LETTER

Full-Index-Embedding Patchwork Algorithm for Audio Watermarking

Hyunho KANG[†a)], ***Member***, **Koutarou YAMAGUCHI**[††], ***Nonmember***, **Brian KURKOSKI**[††], **Kazuhiko YAMAGUCHI**[††], ***Members***, ***and*** **Kingo KOBAYASHI**[††], ***Fellow***

SUMMARY For the digital watermarking patchwork algorithm originally given by Bender et al., this paper proposes two improvements applicable to audio watermarking. First, the watermark embedding strength is psychoacoustically adapted, using the Bark frequency scale. Second, whereas previous approaches leave the samples that do not correspond to the data untouched, in this paper, these are modified to reduce the probability of misdetection, a method called full index embedding. In simulations, the proposed combination of these two proposed methods has higher resistance to a variety of attacks than prior algorithms.

key words: *audio watermarking, patchwork, DWPT (discrete wavelet packet transform), psychoacoustic model*

1. Introduction

This research considers embedding of digital watermarks in audio data that is likely to proliferate via unauthorized sharing particularly over networks. Such watermarks must have resistance to various attacks so that the embedded copyright information cannot be easily altered or removed by malicious parties. However, in watermarking there is an inherent tradeoff between data reliability and embedding strength. Human hearing is more sensitive than vision, and the watermark embedding strength in audio data has a particularly large influence on the perceived quality.

Here, recent research progress in digital audio watermarking is reviewed. Boney et al. introduced a method to perform audio watermarking using spread spectrum [1]. Gruhl and Bender described a technique that adds echo components to audio [2]. Bender, Gruhl, Morimoto and Lu described a data hiding technique that modified the audio phase [3].

An important advancement in audio watermarking by Bender et al., and the embedding algorithm considered in this paper, is the patchwork algorithm, originally applied to image watermarking. This algorithm shifts the host statistics by adding or subtracting a constant to or from the amplitudes in specific sample positions called patches, or index sets [3]. For the binary case, there are two index sets, one which is modified only when the embedded data bit is a 0, and the other which is modified only when the embedded data bit is a 1. Following this original proposal for the patchwork algorithm, Arnold applied it to music in the time domain [4]. Yeo et al. improved it by weighting the embedding strength by the variance, and applying the algorithm to the Fourier transform of the audio, succeeding in giving it greater resistance against various attacks [5]. They called the technique the modified patchwork algorithm (MPA). However, in order to increase the attack resistance of the MPA, information had to be repeatedly embedded. For these techniques, detection of the watermark under low-rate compression and filtering attacks is difficult, and the amount of data which can be embedded is low.

In this paper, performance of the patchwork algorithm is improved in two ways. First, the embedding strength is adaptively changed frame-by-frame using psychoacoustic models [6], whereas Yeo et al. used the samples' variance. Second, whereas prior approaches to the patchwork algorithm modified only samples in the index set corresponding to the embedded data, in this paper, samples in all index sets are modified to reduce the probability of misdetection, a technique referred to as full index embedding. As a result of these two modifications, the proposed Full-Index-Embedding patchwork algorithm obtains a relatively high embedding rate of 43 bits per second. In addition, this algorithm is applied to coefficients of the discrete wavelet packet transform (DWPT), rather than to the time domain or Fourier transform as was done previously. The DWPT is performed over the same audio block size as the watermark embedding, which reduces complexity [7].

The outline of the remainder of the paper is as follows. Section 2 gives an overview of the patchwork algorithm and the MPA variation. The psychoacoustic model as it will be used in the proposed embedding method is explained in Sect. 3. In Sect. 4, the Full-Index-Embedding patchwork algorithm is explained, both how the psychoacoustic model is applied, and the full index embedding method. Section 5 gives simulation results for various attacks, and Sect. 6 gives the conclusions.

2. Patchwork Algorithm for Watermarking

Principles Initially, we consider using the patchwork algorithm to embed one of q distinct symbols; for example $q = 2$ is the binary case. First, in any given frame, for each of the q information symbols, an index set, or patch, $I_j, j = 1, \ldots, q$ is pseudo-randomly generated from a key or

Manuscript received June 12, 2008.
Manuscript revised July 24, 2008.

[†]The author is with the Center for Research and Development Initiative, Chuo University, Tokyo, 112–8551 Japan.
[††]The authors are with the University of Electro-Communications, Chofu-shi, 182–8585 Japan.
a) E-mail: kang@tamacc.chuo-u.ac.jp

DOI: 10.1093/ietisy/e91–d.11.2731

initial seed. Each index set I_j refers to $2N$ samples, where the first half N samples are used for adding and are designated $a_i, i = 1, \ldots, N$. The second half N samples are used for subtraction and are designated $b_i, i = 1, \ldots, N$. Let δ denote the increment level. Then the modified sequence a'_i, b'_i is:

$$\begin{aligned} a'_i &= a_i + \delta, \\ b'_i &= b_i - \delta. \end{aligned} \tag{1}$$

When embedding is not performed, the expected difference of averages of a_i and b_i, $\mathbb{E}[\bar{a} - \bar{b}]$, is zero. However, when embedding is performed, the expected difference is:

$$\mathbb{E}[\bar{a}' - \bar{b}'] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(a_i + \delta) - \frac{1}{N}\sum_{i=1}^{N}(b_i - \delta)\right] \tag{2}$$

$$= 2\delta. \tag{3}$$

This is the basis for detection in the patchwork algorithm.

Embedding If the data to be embedded is $x \in \{1, 2, \ldots, q\}$, then the embedding step (1) is applied to all $a_i, b_i, i = 1, \ldots, N$ belonging to the index set I_x. The samples in the other index positions are not modified.

Detection The detector finds the difference of the averages $\bar{a}' - \bar{b}'$ for all the index sets, denoted $(\bar{a} - \bar{b})_{I_j}$. Then the estimated embedded information symbol $\widehat{x} \in \{1, 2, \ldots, q\}$ is for the symbol corresponding to the index set with the greatest difference of averages. For the example of the binary case with $x \in \{1, 2\}$, if the detector finds $(\bar{a} - \bar{b})_{I_1} > (\bar{a} - \bar{b})_{I_2}$, then the decision is for $\widehat{x} = 1$.

Modified Patchwork Algorithm When the variance of the samples within an index set is large, an impractically large index set size $2N$ must be used in order for $\bar{a} - \bar{b}$ to approach zero. Furthermore, this large variance makes detection difficult. Yeo, et al. proposed the MPA to reduce the effect of large variances, where the embedding strength is proportional to the "pooled sample standard error," S, within each embedding frame. In addition, the algorithm was applied to the Fourier transform of the data, and data was embedded multiple times to reduce the probability of error.

Then, the embedding function is:

$$a'_i = a_i + \mathrm{sgn}(\bar{a} - \bar{b})\,\sqrt{C}\frac{S}{2} \tag{4}$$

$$b'_i = b_i - \mathrm{sgn}(\bar{a} - \bar{b})\,\sqrt{C}\frac{S}{2} \tag{5}$$

Here, the function sgn is "+1" if the argument is positive, and "−1" otherwise, and C is a constant. When done this way, the difference of the averages is at least $\sqrt{C}S$. Figure 1 shows an unshifted (unwatermarked) and shifted (watermarked) distribution. The detector compares the absolute value of the difference of the averages divided by S, and the largest one is the estimated embedded information.

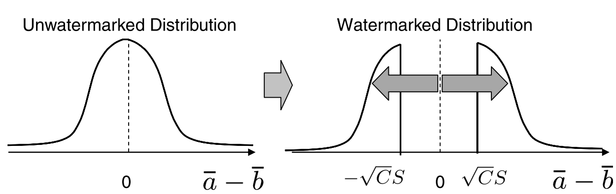

Fig. 1 Left, unshifted (unwatermarked) distribution. Right, shifted distribution due to watermark.

3. Psychoacoustic Model

Human hearing has two properties of interest. One is that the limit on minimally audible energy levels differ depending on frequency. The second is that a sound occurring just before or after a loud sound is particularly difficult to hear. This psychoacoustic model is the basis for state-of-the-art audio compression.

Here, the Bark scale is used, which is adjusted to the characteristics of human hearing, where a Bark with width 1 is called the critical band [6]. In this model, first, the signal energy inside each critical band is calculated. Let z_i be the frequency on the Bark scale, and let $X(z_i)$ be the subband of signal z_i. It is possible to convert from frequency scale to the Bark scale by

$$z(f) = 13\tan^{-1}(0.00076f) + 3.5\tan^{-1}\left(\frac{f}{7500}\right)^2, \tag{6}$$

where f is the frequency in Hertz.

Next, the self-masking level is computed as $v_s(z_i) = -2.025 - 0.175z_i$. The masking level $v_f(z_i, z_j)$ can be found as:

$$v_f(z_i, z_j) = \begin{cases} 17\Delta z - 0.4X(z_i) + 11 & (-3 \le \Delta z < -1) \\ (0.4X(z_i) + 6)\Delta z & (-1 \le \Delta z < 0) \\ -17\Delta z & (0 \le \Delta z < 1) \\ -17\Delta z + 0.15(\Delta z - 1)X(z_i) & (1 \le \Delta z < 8) \end{cases},$$

where $\Delta z = z_i - z_j$.

Using the signal $X(z_j)$, which is a function of the Bark frequency z_i, the masking threshold $M(z_i, z_j)$ is computed as:

$$M(z_i, z_j) = X(z_j) + v_s(z_j) + v_f(z_i, z_j), \tag{7}$$

and finally the general masking threshold $\mathrm{GMT}(z_i)$ is found by:

$$\mathrm{GMT}(z_i) = 10\log_{10}\left(10^{\frac{Tq(z_i)}{10}} + \sum_{j=1}^{28} 10^{\frac{M(z_i,z_j)}{10}}\right). \tag{8}$$

Here, $Tq(z_i)$ is the minimally audible level for Bark frequency z_i.

4. Full-Index-Embedding Patchwork Algorithm

4.1 Psychoacoustically-Adapted Embedding Strength

In the proposed Full-Index-Embedding patchwork algorithm, the embedding strength is scaled by the masking threshold as found in Sect. 3. It is possible to embed inaudibly and strongly by modifying the DWPT coefficients as:

$$a_i' = a_i + \mathrm{sign}(\bar{a} - \bar{b})\mathrm{GMT}(i)$$
$$b_i' = b_i - \mathrm{sign}(\bar{a} - \bar{b})\mathrm{GMT}(i). \quad (9)$$

As with the usual patchwork algorithm detector, the data decision is for the index set with the greatest absolute value of the difference of the averages.

4.2 Full Index Embedding

In the patchwork algorithm, it is assumed that the difference of averages of the samples is close to zero. While this is generally true for large N, using large N increases the number of samples required for embedding one symbol, and thus decreases the total amount of information which can be embedded. On the other hand, decreasing N leads to higher error probability. Thus, there is a trade-off between the total amount of information which can be embedded and the error probability.

Under prior approaches, samples from only one of the q index sets, I_x are modified, and the samples for the other index sets I_j, $j \in \{1, 2, \ldots, q\} \setminus x$ are not modified. However, during detection, if the difference of averages of samples for the unused index set exceeds that for the embedded data's index set, then there will be a detection error. This situation is illustrated in Fig. 2 (a) and (b), where the shaded section in (a) represents the difference of averages, and corresponding probability, where a detection error could occur.

For the proposed system, not only are the samples corresponding to the embedded symbol x modified, as described previously, but the samples for all the other index sets, $j \in \{1, 2 \ldots, q\} \setminus x$ are also modified, as described in the sequel. This is to reduce the probability of false detection, and is illustrated in Fig. 2 (c) and (d). We refer to this as full index embedding.

Define the absolute value of the difference of averages for index I_j as $\Delta_{I_j} = |(\bar{a} - \bar{b})_{I_j}|$. Then the samples in indices which do not correspond to the data symbol are modified as:

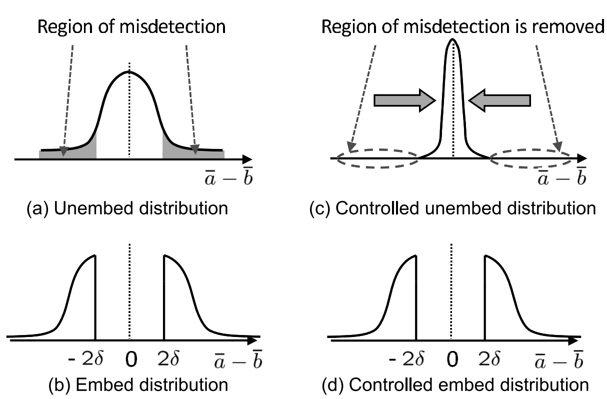

Fig. 2 (a), (b) Prior algorithms' distribution. (c), (d) Proposed algorithm distribution.

$$a_i' = a_i - \mathrm{sign}(\bar{a} - \bar{b})\frac{\Delta_I}{2},$$
$$b_i' = b_i + \mathrm{sign}(\bar{a} - \bar{b})\frac{\Delta_I}{2}. \quad (10)$$

If, for example, $\bar{a} > \bar{b}$, then subtracting the positive Δ_{I_j} from a_i and likewise adding the same constant to b_i, will tend to concentrate the distribution of a_i' and b_i' closer to zero. Note that the samples which do correspond to the data symbol are modified as in Sect. 4.1. Based on this idea, even if the number of embedding locations N is small, the error rate can be held down.

5. Experimental Results and Discussion

In this section, we show that the proposed adaptive patchwork algorithm is effective through simulation of various attacks.

In our system, the audio data is first split into frames, to which the DWPT transform is applied. Then the proposed watermarking algorithm is applied, using the psychoacoustically adapted weights, or full index embedding, or both. Then the inverse DWPT is applied and the frames are concatenated, to obtain the time-domain watermarked sequence. As with prior patchwork algorithm methods, the detector does not require the host, so this is blind watermarking. In the experiments, variations of the proposed algorithm and the MPA were applied to 30-second clips of classical, jazz, pop (female voice) and rock (male voice) music. Further, the index set size is $N = 50$, and the frame size is $L = 1024$ samples. In this section, two-level binary embedding is assumed, $q = 2$, so there are two index sets I_1 and I_2.

The signal-to-noise ratio (SNR) was derived from audio sample from each of the following music genre; pop, classic, jazz and rock. The average SNR of the MPA is 25.5240 dB and our method on the other hand is 24.9555 dB. Even though our results are smaller than that by MPA, our full index embedding reduces misdetection rates, as shown in Table 1. Moreover, we argue that the actual quality of audio is better than the MPA since our method is based on the psychoacoustic model. Subjective quality evaluations of the watermarking method has been done by blind listening tests involving 20 persons that listened to the original and the watermarked audio sequences and were asked to report dissimilarities between the two signals, using the five-grade impairment scale according to ITU-R BS.562 [8] (5: imperceptible, 4: perceptible but not annoying, 3: slightly annoying, 2: annoying, 1: very annoying). The average mean opinion score was 4.675, standard deviation 0.444 as a result of subjective quality evaluation.

5.1 Attack Resistance

Table 1 shows misdetection rates for various attacks and four variations of the patchwork algorithm. Column A indicates misdetection rates using MPA alone, and column B indicates performance of the system that adds full index embedding

Table 1 Misdetection rate (%) (see text for legend).

	A	B	C	D
No Attack	1.36	0.08	0.00	0.00
Echo addition	3.00	1.24	0.25	0.13
Down Sampling	1.81	0.08	0.00	0.00
Requantization	1.92	0.08	0.00	0.00
LPF 12 kHz	3.10	0.08	0.37	0.00
LPF 8 kHz	4.29	2.36	2.23	1.24
White Noise 20 dB	2.23	0.05	1.57	0.00
White Noise 15 dB	4.17	2.44	2.11	0.31
MP3 128 kbps	1.48	0.26	0.00	0.00
MP3 96 kbps	2.25	1.06	0.49	0.20
MP3 64 kbps	5.99	3.42	2.13	0.92
AAC 128 kbps	1.92	0.15	0.00	0.00
AAC 96 kbps	6.62	4.74	3.58	2.04
AAC 64 kbps	15.1	13.7	9.77	8.01

to system A. Column C indicates misdetection rates using psychoacoustically adapted weights only, and column D indicates the performance of the system that adds full index embedding to system C. Further, for the downsampling attack, the sample rate is changed from 44.1 kHz to 22.05 kHz, and the requantization attack changes the amplitude quantization from 16 bits to 8 bits. The echo addition attack adds a signal every 180 ms with 50% amplitude four times, while the white noise attack adds noise to the watermarked signal, with the indicated SNR.

Of all the patchwork algorithms, the proposed algorithm D has the best results. Comparing A and C, it can be seen that using psychoacoustically-adapted embedding embeds more strongly than the variance-based MPA embedding method. Comparing A, B and C, D, we see that full index embedding reduces misdetection rates, and so full index embedding appears effective.

The STEP project, which has the goal of standardization of audio digital watermarking [9], specifies that in 30 seconds, 72 bits of copyright information and in 15 second 2 bits of copy control information, should be able to be embedded and detected. In our experiments, the worst-case ACC 64 kbps had an error rate less than 10%, and 1290 bits can be embedded. So through the use of repetitive embedding, or by using an appropriate error-correcting code, the error rate should become very close to zero.

6. Conclusion

In this paper, two new modifications to the patchwork watermarking algorithm for audio were proposed. One was to adaptively change the embedding strength frame-by-frame using a psychoacoustic model. The second was full index embedding where samples for all index sets, not just the index set corresponding to the embedded information, were modified. It was shown that using psychoacoustically adapted weights gave superior error detection rates to Yeo et al.'s modified patchwork algorithm, and that using full index embedding further improved error rates in both cases. Although embedding of only binary symbols was considered experimentally in this paper, it is reasonable to believe that higher-order alphabets will give additional benefit.

References

[1] L. Boney, A.H. Tewfik, and K.H. Hamdy, "Digital watermarks for audio signals," IEEE Int. Conf. on Multimedia Computing and Systems, pp.473–480, 1996.

[2] D. Gruhl and W. Bender, "Echo hiding," IH'96, Lecture Notes in Computer Science, vol.1174, pp.295–315, Springer-Verlag, 1996.

[3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," IBM Syst. J., vol.35, nos.3&4, pp.313–336, 1996.

[4] M. Arnold, "Audio watermarking: Features, applications and algorithms," IEEE Int. Conf. Multimedia and Expo (II), vol.2, pp.1013–1016, 2000.

[5] I. Yeo and H. Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," IEEE Trans. Speech Audio Process., vol.11, no.4, pp.381–386, 2003.

[6] ISO/IEC 11172-3, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s," 1993.

[7] P. Philippe, F. Moreau de Saint-Martin, and M. Lever, "Wavelet packet filterbanks for low time delay audio coding," IEEE Trans. Speech Audio Process., vol.7, no.3, pp.310–322, 1999.

[8] ITU-R BS.562-3, "Subjective assessment of sound quality," 1990.

[9] Japan Society for the Rights of Authors, Composers and Publishers, "Final selection of technology toward the global spread of digital audio watermarks," http://www.jasrac.or.jp/step2001/presse.htm, 2001.