Reversible Watermarking on Stereo Audio Signals by Exploring Inter-Channel Correlation

Yuanxin Wu, CAS Key Laboratory of Electro-Magnetic Space Information, Chinese Academy of Sciences, University of Science and Technology of China, Hefei, China

Wen Diao, CAS Key Laboratory of Electro-Magnetic Space Information, Chinese Academy of Sciences, University of Science and Technology of China, Hefei, China

Dongdong Hou, CAS Key Laboratory of Electro-Magnetic Space Information, Chinese Academy of Sciences, University of Science and Technology of China, Hefei, China

Weiming Zhang, CAS Key Laboratory of Electro-Magnetic Space Information, Chinese Academy of Sciences, University of Science and Technology of China, Hefei, China

ABSTRACT

A new reversible watermarking algorithm on stereo audio signals is proposed in this article. By utilizing correlations between two channels of audio signal, the authors segment one channel based on another one according to the smoothness. For each segmented sub-host sequence, they estimate its capacity and the corresponding embedding distortion firstly, and then select the optimal combinations of sub-host sequences for embedding. Experimental results indicate that the proposed algorithm can improve SNR (signal to noise ratio) for various kinds of capacity.

KEYWORDS

1. INTRODUCTION

With the rapid development of multimedia and network, the amount of information storage becomes larger and larger. At the same time, editing and copying is so convenient, which speeds up the spread of information. Currently, a lot of digital works are suffering from illegal acquiring and malicious tampering, among which audios are the popular ones. Integrity protection and ownership rights certification of audio files have attracted great attentions, which can be realized with watermarking.

There are two types of watermarking, robust watermarking and fragile watermarking to protect audio files. Robust watermarking (Malvar, 2003; Naderahmadian, 2014; Tian, 2016; Rani, 2014) is used to label copyright information so that protecting copyright. On the contrary, fragile watermarking (Zhang, 2009, 2008; Qin, 2012) is usually used for content integrity authentication, which is asked to be sensitive to the slight change so that editing the content slightly will be detected. Reversible watermarking mainly is used for fragile watermarking, which can restore both the embedded watermark

and the host signal. The reversibility is very important in some special situations, such as high-quality music, legal evidence, military intelligence and criminal investigation.

At the early age, (Barton,1997) proposed the idea of reversible data hiding (RDH). Later on, many efficient methods for images spring up. The current algorithms are divided into five mainstream techniques: lossless compression based schemes (Celik, 2005; Lin, 2011; Zhang, 2012), expansion based schemes (Tian, 2003; Ni, 2006; Thodi, 2007; Hong, 2009; Wang, 2013; Ou, 2013; Li, 2011), content adaptive schemes (Kamstra, 2005; Li, 2013) and integer transform schemes (Alattar, 2004; Chen, 2010; Peng, 2012).

Following the development of RDH in the image field, a lot of efficient reversible watermarking algorithms have been proposed. In the early days, (Michiel, Arno, & Fons, 2003) proposed the reversible audio watermarking, which uses the redundant bits of audio coding to encode the watermark information, and then recover the host signal by restoring the original dynamic range in the decoder. Later on, many methods are proposed in time domain (Yan, 2008; Garcia-Hernandez, 2012; Bradley, 2015; Wang, 2014; Xiang, 2017), compressed domain (Li, 2008; Chen, 2007) and frequency domain (Huang, 2013; Yang, 2009). (Yan & Wang, 2008) refer to the method of extending the prediction error in the image reversible hiding proposed by (Tian, 2003) to construct the appropriate prediction model to realize the reversible data hiding. (Bradley & Alattar, 2015) proposed a high capacity reversible audio watermarking method based on the generalized reversible integer transform. (Xiang & Li, 2017) proposed an alterable prediction order data hiding method based on non-causal prediction, in which the double-embedding strategy in image data hiding is used to divide audio signal for two sets. In the compressed domain, (Li et al., 2008) designed an entropy coding algorithm, the perceptually unimportant indices in one segment of compressed speech bitstream are coded by the algorithm. (Huang et al., 2013) achieved the adaptive embedding of watermark information by processing the DCT coefficients. Particularly, using human auditory perception characteristics can achieve a better effect for reversible watermarking in the audio. (Unoki & Miyauchi, 2011) proposed a kind of non-perceptual audio reversible watermarking based on the delay characteristics of the human ear.

All the above RDH methods are designed for single channel audio. However, to balance the hearing effect, most of the audio files we can see on the Internet are stereo audio. In this paper we proposed a RDH method for stereo audio by exploiting the correlation between the two channels.

The rest of the article is organized as follows. Section 2 and Section 3 give the detailed description of the proposed method, including embedding and extracting procedure. The experiment results and comparisons are presented in Section 4. The article is concluded with a discussion in Section 5.

2. BASIC METHOD ON SINGLE CHANNEL

In this paper, we use the quantized audio samples as covers, and there are two kinds of common quantization bits in standard stereo audio, 8-bit quantization and 16-bit quantization.

For ease of understanding, we give a basic method of watermarking for the right channel signal of stereo audio M firstly.

2.1. Prediction Model

As shown in Figure 1, the length of the signal is N and x_i^R is an integer.

Figure 1. The right channel of audio M

| x_1^R | x_2^R | x_3^R | ... | x_{i-1}^R | x_i^R | x_{i+1}^R | ... | x_{N-1}^R | x_N^R |
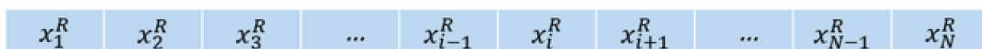
In the right channel, all samples are divided into even set and odd set to avoid that the modified samples affect the prediction of the current sample, and two-round embedding mechanism will be adopted.

In the first round, we only embed data into the PEs of samples in the even set. To generate PEs, the present sample x_{2i}^R is predicted as:

$$\tilde{x}_{2i}^R = u_{-3}^e x_{2i-3}^R + u_{-1}^e x_{2i-1}^R + u_1^e x_{2i+1}^R + u_3^e x_{2i+3}^R \tag{1}$$

$u_q^e \left(q = -3, -1, 1, 3 \right)$ is calculated by solving the following linear regression problem:

$$X_q^e * u_q^e = y_q^e \tag{2}$$

If N is even,

$$X_q^e = \begin{bmatrix} x_1^R & x_3^R & x_5^R & x_7^R \\ x_3^R & x_5^R & x_7^R & x_9^R \\ \vdots & \vdots & \vdots & \vdots \\ x_{2i-3}^R & x_{2i-1}^R & x_{2i+1}^R & x_{2i+3}^R \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-7}^R & x_{N-5}^R & x_{N-3}^R & x_{N-1}^R \end{bmatrix} \tag{3}$$

$$\mathbf{y_q^e} = \begin{bmatrix} x_4^R & x_6^R & \cdots & x_{2i}^R & \cdots & x_{N-4}^R \end{bmatrix} \tag{4}$$

If N is odd,

$$X_q^e = \begin{bmatrix} x_1^R & x_3^R & x_5^R & x_7^R \\ x_3^R & x_5^R & x_7^R & x_9^R \\ \vdots & \vdots & \vdots & \vdots \\ x_{2i-3}^R & x_{2i-1}^R & x_{2i+1}^R & x_{2i+3}^R \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-6}^R & x_{N-4}^R & x_{N-2}^R & x_N^R \end{bmatrix} \tag{5}$$

$$\mathbf{y_q^e} = \begin{bmatrix} x_4^R & x_6^R & \cdots & x_{2i}^R & \cdots & x_{N-3}^R \end{bmatrix} \tag{6}$$

With the least squares method, we can get

$$u_q^e = \left(X_q^e{}' X_q^e + w \right)^{-1} X_q^e{}' y_q^e \tag{7}$$

where w is a regular item to avoid NAN (not a number) problem. The value of w is also a factor that affects the accuracy of prediction model. According to lots of experiments, we can define an empirical value for w:

$$\begin{cases} w = diag\left(\mathbf{I} \right) \\ I = \begin{bmatrix} 1e-5 & 1e-5 & 1e-5 & 1e-5 & 1e-5 \end{bmatrix} \end{cases} \tag{8}$$

In the second round, we embed in the PEs of the odd set. Note that, in this round, the samples in the even set has been modified. The present sample x_{2i+1}^R is predicted as:

$$\tilde{x}_{2i+1}^R = u_{-3}^o \hat{x}_{2i-2}^R + u_{-1}^o \hat{x}_{2i}^R + u_1^o \hat{x}_{2i+2}^R + u_3^o \hat{x}_{2i+4}^R \tag{9}$$

where $\hat{x}_{2i-2}^R$, $\hat{x}_{2i}^R$, $\hat{x}_{2i+2}^R$, $\hat{x}_{2i+4}^R$ are the modified samples. The coefficients $\mathbf{u_q^o} \left(q = -3, -1, 1, 3 \right)$ is calculated by solving the following linear regression problem

$$\mathbf{X_q^o} * \mathbf{u_q^o} = \mathbf{y_q^o} \tag{10}$$

If N is even,

$$X_q^o = \begin{bmatrix} \hat{x}_2^R & \hat{x}_4^R & \hat{x}_6^R & \hat{x}_8^R \\ \hat{x}_4^R & \hat{x}_6^R & \hat{x}_8^R & \hat{x}_{10}^R \\ \vdots & \vdots & \vdots & \vdots \\ \hat{x}_{2i-2}^R & \hat{x}_{2i}^R & \hat{x}_{2i+2}^R & \hat{x}_{2i+4}^R \\ \vdots & \vdots & \vdots & \vdots \\ \hat{x}_{N-6}^R & \hat{x}_{N-4}^R & \hat{x}_{N-2}^R & \hat{x}_N^R \end{bmatrix} \tag{11}$$

$$\mathbf{y_q^o} = \begin{bmatrix} x_5^R & x_7^R & \cdots & x_{2i+1}^R & \cdots & x_{N-3}^R \end{bmatrix} \tag{12}$$

If N is odd,

$$
X_q^o = \begin{bmatrix} \hat{x}_2 & \hat{x}_4 & \hat{x}_6 & \hat{x}_8 \\ \hat{x}_4 & \hat{x}_6 & \hat{x}_8 & \hat{x}_{10} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{x}_{2i-2} & \hat{x}_{2i} & \hat{x}_{2i+2} & \hat{x}_{2i+4} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{x}_{N-7} & \hat{x}_{N-5} & \hat{x}_{N-3} & \hat{x}_{N-1} \end{bmatrix}
\tag{13}
$$

$$
\mathbf{y_q^o} = \begin{bmatrix} x_5^R & x_7^R & \cdots & x_{2i+1}^R & \cdots & x_{N-4}^R \end{bmatrix}
\tag{14}
$$

With the least squares method, we can get

$$
u_q^o = \left(X_q^o{}' X_q^o + w \right)^{-1} X_q^o{}' y_q^o
\tag{15}
$$

2.2. Embedding Procedure

Through the prediction model, we can get the prediction error calculated as:

$$
e_i^R = \tilde{x}_i^R - x_i^R
\tag{16}
$$

According to e_i^R, we embed the watermark bit b into the right channel of the stereo audio M as follows:

$$
\hat{e}_i^R = \begin{cases} 2e_i^R + b, & if \ e_i^R \in [-t, t] \\ e_i^R + t + 1, & if \ e_i^R \in (-t, +\infty) \\ e_i^R - t, & if \ e_i^R \in (-\infty, -t) \end{cases}
\tag{17}
$$

$b \in \{0,1\}$ represents the watermark bit, t is a threshold deciding the capacity.

Adding the modified error $\hat{e}_i^R$ to the current sample, we can get the marked signal:

$$
\hat{x}_i^R = x_i^R + \hat{e}_i^R
\tag{18}
$$

2.3. Extraction and Restoration Procedure

We can extract the watermark bit b as:

$$b = \overset{\wedge R}{e_i} \, mod \, 2 \quad \overset{\wedge R}{e_i} \in \left[-2t, 2t+1\right] \tag{19}$$

To restore the cover signal, it is necessary to recover the original prediction error firstly as:

$$e_i^R = \begin{cases} \overset{\wedge R}{e_i} / 2, & if \ \overset{\wedge R}{e_i} \in \left[-2t, 2t\right) \\ \overset{\wedge R}{e_i} - t - 1, & if \ \overset{\wedge R}{e_i} \in \left(2t, +\infty\right) \\ \overset{\wedge R}{e_i} + t, & if \ \overset{\wedge R}{e_i} \in \left(-\infty, -2t\right) \end{cases} \tag{20}$$

and we can restore the original sample as:

$$x_i^R = \overset{\wedge R}{x_i} + e_i^R \tag{21}$$

3. IMPROVED METHOD WITH INTER-CHANNEL CORRELATION

We then propose an improved method that applies to stereo audio referring to the basic method.

Previous work (Yan, 2008; Garcia-Hernandez, 2012; Bradley, 2015; Wang, 2014) show that efficiently exploiting correlation can increase message embedding capacity in the area of reversible watermarking. There is a strong correlation between two channels in most stereo audio files. In this section, we propose a RDH method on stereo audio by using such correlation. The overview of the proposed method is shown in Figure 2.

We will embed data into the right channel. First, we locate smooth regions of the right channel with inter-channel prediction. And then, in such smooth regions, we generate prediction error (PE) by intra-channel prediction. Finally, watermark is reversibly embedded by modifying the histogram of the PEs.

3.1. Correlation Between Two Channels

We first analyze the correlation of inter channel in stereo audio. We calculate the correlation coefficient of ten stereo audio clips selected in database randomly with Equation (22):

Figure 2. Framework of the reversible watermarking scheme
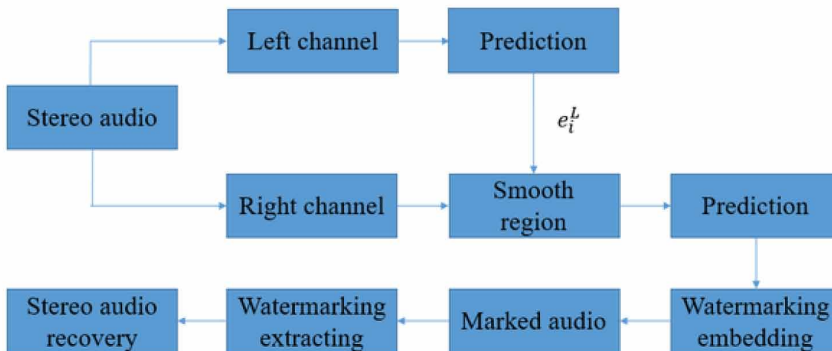
$$q = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}} \tag{22}$$

where X is the left channel signal and Y is the right channel. N is the length of the stereo audio. The correlation coefficients are listed in Table 1, which shows strong correlation between the two channels in most audio clips.

3.2. Prediction Model in Left Channel

In Figure 3, the length of the left channel signal of the stereo audio M is N and x_i^L is an integer.

The left channel is just predicted to select the smooth regions in the right channel. There is no message will be embedded in the left channel.

Usually, x_i^L's has strong local correlation with the context. So the local adjacent points $\{x_{i-k}^L \dots x_{i-3}^L, x_{i-2}^L, x_{i-1}^L, x_{i+1}^L, x_{i+2}^L, x_{i+3}^L \dots x_{i+k}^L\}$ can be used to predict x_i^L. The prediction value $\tilde{x}_i^L$ is given by (23):

$$\tilde{x}_i^L = \sum_{p=-k}^{-1} v_p x_{i-p}^L + \sum_{p=1}^{k} v_p x_{i-p}^L \tag{23}$$

Table 1. Correlation coefficient of ten audio clips

Index	Relevance
7	0.9983
10	0.7463
17	0.8554
24	0.8345
29	0.9603
39	0.5234
45	0.6091
58	0.7974
62	0.7089
70	0.9853

Figure 3. The left channel of audio M

x_1^L	x_2^L	x_3^L	...	x_{i-1}^L	x_i^L	x_{i+1}^L	...	x_{N-1}^L	x_N^L
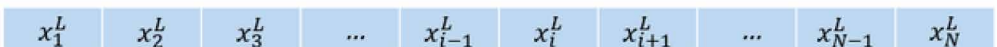
where v_p's are the prediction coefficients. In the prediction model, let $k = 3$, that means a sample point is estimated by the past three samples and the future three samples. The difference between predicted value $\tilde{x}_i^L$ and actual value x_i^L is calculated as:

$$e_i^L = \tilde{x}_i^L - x_i^L \tag{24}$$

We use the least squares regression method to get the best prediction coefficient $\mathbf{v_p}\left(p = -3, -2, -1, 1, 2, 3\right)$ such that:

$$\mathbf{X_p} * \mathbf{v_p} = \mathbf{y_p} \tag{25}$$

where $\mathbf{X_p}$ is a 3×6 matrix:

$$\mathbf{X_p} = \begin{bmatrix} \tilde{x}_{i-4}^L & \tilde{x}_{i-3}^L & \tilde{x}_{i-2}^L & \overline{x}_i^L & x_{i+1}^L & x_{i+2}^L \\ \tilde{x}_{i-3}^L & \tilde{x}_{i-2}^L & \tilde{x}_{i-1}^L & x_{i+1}^L & x_{i+2}^L & x_{i+3}^L \\ \tilde{x}_{i-2}^L & \tilde{x}_{i-1}^L & \overline{x}_i^L & x_{i+2}^L & x_{i+3}^L & x_{i+4}^L \end{bmatrix} \tag{26}$$

and $\mathbf{v_p} = \begin{bmatrix} v_{-1} & v_{-2} & v_{-3} & v_1 & v_2 & v_3 \end{bmatrix}'$, $\mathbf{y_p} = \begin{bmatrix} \tilde{x}_{i-1}^L & \overline{x}_i^L & x_{i+1}^L \end{bmatrix}'$.

We use the approximate value $\overline{x}_i^L$ to replace x_i^L as shown in (27) so that the data can be lossless recovered at the receiver side:

$$\overline{x}_i^L = \left(\tilde{x}_{i-1}^L + x_{i+1}^L\right) / 2 \tag{27}$$

According to the least-square method, the best prediction coefficients are given by:

$$\mathbf{v_p} = \left(\mathbf{X_p'}\mathbf{X_p} + \mathbf{w}\right)^{-1}\mathbf{X_p'}\mathbf{y_p} \tag{28}$$

3.3. Payload Assignment

We adaptively assign the payloads according to the degree of smoothness of the right channel, which is estimated by the information from the left channel. With the method described in Subsection prediction model in left channel, we can get the PEs of the left channel such that:

$$E = \left(e_1^L, e_2^L, e_3^L, e_4^L, \ldots, e_N^L\right) \tag{29}$$

We define the set of the smooth samples in the right channel as:

$$S^R = \{x_i^R \mid \left|e_i^L\right| < tr\} \tag{30}$$

where tr is a threshold and we select tr as an integer. The set of S^R is then divided into a series of subsets according to the degrees of smoothness such that:

$$l_j^R = \{x_i^R \mid j-1 \leq \left|e_i^L\right| < j\} \ 1 \leq j \leq tr \tag{31}$$

the set of S^R can be represented equally as:

$$S^R = \left(l_1^R, l_2^R, l_3^R, \ldots, l_j^R, \ldots, l_{tr}^R\right) \tag{32}$$

Then, we use a randomly generated message to do a tentative embedding in each subset l_j^R using the basic method in single channel, by which we can estimate the capacity c_j and corresponding distortion d_j of the subset l_j^R. The c_j is the number of embeddable watermark bits, and d_j is calculated as (33):

$$d_j = \frac{\sum_{i=1}^{N} \left(\hat{x}_i^R - x_i^R\right)^2}{\sum_{i=1}^{N} \left(x_i^R\right)^2} \tag{33}$$

where $x_i^R \in l_j^R$, and $\hat{x}_i^R$ is the sample after tentative embedding. With such method, we get the estimated capacity in each subset denoted as:

$$C^R = c_1, c_2, c_3, \ldots c_j, \ldots, c_{tr} \tag{34}$$

and the corresponding distortion set denoted as:

$$D^R = d_1, d_2, d_3, \ldots d_j, \ldots, d_{tr} \tag{35}$$

For a given message length C, partial sample subsets are enough for accommodate the message. To get the optimal combination of subsets for C, we solve the following 0-1 programming problem:

$$\begin{cases} minimize \quad \sum_{j=1}^{tr} h_j * d_j \\ \quad\quad\quad\quad\quad\quad\quad\quad h_j \in \{0,1\} \\ subject\ to \ \sum_{j=1}^{tr} h_j * c_j \geq C \end{cases} \tag{36}$$

The optimal solution is denoted as:

$$S_o^R = \left(l_{o1}^R, l_{o2}^R, \ldots, l_{oh}^R \right) (0 < o1 < o2 < \ldots < oh \le tr) \tag{37}$$

Finally, with the basic method in single channel, we embed message into S_o^R in turn according to the location in the right of audio M.

3.4. Embedding Procedure

There are some auxiliary information should be embedded into the cover audio (the right channel of audio M) for extraction and restoration. The parameters t, S_o^R, u_q^e and u_q^o are necessary to be embedded. The required space is calculated as follows:

1. Usually, $0 < t < 15$ is enough for the method, and it needs 4 bits;
2. $S_o^R = \left(l_{o1}^R, l_{o2}^R, \ldots, l_{oh}^R \right)$ $(0 < o1 < o2 < \ldots < oh \le tr)$. Usually $tr = 40$ is enough, and thus we use 40 bits to label which subset is chosen for embedding with the bit "1" represents selected subset;
3. u_q^e and u_q^o are decimals range from -1 to 1, and they need 30 bits.

The total size of the auxiliary information is 74 bits which occupies only a small amount of samples in audio M. To ensure adequate space, we use the last 80 samples in the right channel of audio M to embed the auxiliary information.

The details of embedding procedure are stated as follows:

Step 1: Replace the LSBs of the last 80 samples in the right channel of audio M with parameters t, S_o^R, u_q^e and u_q^o, and then the LSBs will be reversibly embedded as part of the watermark message.

Step 2: Select the optimal embedding samples S_o^R in the right channel of audio M according to the prediction error in the left channel of audio M.

Step 3: For the selected samples S_o^R, we call the algorithm in Section 2 for embedding.

3.5. Extraction and Restoration Procedure

In the extracting procedure, our purpose is to get watermark message from the marked audio and restore the original audio. The details are stated as follows:

Step 1: Preprocess the marked audio. Read the last 80 samples in right channel to get threshold t in (17), chosen region S_o^R in (37), coefficients u_q^e and u_q^o.

Step 2: Calculate the prediction value of left channel and the prediction error $E = \left(e_1^L, e_2^L, e_3^L, e_4^L, \ldots, e_N^L \right)$. Determine the embedding region according to S_o^R.

Step 3: Recover $\hat{e}_i^R$ of right channel according to u_q^e and u_q^o, first odd set, then even set. Extract *watermark* and the LSB using (19) and restore the audio M' using (21).

Step 4: Use the LSB to replace the last 80 samples in right channel to get original audio M.

4. EXPERIMENTAL RESULTS

In the testing, we get 4 audio files for example in audio database to evaluate the performance of the proposed algorithm. Uniformly, all of the clips are standard stereo audio and the sampling frequency is 44.1k. We intercept 200000 audio points for the audio files in order to perform intuitively. Embedding capacity and audio distortion are two important criterions to be calculated. The embedding capacity is represented by the amount of data embedded in audio. And the signal to noise ratio (SNR) is used to measure the watermark distortion:

$$\text{SNR} = 10\lg\left(\frac{\sum_{i=1}^{N}\left(\hat{x}_i^{R} - x_i^{R}\right)^2}{\sum_{i=1}^{N}\left(x_i^{R}\right)^2}\right) \tag{38}$$

We compare the proposed algorithm with three existing work in reversible audio data hiding algorithm (Xiang, 2017; Nishimura, 2011; Xiang, 2012) for clip 08, 16, 53, 70. Figure 4, 5, 6, 7 show the experimental performance of four algorithm on the example clips.

Figure 8 shows the average SNR of 70 audio clips in database for different capacity. The result of these methods are listed in Table 2. We can observe that the proposed method outperforms the existing work in reversible audio data hiding (Xiang, 2017; Nishimura, 2011; Xiang, 2012).
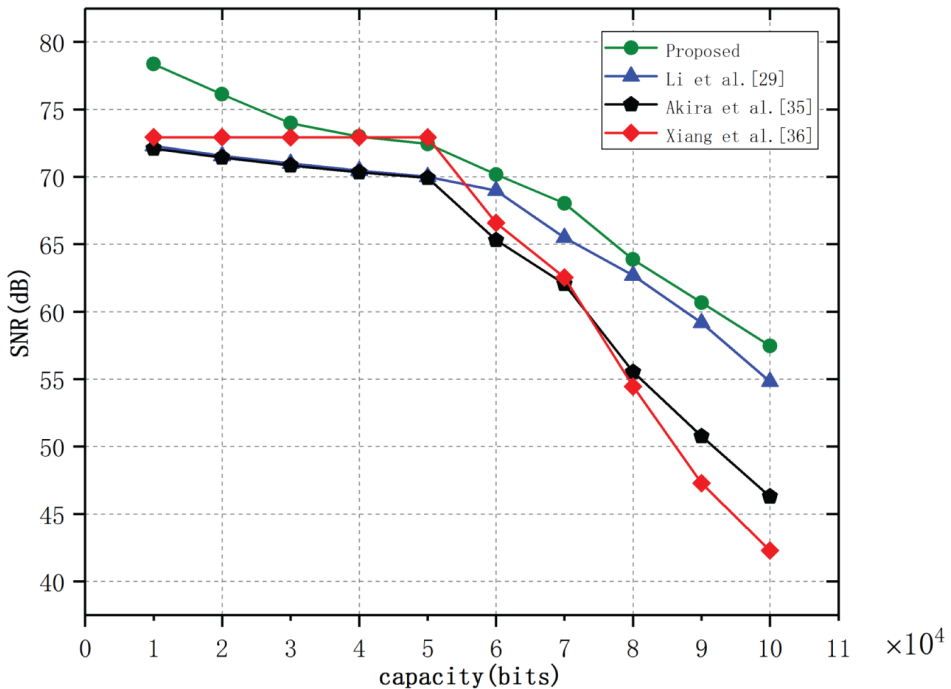
Figure 4. Distortion comparison for clip 08
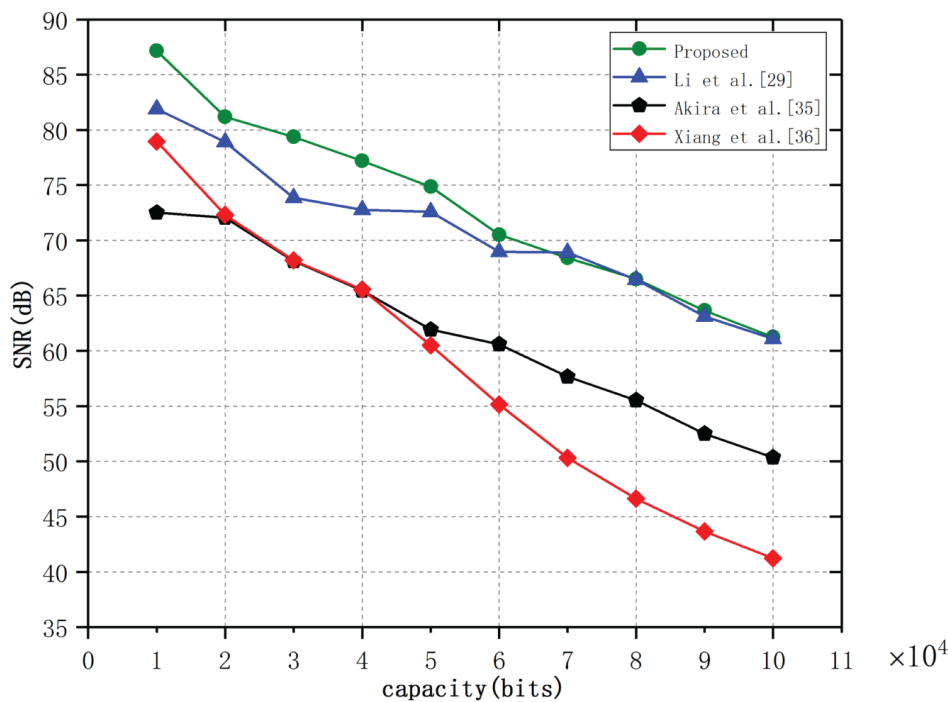
Figure 5. Distortion comparison for clip 16

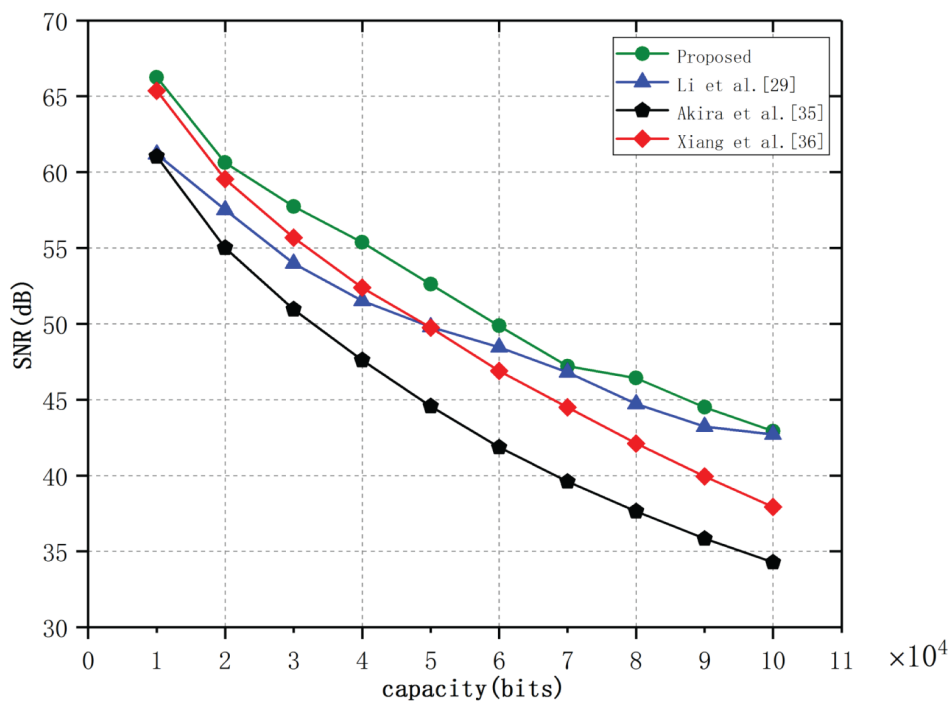**Figure 6. Distortion comparison for clip 53**

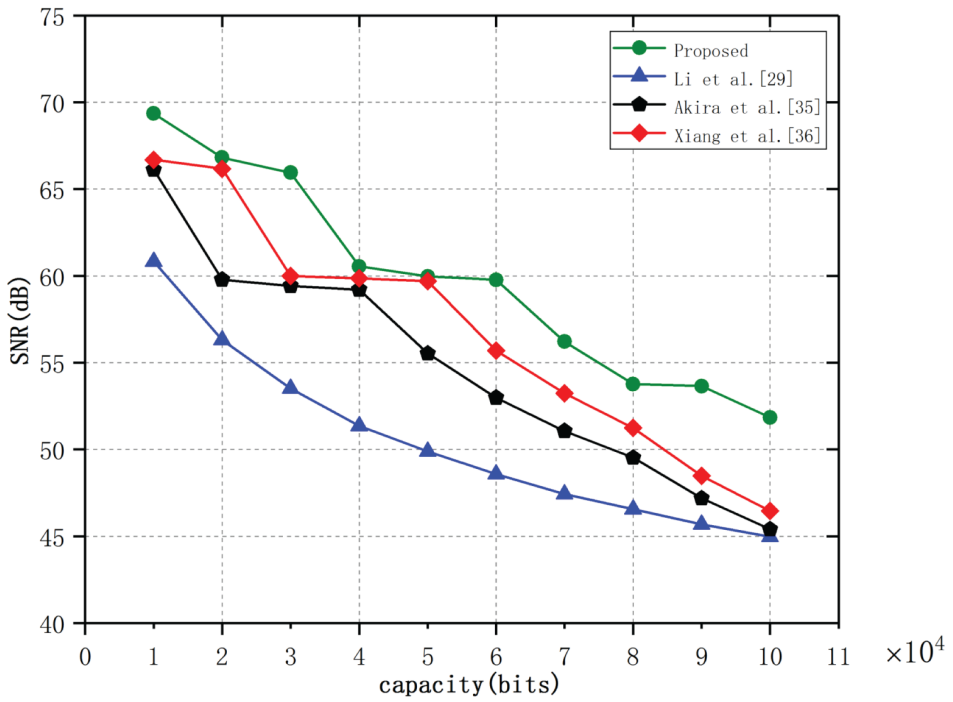**Figure 7. Distortion comparison for clip 70**

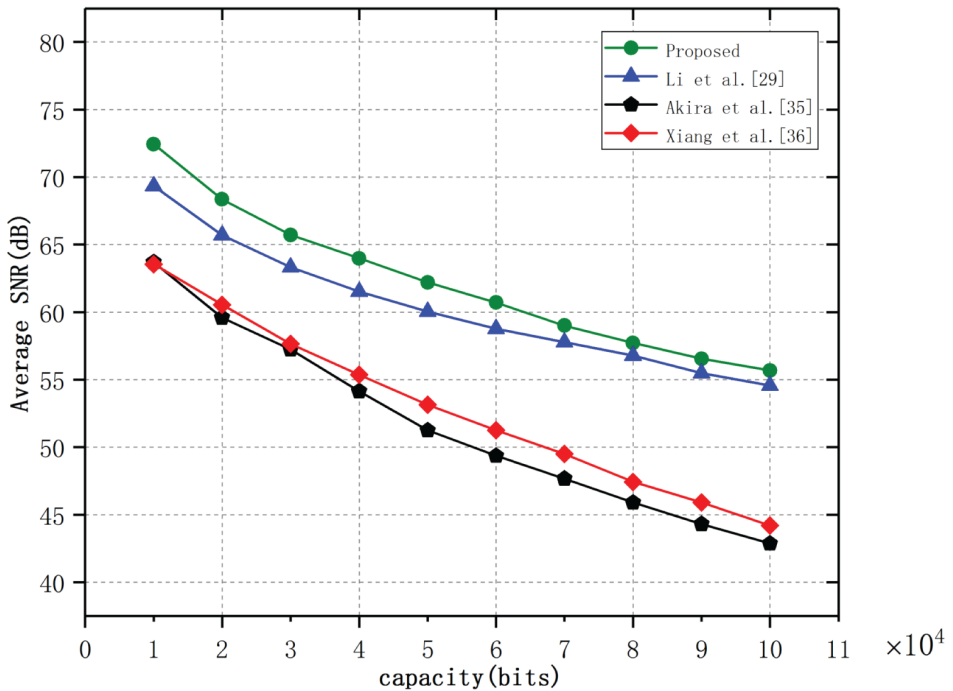**Figure 8. Average SNR for stereo audio database for different capacity**

Table 2. Average SNR for stereo audio database for different capacity

Capacity×10⁴	1	2	3	4	5	6	7	8	9	10
Proposed	72.45	68.35	65.72	63.98	62.20	60.71	59.01	57.73	56.55	55.68
Li et al.	69.35	65.69	63.34	61.52	60.04	58.77	57.79	56.78	55.50	54.58
Akira et al.	63.69	59.60	57.24	54.14	51.25	49.37	47.67	45.90	44.31	42.88
Xiang et al.	63.55	60.55	57.65	55.36	53.14	51.25	49.50	47.44	45.90	44.21

5. CONCLUSION

In this paper, we proposed a reversible watermarking algorithm for stereo audio based on the inter-channel correlation. The message is only embedded into the right channel while the embedding regions are determined with the help of the smoothness degrees in the left channel. By exploring such inter-channel correlation, we can effectively avoid region that may introduce large modification costs. Experimental results illustrate that the proposed method achieves lower distortion than several traditional methods that use single audio channel.

ACKNOWLEDGMENT

REFERENCES

Alattar, A. M. (2004). Reversible watermark using the difference expansion of a generalized integer transform. *IEEE Transactions on Image Processing*, *13*(8), 1147–1156. doi:10.1109/TIP.2004.828418 PMID:15326856

BartonJ. M. (1997). US Patent 5.

Bradley, B., & Alattar, A. (2015). High-capacity invertible data-hiding algorithm for digital audio. In Security, Steganography, and Watermarking of Multimedia Contents VII (pp. 789-801).

Celik, M. U., Sharma, G., Tekalp, A. M., & Saber, E. (2005). Lossless generalized-LSB data embedding. *IEEE Transactions on Image Processing*, *14*(2), 253–266. doi:10.1109/TIP.2004.840686 PMID:15700530

Chen, O. T. C., & Liu, C. H. (2007). Content-dependent watermarking scheme in compressed speech with identifying manner and location of attacks. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(5), 1605–1616. doi:10.1109/TASL.2007.896658

Chen, X., Li, X., Yang, B., & Tang, Y. (2010). Reversible image watermarking based on a generalized integer transform. In *Proc. IEEE ICASSP* (pp. 2382–2385). doi:10.1109/ICASSP.2010.5496175

EBU Committee. (n.d.). Sound quality assessment material recordings for subjective tests. Retrieved from https://tech.ebu.ch/publications/sqamcd

Garcia-Hernandez, J. J. (2012). Exploring reversible digital watermarking in audio signals using additive interpolation-error expansion. In *2012 Eighth International Conference on* Intelligent Information Hiding and Multimedia Signal Processing *(IIH-MSP)*. doi:10.1109/IIH-MSP.2012.40

Hong, W., Chen, T. S., & Shiu, C. W. (2009). Reversible data hiding for high quality images using modification of prediction errors. *Journal of Systems and Software*, *82*(11), 1833–1842. doi:10.1016/j.jss.2009.05.051

Huang, X., Ono, N., Echizen, I., & Nishimura, A. (2013). Reversible Audio Information Hiding Based on Integer DCT Coefficients with Adaptive Hiding Locations. In *International Workshop on Digital Watermarking* (pp. 376-389). Springer.

Kamstra, L., & Heijmans, H. J. A. M. (2005). Reversible data embedding into images using wavelet techniques and sorting. *IEEE Transactions on Image Processing*, *14*(12), 2082–2090. doi:10.1109/TIP.2005.859373 PMID:16370461

Li, M., Jiao, Y., & Niu, X. (2008). Reversible Watermarking for Compressed Speech. In *Eighth International Conference on Intelligent Systems Design and Applications ISDA '08* (pp. 197–201). doi:10.1109/ISDA.2008.295

Li, X., Li, J., Li, B., & Yang, B. (2013). High-fidelity reversible data hiding scheme based on pixel-value-ordering and prediction-error expansion. *Signal Processing*, *93*(1), 198–205. doi:10.1016/j.sigpro.2012.07.025

Li, X., Yang, B., & Zeng, T. (2011). Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection. *IEEE Transactions on Image Processing*, *20*(12), 3524–3533. doi:10.1109/TIP.2011.2150233 PMID:21550888

Lin, S. J., & Chung, W. H. (2011). The scalar scheme for reversible information-embedding in gray-scale signals: Capacity evaluation and code constructions. *IEEE Transactions on Information Forensics and Security*, *7*(4), 1155–1167. doi:10.1109/TIFS.2012.2197614

Malvar, H. S., & Florencio, D. A. F. (2003). Improved spread spectrum: A new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, *51*(4), 898–905. doi:10.1109/TSP.2003.809385

Michiel, V. D. V., Arno, V. L., & Fons, B. (2003). Reversible Audio Watermarking. Audio Engineering Society.

Naderahmadian, Y., & Hosseini-Khayat, S. (2014). Fast and robust watermarking in still images based on QR decomposition. *Multimedia Tools and Applications*, *72*(3), 2597–2618. doi:10.1007/s11042-013-1559-9

Ni, Z., Shi, Y. Q., Ansari, N., & Su, W. (2006). Reversible data hiding. *IEEE Transactions on Circuits and Systems for Video Technology*, *16*(3), 354–362. doi:10.1109/TCSVT.2006.869964

Nishimura, A. (2011). Reversible audio data hiding using linear prediction and error expansion. In *Proc. of IIHMSP2011* (pp. 318–321). doi:10.1109/IIHMSP.2011.76

Ou, B., Li, X., Zhao, Y., Ni, N., & Shi, Y. Q. (2013). Pairwise prediction-error expansion for efficient reversible data hiding. *IEEE Transactions on Image Processing*, *22*(12), 5010–5021. doi:10.1109/TIP.2013.2281422 PMID:24043388

Peng, F., Li, X., & Yang, B. (2012). Adaptive reversible data hiding scheme based on integer transform. *Signal Processing*, *92*(1), 54–62. doi:10.1016/j.sigpro.2011.06.006

Qin, C., Chang, C. C., & Chen, P. Y. (2012). Self-embedding fragile watermarking with restoration capability based on adaptive bit allocation mechanism. *Signal Processing*, *92*(4), 1137–1150. doi:10.1016/j.sigpro.2011.11.013

Rani, A., Raman, B., & Kumar, S. (2014). A robust watermarking scheme exploiting balanced neural tree for rightful ownership protection. *Multimedia Tools and Applications*, *72*(3), 2225–2248. doi:10.1007/s11042-013-1528-3

Thodi, D. M., & Rodriguez, J. J. (2007). Expansion embedding techniques for reversible watermarking. *IEEE Transactions on Image Processing*, *16*(3), 721–730. doi:10.1109/TIP.2006.891046 PMID:17357732

Tian, H., Xiao, Y., Cao, G., Ding, J., & Ou, B. (2016). Robust watermarking of mobile video resistant against barrel distortion. *China Communications*, *13*(9), 131–138. doi:10.1109/CC.2016.7582305

Tian, J. (2003). Reversible data embedding using a difference expansion. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(8), 890–896. doi:10.1109/TCSVT.2003.815962

Unoki, M., & Miyauchi, R. (2011). *Reversible Watermarking for Digital Audio Based on Cochlear Delay Characteristics*. In Intelligent Information Hiding and Multimedia Signal Processing. doi:10.1109/IIHMSP.2011.99

Wang, F., Xie, Z., & Chen, Z. (2014). *High Capacity Reversible Watermarking for Audio by Histogram Shifting and Predicted Error Expansion*. The Scientific Word Journal.

Wang, J., & Ni, J. (2013). A GA optimization approach to HS based multiple reversible data hiding. In *Proc. IEEE WIFS* (pp. 203–208). doi:10.1109/WIFS.2013.6707819

Xiang, S. (2012). Non-integer expansion embedding for prediction-based reversible watermarking. In *Proc. 14th Int. Conf.* (pp. 224-239).

Xiang, S., & Li, Z. (2017). Reversible audio data hiding algorithm using noncausal prediction of alterable orders. *EURASIP Journal on Audio, Speech, and Music Processing*, (1), 4. doi:10.1186/s13636-017-0101-9

Yan, D., & Wang, R. (2008). Reversible Data Hiding for Audio Based on Prediction Error Expansion. In *Intelligent Information Hiding and Multimedia Signal Processing* (pp. 249-252).

Yang, Y., Huang, R., & Xu, M. (2009). A Novel Audio Watermarking Algorithm for Copyright Protection Based on DCT Domain. In *Second International Symposium on Electronic Commerce and Security ISECS '09* (pp. 184-188).

Zhang, W., Chen, B., & Yu, N. (2012). Improving various reversible data hiding schemes via optimal codes for binary covers. *IEEE Transactions on Image Processing*, *21*(6), 2991–3003. doi:10.1109/TIP.2012.2187667 PMID:22345532

Zhang, X., & Wang, S. (2008). Fragile watermarking with error-free restoration capability. *IEEE Transactions on Multimedia*, *10*(8), 1490–1499. doi:10.1109/TMM.2008.2007334

Zhang, X., & Wang, S. (2009). Fragile watermarking scheme using a hierarchical mechanism. *Signal Processing*, *89*(4), 675–679. doi:10.1016/j.sigpro.2008.10.001

Yuanxin Wu is a graduate student at CAS Key Laboratory of Electro-magnetic Space Information, Chinese Academy of Sciences, University of Science and Technology of China, Hefei, China. Her research interests include reversible audio watermarking and audio tamper detection.

Wen Diao is currently a graduate student at CAS Key Laboratory of Electro-magnetic Space Information, Chinese Academy of Sciences, University of Science and Technology of China, Hefei, China. His research interests include audio watermarking and its application.

Dongdong Hou received the BS degree from the Hefei University of Technology, Hefei, China, in 2014, and is currently working toward the PhD degree in electronic engineering at the University of Science and Technology of China, Hefei, China. His research interests include multimedia security, image processing, and information hiding.

Weiming Zhang received his BS degree and PhD degree in 1999 and 2005 respectively from Information Engineering University, Zhengzhou, China. Currently, he is a full professor with the School of Information Science and Technology, University of Science and Technology of China. His research interests include information hiding, multimedia security, and privacy-preserving data searching and analysis. He has published more than 80 papers in journals and conferences including IEEE Trans. Information Theory, IEEE Trans. Image Processing,IEEE Trans. Inf. Foren. & Sec., and Information Hiding.