Derivative-Based Steganographic Distortion and Its Non-additive Extensions for Audio

Kejiang Chen, Hang Zhou, Weixiang Li, Kuan Yang, Weiming Zhang, and Nenghai Yu

***Abstract*— Steganography is the art of covert communication, which aims to hide the secret messages into cover medium while achieving high undetectability. To this end, the framework of minimal distortion embedding is widely adopted for adaptive steganography, where a well-designed distortion function is significant. In this paper, inspired by the phenomenon that the modification of audio samples with the low amplitude will be easily detected, a novel distortion is presented for audio steganography. Taking the fragility of the low amplitude audio samples into account, the proposed distortion is inversely proportional to the amplitude. Furthermore, in order to resist the strong steganalysis, the derivative filter is utilized for acquiring the residual of audio, which plays an important role in distortion definition. The experimental results show that the proposed distortion outperforms the state-of-the-art methods defending strong steganalytic methods. To take a step forward, considering the mutual impact caused by embedding modification, the non-additive extensions of the proposed methods are put forward. The extending experiments show that in most cases, the proposed non-additive extensions can achieve higher level of security than the original methods.**

***Index Terms*— Steganography, derivate filter, non-additive, amplitude.**

I. Introduction

STEGANOGRAPHY is the art of covert communication, which hides messages into digital media so that no one apart from sender and the intended recipients can cognize the existence of the secret [1]. Nowadays, lossless audio in WAV format is widely spread in social media, such as raw music share websites, audio synthesis applications. And audio steganography on WAV format has developed a lot in the past years, including conventional and content-adaptive schemes. In conventional audio steganography, the secret message is embedded by replacing the least significant bits of audio, which means every element in cover audio possesses the same modification priority. In this way, some sensitive parts in cover audio, such as mute segments, will be modified, which possibly exposes to steganalysis [2]. Accordingly, under the framework of minimizing distortion steganography, a content-adaptive scheme [3] has been proposed. The distortion was built on the residual between the original audio and the compressed audio using Advanced Audio Coding (AAC). With the methodology of syndrome-trellis codes (STCs) [4], the information embedding can be well implemented. The aforementioned distortion function is additive. Under additive distortion model, the total distortion caused by the embedding can be expressed as the sum of embedding distortion over all elements. In other words, the modification of the current element do not impact the modification priority of neighbour elements. Zhang *et al.* [5] and Li *et al.* [6] proposed non-additive schemes (*DeJoin* and *UpDist*) for considering the mutual impact of modification, and show considerable performance in image steganography.

The oppose of steganography, steganalysis, aims at detecting and analyzing the hidden information in digital media. In the past few years, several steganalytic methods have been proposed for detecting the existence of secret message in audio steganographic systems. Kraetzer et al. calculated steganalytic feature on Mel-frequency cepstral coefficients, which are widely used in speech recognition, and it delivers good performance [7]. Liu et al. improved Kraetzer et al.'s work by building steganalytic feature on derivative filter residual of the Mel-cepstrum coefficients [8]. Luo et al. further enhanced the steganalytic performance by extracting effective features from both the time and frequency domains, which owns the state-of-the-art performance [9].

The strong steganalytic methods show the weakness of the current steganography for audio in time domain, namely, the distortion designed in [3] is not precise enough, which requires us to design new secure steganographic algorithms. In this paper, considering the property of audio as well as the process of steganalytic feature, we design a new distortion function for content-adaptive audio steganography, which is related to the amplitude of audio and the residual obtained by derivative filter. Detailly, the audio sample for low amplitude is not suitable for embedding for they are voiceless sounds. There is no fundamental frequency in voiceless sounds, which can be easily modeled by Sinusoidal Model [10]. Based on the analysis, we propose a rule named "large-amplitude-first", which will assign low modification distortion to large amplitude audio samples.

Furthermore, the state-of-the-art steganalysis are designed based on the derivative signal [8], [9]. In order to defend against them, the derivative filter residual is combined into the distortion function. The distortion designed is still additive,

Manuscript received March 18, 2019; revised May 3, 2019; accepted May 16, 2019. Date of publication May 23, 2019; date of current version July 2, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1636201 and Grant 61572452 and in part by the Fundamental Research Funds for the Central Universities under Grant WK6030000135 and Grant WK6030000136. This paper was recommended by Associate Editor X. Cao. *(Corresponding author: Weiming Zhang.)*

The authors are with the CAS Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China, Hefei 230026, China (e-mail: zhangwm@ustc.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2019.2918511

which means the modification impact is mutually independent. However, it is obvious that the modification will interact with neighbour elements. Consequently, we introduce non-additive extension (*DeJoin* and *UpDist*) schemes for proposed distortion. Experiments demonstrate the effectiveness of our proposed steganographic distortion and its non-additive extensions.

The main contributions of our proposed approach can be summarized as follows:

- Considering the fragility of modifying low amplitude samples, the "large-amplitude-first" rule is proposed for audio steganography.
- The derivative filter is utilized for generating predicting residual in distortion function for defending the state-of-the-art steganalysis.
- Based on the proposed distortion, non-additive extensions using *DeJoin* [5] and *UpDist* [6] are presented to further enhance the security, respectively.

The rest of this paper is organized as follows. Section II reviews some representative related work. Section III and IV present our proposed steganographic distortion and its non-additive extensions. Section V shows the experiments for verifying the effectiveness of the proposed scheme, and Section VI concludes this paper.

II. Related Works

In this section, we briefly introduce some related works from two aspects: minimal distortion steganography and AACbased distortion.

A. Minimal Distortion Steganography

The minimal distortion steganography model is established in [4], where the distortion of changing x_i to y_i can be denoted by $d_i(\mathbf{x}, y_i)$. It is supposed that no modification means no distortion, and the modification distortion is same whether $+1$ or -1. Namely, $d_i(\mathbf{x}, x_i) = 0$ and $d_i(\mathbf{x}, x_i - 1) = d_i(\mathbf{x}, x_i + 1) = d_i \in [0, \infty)$. The overall distortion is the sum of the distortion of every element:

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} d_i |x_i - y_i|. \tag{1}$$

Denote $\pi(y_i)$ as the probability of changing x_i to y_i. Given message m, the sender wants to minimize the average distortion Eq. (1). The optimal scheme is allocating the modification probability of every element following the Gibbs distribution [11]:

$$\pi(y_i) = \frac{\exp(-\lambda d_i(\mathbf{x}, y_i))}{\sum_{y_i \in I_i} \exp(-\lambda d_i(\mathbf{x}, y_i))}, \quad 1 \le i \le n, \tag{2}$$

where the scalar parameter $\lambda > 0$ is determined by the payload constraint

$$m = \sum_{i=1}^{n} \sum_{y_i \in I_i} \pi(y_i) \log \frac{1}{\pi(y_i)}. \tag{3}$$

Given additive distortion, STCs [4] can accomplish the message embedding with the minimal distortion nearing the theoretical bound.

B. AACbased Distortion

As for adaptive steganography, the cover elements in the complex region which are hard to be predicted by neighbour elements will be assigned low distortion. This strategy is called complexity-first principle [12]. According to the principle, Luo et al. designed a distortion definition method [3], which assigns high distortion to samples whose differences between original audio and the reconstructed audio under AAC compression and decompression are large. The difference between the original audio and the reconstructed audio can be denoted as:

$$r(i) = x(i) - x'(i), \tag{4}$$

where $r(i)$ is the i-th difference, and $x(i)$, $x'(i)$ are the i-th element of the original audio and the reconstructed audio after AAC compression with a high bitrate and decompression using NeroAAC tool,[1] respectively. Then the modification distortion is calculated as:

$$\rho_i^+ = \begin{cases} 1/|r(i)|, & \text{if } r(i) < 0 \\ 10/|r(i)|, & \text{if } r(i) > 0 \\ 10 & \text{if } r(i) = 0, \end{cases} \tag{5}$$

$$\rho_i^0 = 0, \tag{6}$$

$$\rho_i^- = \begin{cases} 10/|r(i)|, & \text{if } r(i) < 0 \\ 1/|r(i)|, & \text{if } r(i) > 0 \\ 10 & \text{if } r(i) = 0, \end{cases} \tag{7}$$

where ρ^+, ρ^-, ρ^0 represent the cost of the i-th element with modification of $+1$, -1, 0, respectively.

III. The Proposed Approach

A. Motivation

The complexity-first principle indicates that those elements easily predicted by neighbours will be assigned high cost, and elements in complex region will be assigned low cost. Most existing adaptive steganographic schemes follow the rule to design cost functions, including the aforementioned AAC-based distortion. However, when it comes to different digital media, the property of the media should be fully considered, otherwise the security is rather limited.

The modification distribution of AACbased at 0.4 bps (bit per sample) is presented in Fig. 1. It can be seen that there exist many modifications in the low amplitude samples (red rectangle in Fig. 1), which is voiceless sound. For purely voiceless sounds, there is no fundamental frequency in excitation signal and therefore no harmonic structure either and the excitation, and can be well modeled by Sinusoidal Model [10], which indicates that this part of audio is not suitable for embedding message. To verify such conjecture, we further conduct a simulation with the following steps.

1) Collect 1000 audio clips, and then calculate the residual of each audio sample x according to Eq. (4).
2) Select the residuals r_s which are larger than the median of the residuals, and count the frequency P of these residuals.

[1] Nero AAC Codec is available at: https://nero-aac-codec.en.lo4d.com.

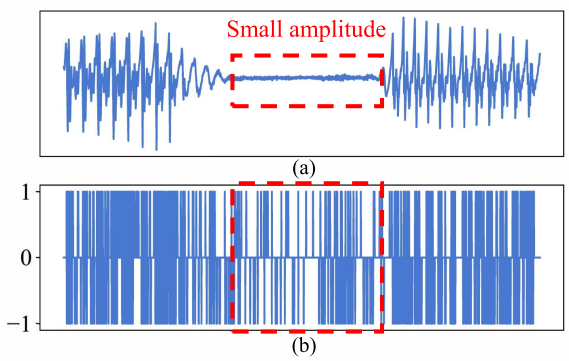


Fig. 1. The modification distribution of AACbased algorithm. (a) Audio samples. (b) Modification points.

TABLE I
THE AVERAGE MMD VALUES AND THE AVERAGE DETECTION ERROR $\overline{P}_{\mathrm{E}}$ OF DIFFERENT PAIRS

Pairs	Average MMD	$\overline{P}_{\mathrm{E}}$
$F(x), F(y_h)$	0.0052±0.0018	0.0714±0.0074
$F(x), F(y_l)$	0.0034±0.0012	0.1664±0.0124

3) For audio samples whose frequency P is a multiple of 2, we will divide them equally into two parts: high amplitude x_h and low amplitude x_l.
4) The random noise is added into audio samples where x_h and x_l locate in cover audio x to simulate data embedding, respectively. And two stego audios are obtained for the cover audio x, denoted by y_h and y_l.
5) Compute the 585-Dimensional MFCCF steganalytic feature [9] for cover and stego, $F(x), F(y_h), F(y_l)$. Then MMD (maximum mean discrepancy) [13] and steganalysis results are calculated, which quantify the security performance through distance between the feature sets (cover and stego) and the classifying error rate, respectively.

The MMD values and steganalysis results are computed on two pairs: $F(x), F(y_h)$ and $F(x), F(y_l)$. As for the setting of steganalysis, the reader can refer to the Section V for detail, and the only difference is that the audio clips are splitted into 500/500. A smaller value of MMD or a higher value of testing error P_{E} leads to less statistical detectability. The security measurement is implemented for 10 times to avoid outlier values, and the results are shown in Table I. It can be observed that $\mathrm{MMD}\left(F(x), F(y_h)\right) < \mathrm{MMD}\left(F(x), F(y_l)\right)$ and $\overline{P}_{\mathrm{E}}\left(F(x), F(y_h)\right) > \overline{P}_{\mathrm{E}}\left(F(x), F(y_l)\right)$, indicating that making modifications on the audio samples with large amplitude is more secure. In this sight, we propose a rule named "large-amplitude-first", which means the modification should be inversely proportional to the amplitude.

B. The Proposed Method

Combining complexity-first rule with large-amplitude-first rule, we propose a novel distortion function for uncompressed WAV audio. As for complexity-first rule, we make that the distortion is inversely proportional to the absolute value of residual, where the residual is obtained by the derivative filter, describing the complexity degree of audio samples. The reason why the derivative filter is selected as the texture descriptor is to defend against the state-of-the-art steganalytic features that are calculated on the derivative filter. When it comes to large-amplitude-first rule, similarly, the proposed distortion will be inversely proportional to the amplitude. Before giving the definition of the distortion, we will introduce the derivate filer first.

1) Obtaining Derivate Filter Residual: For an audio x, the first and the n-th partial derivative can be respectively defined as

$$\frac{\partial x(i)}{\partial i} = x(i) - x(i+1), \tag{8}$$

$$\frac{\partial^n x(i)}{\partial i^n} = \frac{\partial^{n-1} x(i)}{\partial i^{n-1}} - \frac{\partial^{n-1} x(i)}{\partial (i+1)^{n-1}}. \tag{9}$$

Let f_n denote the derivative filter of the order n. For example,

$$f_1 = [-1 \quad 1], \quad f_2 = [-1 \quad 2 \quad -1], \quad f_3 = [-1 \quad 3 \quad -3 \quad 1]. \tag{10}$$

The audio filter residual can be obtained by convolving an audio with a filter f, and is denoted as r_f:

$$r_f = x \otimes f, \tag{11}$$

where the symbol $\otimes$ denotes the convolution with x mirror-padded so that $x \otimes f$ has the same dimension as x.

2) Distortion Definition: Owning the residual of audio, combining with the large-amplitude-first rule, the proposed steganographic distortion, named DFR, is designed as follows:

$$\rho^+ = \rho^- = \frac{1}{|r_f| + |x| + \sigma} = \frac{1}{|x \otimes f| + |x| + \sigma}, \tag{12}$$

where r_f is the derivative filter residual, corresponding to the complexity-first rule, and x is the amplitude of the audio sample, corresponding to the large-amplitude-first rule. $\sigma = 2^{-10}$ is a stabilizing constant introduced to avoid dividing by zero.

IV. NON-ADDITIVE EXTENSIONS

The proposed distortion is additive, which means the modification of current sample will not impact the modification cost of other samples. Intuitively, the changes on adjacent audio samples will interact, and thus non-additive steganography will be more suitable for adaptive steganography. There exist two ways for implementing non-additive steganography, *DeJoin* and *UpDist*, which will be adopted to further enhance the security of proposed additive distortion. In the following subsections, we will introduce how to carry out the non-additive steganography for audio.

A. DeJoin for Audio

We first define the initial distortion on audio sample using proposed distortion method, and then the joint distortion is calculated for each audio sample block following the synchronizing modification direction principle [12] based on the initial distortion. In detail, the audio is divided into 1×2

$\tau(l,r)$:

l \ r	-1	0	+1
-1	1	α	2α
0	α	1	α
+1	2α	α	1

Fig. 2. Scaling function for audio steganography.

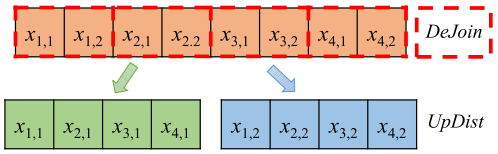


Fig. 3. An example of the composition of blocks and division of an audio into two disjoint sub-audios. The adjacent two elements make up a block for *DeJoin*. The odd sample points $x_{1,1}, x_{2,1}, x_{3,1}, x_{4,1}$ and even sample points $x_{1,2}, x_{2,2}, x_{3,2}, x_{4,2}$ compose two subgroups for *UpDist*.

non-overlapped sequences, shown as Fig. 3. For the sequence $S_i = (x_{i,1}, x_{i,2})$, we denote the initial distortion on $x_{i,1}$ by $\rho_1^{(i)}(l)$ for $l \in \{-1, 0, +1\}$ and the distortion on $x_{i,2}$ by $\rho_2^{(i)}(r)$ for $r \in \{-1, 0, +1\}$. Then the distortion on S_i is defined as:

$$\rho^{(i)}(l,r) = \tau(l,r) \times \left(\rho_1^{(i)}(l) + \rho_2^{(i)}(r)\right), \tag{13}$$

where the scaling factor $\tau(l,r)$ depends on α as shown in Fig. 2, meaning that the modification distortions in the same direction are promoted by multiplying smaller scaling factors. *DeJoin* implements a two-round embedding strategy by decomposing the joint probability into marginal probability and conditional probability for individual element following the chain rule, so that STCs can be adopted to embed message efficiently [5].

B. UpDist for Audio

UpDist first defines the distortion on cover according to the proposed distortion function, and then the cover is divided into two sub-cover. Afterwards, the distortion is individually minimized in each sub-cover while the costs of cover elements within each sub-cover are dynamically updated [6], which motivates the neighbour modified elements have the same modification direction. Take two subgroups as an example in Fig. 3, the first group is embedded first, and then the distortion ρ_i of the second group will be updated as follows:

$$\rho_i^+ = \rho_i^+ / \beta \quad \text{if } e_i > 0, \tag{14}$$

and

$$\rho_i^- = \rho_i^- / \beta \quad \text{if } e_i < 0, \tag{15}$$

where e_i is the difference between cover and stego of the i-th element in the first subgroup, and β is the factor playing the similar role as τ in *DeJoin*.

TABLE II
THE AVERAGE TESTING ERROR $\overline{P}_{\mathrm{E}}$ OF DIFFERENT ORDERS OF DERIVATIVE FILTER

Order	Filter operator	$\overline{P}_{\mathrm{E}}$
1	[-1 1]	0.1865±0.0020
2	[-1 2 -1]	0.1953±0.0027
3	[1 -3 3 -1]	0.1966±0.0016
4	**[-1 4 -6 4 -1]**	**0.2052±0.0028**
5	[-1 5 -10 10 -5 1]	0.1974±0.0028
6	[-1 6 -15 20 -15 6 -1]	0.1946±0.0021

V. EXPERIMENT

A. Setups

In this paper, we collect 20,000 mono 16 kHz 16-bit speech clips from CMU_ARCTIC[2] and LibriSpeech,[3] where 10,000 clips (CMU_ARCTIC) belong to training and testing set, and the other 10,000 clips (LibriSpeech) are the validation set for parameter setting. All the clips with various contents have the same duration of 3 seconds and are stored in WAV format. Two steganographic algorithms in the time domain, including AACbased and the proposed DFR are compared. All tested embedding algorithms are simulated at their corresponding payload-distortion bound for payloads $R \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ bps. The state-of-the-art feature sets MFCCF [9] and D-MC [8] are selected for steganalysis of audio.

The statistical undetectability is qualified with the total classification error probability on the testing set under equal priors $P_{\mathrm{E}} = min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$, where P_{FA} and P_{MD} are the false-alarm probability and the missed-detection probability. In order to reduce the occasional error, we implement the classification for ten times on different shuffle 5000/5000 database splits and obtain the average testing error $\overline{P}_{\mathrm{E}}$. Larger $\overline{P}_{\mathrm{E}}$ represents stronger security.

B. Investigation on Derivative Filter

We investigate the effect of different orders for derivative filter. The steganalytic results at 0.4 bps against MFCCF on validation sets are shown in Table II. The $\overline{P}_{\mathrm{E}}$ of the proposed method using the 4-th order filter f_4 is the highest, meaning that it is the most undetectable. As a result, the 4-th order filter f_4 is selected as the final filter.

We would also like to show the superiority of derivative filters comparing to the AAC in terms of generating residual. In implementation, the security performance of derivative filter is obtained by merely replacing the residual in AACbased with the residual calculated by derivative filter with the 4-th order, named DFRbased. Table III shows the results, and it can be observed that the DFRbased method outperforms AACbased method, which verifies the conclusion that calculating residual by derivative filter owns more secure performance.

C. Determining the Scaling Factor of Non-Additive Extensions

We investigate the effects of different scaling factors for non-additive extensions *DeJoin* and *UpDist*. DFR is chosen as

[2]The CMU_ARCTIC can be downloaded at http://festvox.org/cmu_arctic/
[3]The LibriSpeech Dataset can be downloaded at http://www.openslr.org/resources/12

TABLE III

THE AVERAGE TESTING ERROR $\overline{P}_{\mathrm{E}}$ OF AACBASED AND DFRBASED AGAINST MFCCF

Methods \ Payloads	0.1	0.2	0.3	0.4	0.5
AACbased	0.3435	0.2146	0.1417	0.0972	0.0689
DFRbased	**0.4071**	**0.3060**	**0.2125**	**0.1446**	**0.1015**

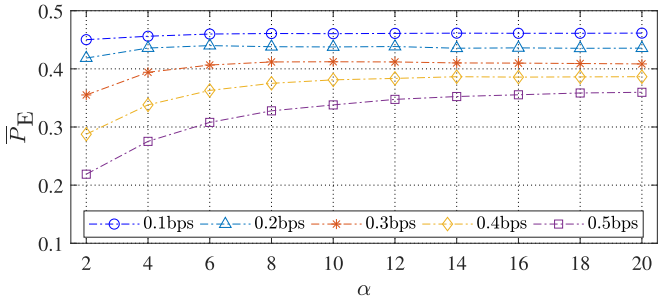


Fig. 4. Investigation on the effect of different α in *DeJoin* schemes.

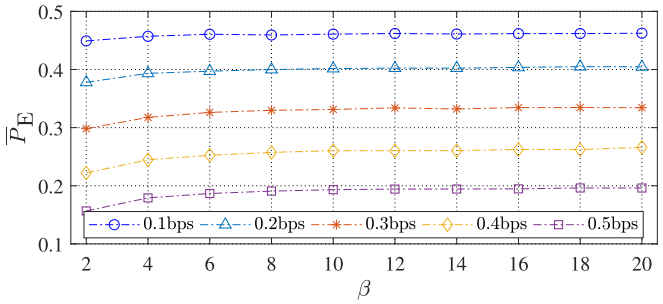


Fig. 5. Investigation on the effect of different β in *UpDist* schemes.

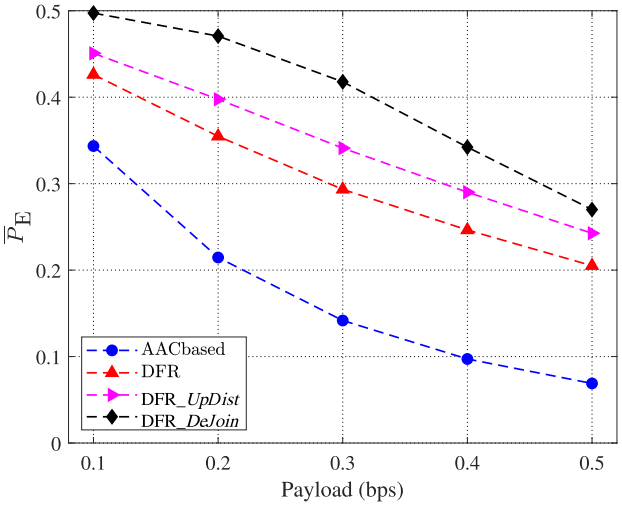


Fig. 6. Steganalytic performance (MFCCF) for steganographic methods with the optimal embedding simulator on the CMU_ARCTIC database.

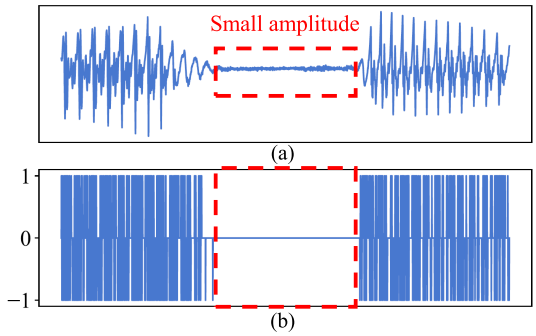


Fig. 7. The modification distribution of DFR algorithm.

the seed algorithm. The exploring results are shown in Fig. 4 and Fig. 5. For *DeJoin* and *UpDist*, the $\overline{P}_{\mathrm{E}}$ turns to be large and steady with the increment of α and β versus different payloads, meaning that the non-additive extensions do improve the security performance of the additive distortion. According to the improved tendency, we fix $\alpha = 20$, $\beta = 20$ in our following experiments.

D. Comparison of Security Level

AACbased, DFR and the non-additive extensions are chosen as the steganographic methods. The non-additive extensions adopting *DeJoin* and *UpDist* are named by suffixing the original name with "*_UpDist*" and "*_DeJoin*", such as DFR_*UpDist*, DFR_*DeJoin*. In this paper, only two elements are mutually considered together. The steganalytic results are demonstrated in Fig. 6. Since the results under D-MC steganalytic features are all close to 50%, which means it is too weak to detect adaptive steganographic schemes, only the steganalytic results of MFCCF are presented. We can observe that DFR outperforms AACbased and the improvements become large with the increment of the payload. The non-additive extensions including DFR_*UpDist*, DFR_*DeJoin*, show superior secure performance than the seed algorithm DFR. The improvement of DFR_*UpDist* with respect to DFR is less significant than DFR_*DeJoin*, which is due to DFR_*DeJoin* considers mutual impact more adequately than DFR_*UpDist*.

E. Visualizing Embedding Changes

To verify whether the proposed distortion can effectively avoid low amplitude samples to be modified, we give an example to visualize the embedding changes in Fig. 7. It is obvious that the low part of the audio sample is rarely modified, which shows the modification of elements with lower amplitude is restrained under the proposed scheme.

VI. CONCLUSIONS

In this paper, considering the fragility of the low amplitude sample as well as the perspective of construction of the state-of-the-art steganalysis, a new distortion for adaptive audio steganography is proposed. In order to further improve the security performance, non-additive extension of the proposed distortion is presented. The experimental results demonstrate that the proposed distortion outperforms the current adaptive distortion with a large margin, and the non-additive extensions do improve the secure level of the seed distortion.

In our future work, we would like to model the distribution of the residual, and design model-based steganography

for audio. In addition, the stereo audio will be explored in the non-additive way.

REFERENCES

[1] T. Pevný and J. Fridrich, "Benchmarking for steganography," in *International Workshop on Information Hiding*. Berlin, Germany: Springer, 2008, pp. 251–267.

[2] Q. Liu, A. H. Sung, and M. Qiao, "Temporal derivative-based spectrum and mel-cepstrum audio steganalysis," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 3, pp. 359–368, Sep. 2009.

[3] W. Luo, Y. Zhang, and H. Li, "Adaptive audio steganography based on advanced audio coding and syndrome-trellis coding," in *Proc. Int. Workshop Digit. Watermarking*, Jul. 2017, pp. 177–186.

[4] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 920–935, Sep. 2011.

[5] W. Zhang, Z. Zhang, L. Zhang, H. Li, and N. Yu, "Decomposing joint distortion for adaptive steganography," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2274–2280, Oct. 2017.

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

[7] C. Kraetzer and J. Dittmann, "Mel-cepstrum-based steganalysis for VoIP steganography," *Proc. SPIE*, vol. 6505, Mar. 2007, Art. no. 650505.

[8] Q. Liu, A. H. Sung, and M. Qiao, "Derivative-based audio steganalysis," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 7, no. 3, Aug. 2011, Art. no. 18.

[9] W. Luo, H. Li, Q. Yan, R. Yang, and J. Huang, "Improved audio steganalytic feature and its applications in audio forensics," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 2, May 2018, Art. no. 43.

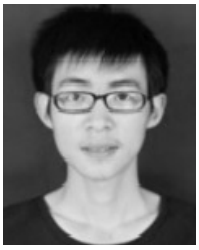
[10] G. P. Kafentzis and Y. Stylianou, "High-resolution sinusoidal modeling of unvoiced speech," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 4985–4989.

[11] T. Filler and J. Fridrich, "Gibbs construction in steganography," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 705–720, Dec. 2010.

[12] B. Li, S. Tan, M. Wang, and J. Huang, "Investigation on cost assignment in spatial image steganography," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 8, pp. 1264–1277, Aug. 2014.

[13] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, and B. Schölkopf, A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.

Kejiang Chen received the B.S. degree from the School of Communication and Information Engineering, Shanghai University, in 2015. He is currently pursuing the Ph.D. degree in information security with the University of Science and Technology of China (USTC). His research interests include information hiding, image processing, and deep learning.



Hang Zhou received the B.S. degree from the School of Communication and Information Engineering, Shanghai University, in 2015. He is currently pursuing the Ph.D. degree in information security with the University of Science and Technology of China (USTC). His research interests include information hiding, image processing, and computer graphics.

Weixiang Li received the B.S. degree from Xidian University, Xi'an, China, in 2016. He is currently pursuing the Ph.D. degree with the University of Science and Technology of China. His research interests include steganography and steganalysis. He was a recipient of the Best Student Paper Award of 6th ACM IH&MMSec in 2018.

Kuan Yang received the B.S. degree from the University of Science and Technology of China, in 2016, where he is currently pursuing the M.E. degree in electronics and communication engineering. His research interests include information hiding, image processing, and deep learning.

Weiming Zhang received the M.S. and Ph.D. degrees from the Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2002 and 2005, respectively. He is currently a Professor with the School of Information Science and Technology, University of Science and Technology of China. His research interests include multimedia security, information hiding, and privacy protection.

Nenghai Yu received the B.S. degree from the Nanjing University of Posts and Telecommunications in 1987, the M.E. degree from Tsinghua University in 1992, and the Ph.D. degree from the University of Science and Technology of China in 2004. He is currently a Professor with the University of Science and Technology of China. His research interests include multimedia security, multimedia information retrieval, video processing, information hiding and security, and privacy and reliability in cloud computing.