IEICE TRANS. INF. & SYST., VOL.E101–D, NO.1 JANUARY 2018

LETTER *Special Section on Enriched Multimedia — Potential and Possibility of Multimedia Contents for the Future —*

Tolerance Evaluation of Audio Watermarking Method Based on Modification of Sound Pressure Level between Channels

Harumi MURATA[†a)], **Akio OGIHARA**[††], *Members*, *and* **Shigetoshi HAYASHI**, *Nonmember*

SUMMARY We have proposed an audio watermarking method based on modification of sound pressure level between channels. This method is focused on the invariability of sound localization against sound processing like MP3 and the imperceptibility about slightly change of sound localization. In this paper, we investigate about tolerance evaluation against various attacks in reference to IHC criteria.

key words: *audio watermarking, sound pressure level, IHC criteria*

1. Introduction

Recently, copyright infringement has become a social problem such that the illegal reproduction is distributed on the Internet. Hence, audio watermarking methods, which embed proprietary data into digital audio data, have attracted attention as prevention techniques against copyright infringement.

Y. Li et al. have proposed an underdetermined blind source separation method [1]. In this method, source separation is achieved by estimating mixing matrix. This matrix corresponds to a ration of sound pressure level among each channel. In other words, mixing matrix represents approximate source localization.

We focus on two characteristics of sound source. One is location of a dominant sound source will be maintained even if signal processing like MP3 is performed. Two is deterioration of sound quality is imperceptible even if location of a dominant sound source is modified slightly. Therefore, we have proposed an embedding method based on modification of sound pressure level between channels [2]. It has been confirmed that we can embed the watermarks with high sound quality and tolerance against MP3.

However, tolerance against various attacks except for MP3 is not evaluated. Moreover, it is hard to say that the tolerance against MP3 is enough. Hence, in this paper, we introduce BCH code which is one of error correcting code for the tolerance improvement as with [3], and we investigate about tolerance evaluation against various attacks in reference to IHC criteria [4], [5].

Manuscript received April 5, 2017.
Manuscript revised July 26, 2017.
Manuscript publicized October 16, 2017.

[†]The author is with the Department of Information Engineering, School of Engineering, Chukyo University, Toyota-shi, 470–0393 Japan.
[††]The author is with the Department of Informatics, Faculty of Engineering, Kindai University, Higashi-Hiroshima-shi, 739–2116 Japan.
a) E-mail: murata_h@sist.chukyo-u.ac.jp
DOI: 10.1587/transinf.2017MUL0003

2. Audio Watermarking Method Based on Modification of Sound Pressure Level between Channels

In this section, we explain about estimating mixing matrix [1] and embedding and extracting of watermarks [2].

2.1 Estimating Mixing Matrix

Underdetermined blind source separation can be achieved two-stage sparse representation approach [1]. The first challenging task of this approach is to estimate precisely the unknown mixing matrix. The second task of the two-stage approach is to estimate the source matrix using a standard linear programming algorithm.

In this paper, we focus on the first task and we consider the following noise-free model:

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where the mixing matrix $\mathbf{A} \in R^{n\times m}$ is unknown. The matrix $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_K] \in R^{m\times K}$ is composed of the m unknown sources, and the only observable $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_K] \in R^{n\times K}$ is a data matrix that has rows containing mixtures of sources.

Below, we describe about detailed algorithm for estimating the mixing matrix $\mathbf{A}$, and the output estimated matrix is denoted as $\bar{\mathbf{A}}$. Also, let a matrix $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_{N_0}]$ be a matrix to store the estimated columns during intermediate steps. Initially, $\mathbf{E}$ is an empty matrix.

Step A1. Apply a wavelet packets transformation to every row of the matrix $\mathbf{X} \in R^{n\times K}$. A time-frequency representation matrix $\tilde{\mathbf{X}}$ is obtained.
Step A2. Find a submatrix $\hat{\mathbf{X}}$ of $\tilde{\mathbf{X}}$ such that the norm of its each column is greater than ξ_1, where ξ_1 is a positive constant chosen in advance.
Step A3. For $n_1 = 1$ to n, do the following.

Step A3.1. If the absolute value of an entry of the n_1th row of $\hat{\mathbf{X}}$ is less than a preset positive constant ξ_2, then we shall remove the corresponding column of $\hat{\mathbf{X}}$ containing the entry. Suppose that there are K_1 columns of $\hat{\mathbf{X}}$ left and denote their indexes as $q_1, \ldots, q_{K_1}$. A new ratio matrix using the left K_1 columns of $\hat{\mathbf{X}}$ is constructed.

$$\tilde{\tilde{\mathbf{X}}} = \begin{bmatrix} \frac{\hat{x}_1(q_1)}{\hat{x}_{n_1}(q_1)} & \cdots & \frac{\hat{x}_1(q_{K_1})}{\hat{x}_{n_1}(q_{K_1})} \\ \vdots & \vdots & \vdots \\ \frac{\hat{x}_n(q_1)}{\hat{x}_{n_1}(q_1)} & \cdots & \frac{\hat{x}_n(q_{K_1})}{\hat{x}_{n_1}(q_{K_1})} \end{bmatrix} \tag{2}$$

Step A3.2. For $n_2 = 1$ to n, $n_2 \neq n_1$, do the following.

Step A3.2.1. Find the minimum $\tilde{r}_{n_2}$ and maximum $\tilde{R}_{n_2}$ of $\tilde{\tilde{\mathbf{x}}}_{n_2}$, the n_2th row of $\hat{\mathbf{X}}$. Divide the entry range (interval) $[\tilde{r}_{n_2}, \tilde{R}_{n_2}]$ of $\tilde{\tilde{\mathbf{x}}}_{n_2}$ equally into M_0 subintervals (bins), where M_0 is a chosen large positive integer. Then, divide the matrix $\tilde{\tilde{\mathbf{X}}}$ into M_0 submatrices, donated as $\tilde{\tilde{\mathbf{X}}}_1, \ldots, \tilde{\tilde{\mathbf{X}}}_{M_0}$, such that all entries of the n_2th row of $\tilde{\tilde{\mathbf{X}}}_k$ are in the kth bin, $k = 1, \ldots, M_0$.

Step A3.2.2. From the submatrix set $\{\tilde{\tilde{\mathbf{X}}}_k, k = 1, \ldots, M_0\}$, delete those submatrices of which the number of columns is less than J_1, where J_1 is a chosen positive integer. The new set of submatrices is denoted as $\{\tilde{\tilde{\mathbf{X}}}_{j_k}, k = 1, \ldots, N_1\}$.

Step A3.3. Calculating the mean of all the column vector of the matrix $\tilde{\tilde{\mathbf{X}}}_{j_k}$ and normalizing the averaged column vector to the unit norm, we obtain an estimated column vector, denoted as $\mathbf{e}_i$, of the mixing matrix $\mathbf{A}$.

Step A4. After carrying out the loops above, we can obtain a set of estimated columns, denoted as $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_{N_0}]$. In the final step, we need to remove the duplication of column vectors in $\mathbf{E}$. Finally, the obtained matrix, denoted as $\bar{\mathbf{A}}$, is taken as the estimate of the original mixing matrix $\mathbf{A}$.

In this algorithm, ξ_1, ξ_2 are related to the amplitude of the entries of the data matrix domain. Let Q_1 denotes the maximum of the norms of all columns of the data matrix, and Q_2 denotes the maximum of the amplitude of the entries of the data matrix. Then, ξ_1, ξ_2 are calculated as following equations.

$$\xi_1 = th_1 \times Q_1 \tag{3}$$

$$\xi_2 = th_2 \times Q_2 \tag{4}$$

where th_1 and th_2 are variables.

2.2 Embedding of Watermarks

Step B1. Host signal is divided into consecutive N-length segment.

Step B2. Location of the most dominant source in the segment is extracted based on algorithm for estimating the mixing matrix [1]. We divide angle of arrival equally into M_0 equal part and we call these part as "band".

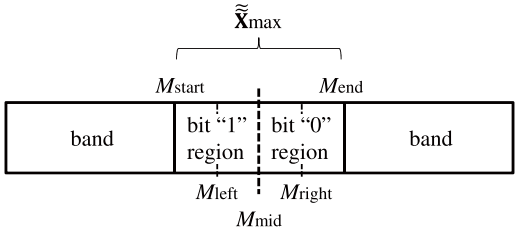

Fig. 1 Variable definition and embedding region.

Step B2.1 Submatrix set $\tilde{\tilde{\mathbf{X}}}_{j_k}(k = 1, \ldots, N_1)$ is calculated from Step A1. to Step A3.2.2..

Step B2.2 The maximum number of columns is defined as $\tilde{\tilde{\mathbf{X}}}_{\max}$. Suppose that there are K columns of $\tilde{\tilde{\mathbf{X}}}_{\max}$ left and denote their indexes as $P(k)$ $(k = 2, 3, \ldots, K)$.

Step B3 Watermarks are embedded by changing a sound source location. Subsequently, $\tilde{\tilde{\mathbf{X}}}_{\max}$ is divided in half again as shown in Fig. 1. The left side of the band corresponds to bit "1" and the other side corresponds to bit "0". Then, the location of the dominant source is modified according to watermark bit. Indexes of the left side are $P_{\text{left}}(k_2)$ $(k_2 = 1, 2, \ldots, K_2)$, and indexes of the right side are $P_{\text{right}}(k_3)$ $(k_3 = 1, 2, \ldots, K_3)$. Update formula is shown below.

$$\begin{cases} \hat{x}_1(P(k)) \leftarrow (1+\alpha) \times \hat{x}_1(P(k)) \ (k = 1, \ldots, K) \\ \hat{x}_2(P(k)) \leftarrow (1-\alpha) \times \hat{x}_2(P(k)) \ (k = 1, \ldots, K) \end{cases} \tag{5}$$

If bit "1" is embedded, $P(k)$ is $P_{\text{right}}(k_3)$. Otherwise, $P(k)$ is $P_{\text{left}}(k_2)$. Also, α is calculated from the following equation.

$$\alpha = \begin{cases} \frac{X_n(P_{\text{right}}(k)) - M_{\text{left}}}{X_n(P_{\text{right}}(k)) + M_{\text{left}}} & \text{to embed "1"} \\ \frac{X_n(P_{\text{left}}(k)) - M_{\text{right}}}{X_n(P_{\text{left}}(k)) + M_{\text{right}}} & \text{to embed "0"} \end{cases} \tag{6}$$

where $X_n(P(k)) = \hat{x}_2(P(k))/\hat{x}_1(P(k))$. Sound source localization can be transfered by above operation.

Step B4. For all segments, Step B2. and Step B3. are repeated.

2.3 Extracting of Watermarks

Stego signal is divided into consecutive N-length segment as same as the embedding process. Moreover, we divided angle of arrival equally into M_0 equal part and the band in which the most dominant source is included is divided in half again. If the number of elements of the left side is larger than that of the right side, the watermark bit "1" is extracted. Otherwise, the watermark bit "0" is extracted.

2.4 Tolerance Improvement Using BCH Code

In [2], we investigated only the tolerance against MP3. Moreover, it is hard to say that the tolerance against MP3 is enough. Hence, we introduce BCH code which is one of error correcting code for the tolerance improvement as with [3].

The BCH code is a representative cyclic code compatible with various error correcting requirements and code lengths and is relatively easy to encode and decode. Denoting the code length and number of information bits by l and p, respectively, we refer to such a code as a BCH(l, p). The BCH encodes p consecutive bits of an embedded bit sequence. BCH code with $2t$ roots can correct up to t errors. In this paper, decoding is performed by a Euclidean method.

The proposed method embeds payloads and synchronization codes as watermarks. Payloads are encoded by BCH(31, 11) and BCH(15, 5) codes, and their inclusion is preceded by synchronization codes. The BCH(31, 11) and BCH(15, 5) codes can correct errors up to 5 bits and 3 bits, respectively.

3. Experimental Results

In order to confirm the validity of the proposed method, we examined bit error rate (BER) and objective difference grade (ODG). For testing, we used 8 music data selected from "SQAM recordings for subjective test [5]" and 12 music data selected from "RWC music database: music genre [4]," 60 seconds duration, at a 44.1 kHz sampling rate, with stereo channel. 263-bit BCH-encoded payloads and 63-bit M-sequence of synchronization code per 15 seconds were embedded into each music data.

The embedding parameters were set $N = 1024$, $M_0 = 400$, $J_1 = 100$, $th_1 = 0.3$, $th_2 = 0.1$, and $[\tilde{r}_{n_2}, \tilde{R}_{n_2}] = [-100, 100]$. These parameters were determined by preliminary experiment. Moreover, interval $[\tilde{r}_{n_2}, \tilde{R}_{n_2}]$ was changed by attacks, and these values were fixed.

3.1 Tolerance Evaluation against Attacks

We evaluated for the tolerance against the following attacks.

- MP3 128 kbps (joint stereo)
- A series of attacks that mimic D/A and A/D conversions
- MP3 128 kbps (joint stereo) tandem coding
- MPEG4 HE-AAC 96 kbps
- Gaussian noise addition (overall average SNR 36 dB)

BER of watermarks was defined as

$$\text{BER} = \frac{\text{number of error bits}}{180 \text{ bits}} \cdot 100\ [\%] \tag{7}$$

where the denominator represents number of watermarks for 30 seconds interval [5]. The payloads were extracted from consecutive 45 seconds of stego data from which the initial sample is randomly chosen in the initial 15 seconds. BER is defined as the number of mismatched bits between the embedded and extracted payloads relative to the 180 bits that are embedded into 15 to 45 seconds of the stego data.

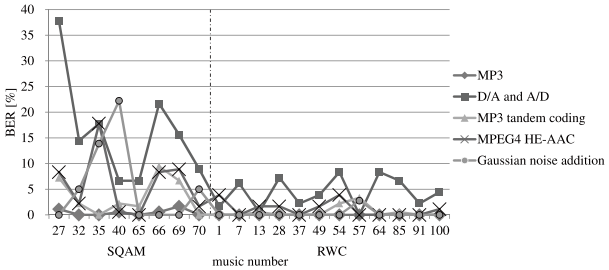

Fig. 2 The results of BER [%].

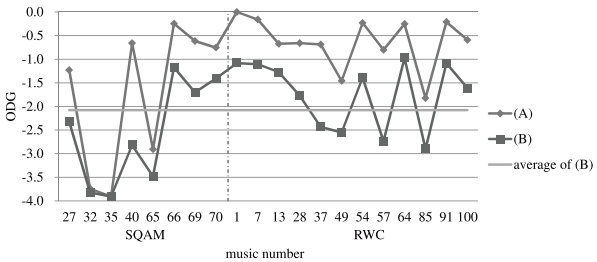

Fig. 3 ODG: (A) comparing host signals to stego signals, (B) comparing host signals to degraded signals.

Figure 2 shows the results of BER. However, synchronization detection was not operated in this paper. From the results, music data of RWC were satisfied with the criteria for all attacks and music data of SQAM were high tolerance against MP3. However, the tolerance against Gaussian noise addition was more than 10% in 2 pieces of music. These music data have a lot of period close to the no sound data period, and it is considered that sound localization was changed by Gaussian noise addition. In addition, there were no tolerances against D/A and A/D conversions in music data of SQAM.

3.2 Objective Sound Quality Evaluation

We evaluated the objective sound quality by PEAQ [6]. PEAQ uses some features of both host and stego signals and represents the quality comparison result as ODG. The ODG ranges from 0 to −4, with higher values indicating greater watermark transparency. Figure 3 shows the ODG results, which are calculated as follows.

(A) ODGs between the original PCM host signals and stego signals. The ODG values should be more than −2.5.
(B) ODGs between the original PCM host signals and MP3-coded stego signals. The arithmetic mean of the 20 ODGs should be more than −2.0.

The ODGs were satisfied with the criteria in music data of RWC. However, 3 pieces of music were not satisfied with the criteria in music data of SQAM. It is considered that

sound quality degradation was caused by an embedding process for a period close to the no sound data period. Hence, it needs to modify the embedding position such that watermarks are not embedded into a period close to the no sound data period.

4. Conclusion

We proposed an embedding method based on modification of sound pressure level between channels and investigated about tolerance evaluation against various attacks in reference to IHC criteria. From the experimental results, it is confirmed that watermarks could be embedded with high tolerance and high sound quality in music data of RWC. However, there were no tolerances against attacks except for MP3 in music data of SQAM. Hence, tolerance should be improved in music data of SQAM. Furthermore, synchronization detection will be experimented in future works.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP26870681, JP26330214.

References

[1] Y. Li, S. Amari, A. Cichocki, D.W.C. Ho and S. Xie, "Underdetermined blind source separation based on sparse representation," IEEE Trans. Signal Processing, vol.54, no.2, pp.423–437, 2006.
[2] H. Murata, A. Ogihara and S. Hayashi, "An audio watermarking method based on modification of sound pressure level between microphones," Proc. the 28th International Technical Conference on Circuits/Systems, Computers and Communications, pp.669–672, 2013.
[3] H. Murata, A. Ogihara and M. Uesaka, "Sound quality evaluation for audio watermarking based on phase shift keying using BCH Code," IEICE Trans. on Information and Systems, vol.E98-D, no.1, pp.89–94, 2015.
[4] http://www.ieice.org/iss/emm/ihc/audio/audio2013v2.pdf, Accessed Oct. 1, 2013.
[5] http://www.ieice.org/iss/emm/ihc/IHC_criteriaVer5.pdf, Accessed March 30, 2017.
[6] ITU-R Rec. Bs.1387, "Method for objective measurements of perceived audio quality," 2001.