audio Steganography: A Survey on Recent Approaches," World Applied Programming, vol.2 no.3, pp: 202-205, Mar, 2012.

[3] Y. Erfani, S. Siahpoush, "Robust audio watermarking using improved TS echo hiding," Digital Signal Processing, vol.19,pp.809-814,2009.

[4] B.S. Ko, R. Nishimura, Y. Suzuki, "Time-spread echo method for digital audio watermarking," IEEE Trans, Multimedia, vol. 7(2), pp.212-221, 2005.

[5] A. Delforouzi and M. Pooyan, "Adaptive digital audio steganography based on integer wavelet transform," Circuits, Syst. Signal Process, vol.27, no.2, pp.247-259, Apr.2008.

[6] S.Rekik,D. Guerchi, S.A. Selouani, and H. Hamam, "Speech steganography using wavelet and fourier transforms," *EURASIP J. Audio, Speech, Music Process*, No.1,pp.1-4, Aug.2012.

[7]S. Ahani, S Ghaemmaghami, Z. J Wang, "A Sparse Representation-Based Wavelet Domain Speech Steganography Method" [J]. *IEEE/ACM Trans.* Audio, Speech, and Language Processing, vol. 23, No.1, Jan, 2015.

[8] X. P. Huang, N. Ono, I. Echizen and A. Nishimura, "Reversible Audio Information Hiding Based on Integer DCT Coefficients with Adaptive Hiding Locations," in *Proc. IWDW, 2013, LNCS,* vol. 8389. pp. 376-389.

[9]T. Filler, J Judas, J Fridrich. "Minimizing Additive Distortion in Steganography using Syndrome-Trellis Codes," *IEEE Trans. Inf.* Forensics Security, vol. 6, no.3, pp.920-935, Mar. 2011.

[10]B. Li, S. Q. Tan, M. Wang, J.W Huang. "Investigation on Cost Assignment in Spatial Image Steganography," *IEEE Trans. Inf. Forensics Security*, vol.9, no.8, pp.1264-1277, Aug, 2014.

[11] H. Sasaki, M U. Gutmann, H. Shouno, A. Hyvärinen. "Correlated topographic analysis: estimating an ordering of correlated components," Machine Learning,vol.92, pp: 285-317.

[12] Q. Liu, A. H. Sung and M. Qiao, "Derivative-based audio steganalysis," *ACM Trans*. Multimedia Computing, Commun, Applicat.(TOM-CCAP),vol.7,no.3, pp.1-19, Aug, 2011.

[13] J. Kondovsky, J. Fridrih, "Calibration revisited," Proceedings of the 11th ACM multimedia and security workshop, Princeton, NJ, September,7-8,2009.