

Robust Patchwork-Based Watermarking Method for Stereo Audio Signals

Iynkaran Natgunanathan, Yong Xiang*, Yue Rong, *Senior Member, IEEE*, and Dezhong Peng, *Member, IEEE*

Abstract—This paper presents a patchwork-based watermarking method for stereo audio signals, which exploits the similarity of the two sound channels of stereo signals. Given a segment of stereo signal, we first compute the discrete Fourier transforms (DFTs) of the two sound channels, which yields two sets of DFT coefficients. The DFT coefficients corresponding to certain frequency range are divided into multiple subsegment pairs and a criterion is proposed to select those suitable for watermark embedding. Then a watermark is embedded into the selected subsegment pairs by modifying their DFT coefficients. The exact way of modification is determined by a secret key, the watermark to be embedded, and the DFT coefficients themselves. In the decoding process, the subsegment pairs containing watermarks are identified by another criterion. Then the secret key is used to extract the watermark from the watermarked subsegments. Compared to the existing patchwork methods for audio watermarking, the proposed method does not require knowledge of which segments of the watermarked audio signal contain watermarks and is more robust to conventional attacks.

Index Terms—Audio watermarking, patchwork, stereo audio signal, discrete Fourier transform.

I. INTRODUCTION

The past decade has seen an unprecedented surge in the production and distribution of digital media, facilitated by the significant advances in communication networks, computers and multimedia technology. This inevitably leads to strong demand for copyright protection. Traditionally, copyright information (such as publisher’s name, signature, logo, ID number, etc.) is embedded into the header of the media files. However, the copyright data hidden in the header can be easily changed or removed by using commercial audio processing softwares. Due to copyright infringement, the multimedia publishing industry loses many millions of dollars every year. Digital watermarking is an important technology to deal with this problem [1]-[11], aiming to hide watermark data (e.g., copyright information) into the actual media object without affecting its normal usage. While digital watermarking can be applied to various media data such as audio, image and video, we limit our attention to audio watermarking in this paper.

* Corresponding author.

This work was supported in part by the Australian Research Council under grants DP1095498 and DP110102076.

I. Natgunanathan and Y. Xiang are with the School of Information Technology, Deakin University, Burwood Campus, Melbourne, VIC 3125, Australia (Tel: +61 3 92517740, fax: +61 3 92446831, e-mail: inat@deakin.edu.au, yxiang@deakin.edu.au).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (Tel: +61 8 92667398, fax: +61 8 92662584, e-mail: y.rong@curtin.edu.au).

D. Peng is with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, China (Tel: +86 28 85400618, fax: +86 28 85400618, e-mail: pengdz@scu.edu.cn).

The effectiveness of an audio watermarking method is mainly assessed from three aspects: imperceptibility, robustness and security. Imperceptibility refers to that a normal listener cannot distinguish the difference between the host audio signal and the watermarked signal. Robustness indicates the ability of preventing the embedded watermarks from being removed or altered by various attacks such as noise addition, compression, and re-sampling. Security means that an unauthorized user cannot extract the watermark data from the watermarked signal without using a secret key. In addition to these aspects, the computation complexity and watermark embedding rate should also be considered [12], [13]. Furthermore, with respect to decoding, blind methods that can extract watermarking data without resort to host audio signal is desirable as semi-blind and non-blind methods are not applicable to most practical applications [1]. Over the past decade, many audio watermarking methods have been developed by using different techniques such as spread-spectrum [14]-[16], support vector regression [17]-[19], echo-hiding [5]-[7], [20]-[21], and patchwork [1], [8], [22]. The watermarking methods based on patchwork technique are very promising due to their remarkable robustness against conventional attacks. They also have good imperceptibility and high level of security.

Patchwork technique was originally proposed by Bender *et al.* for image watermarking [23] and then Arnold applied this technique to audio watermarking [22]. After that, the modified patchwork algorithm (MPA) was proposed by Yeo and Kim to improve watermarking performance [8]. In MPA, the digital cosine transform coefficients obtained from one audio segment are used to form four patches. Two of these patches are used for embedding watermark bit “1” and the other two patches are utilized for embedding watermark bit “0”. The MPA requires that the selected patches have the same statistical characteristic. This requirement cannot be guaranteed in practice as each patch only has a limited number of samples and increasing the length of the patches will result in low watermark embedding rate [1]. Kalantari *et al.* proposed a multiplicative patchwork method in [1] to deal with this problem. In [1], two patches are constructed by using the wavelet transform coefficients of one host audio segment. A host audio segment is chosen for watermark embedding only if the two patches associated with it have comparable statistical characteristics. Based on this segment selection criterion, a substantial percentage of audio segments are not used to embed watermarks. Since watermarks are embedded in selected host audio segments, in the decoding process one needs to know which segments of the watermarked signal contain watermarks. Without this information, a large number of false watermarks will be “extracted” from the un-

watermarked audio segments. However, [1] does not provide an answer to this question. While an approach was proposed in [24] to estimate the indices of selected image frames, it cannot be directly applied nor simply modified to identify the watermarked segments encountered in [1].

Although nowadays most audio signals are stereo signals, the above-mentioned watermarking methods are developed only for mono audio signals. The watermarking methods that are specifically designed for stereo audio signals are scarce [25]-[28]. In [25], three watermarking schemes are proposed for stereo signals but two of them are non-blind and none of them is secure since they do not use secret key in the embedding and decoding processes. Similarly, the method in [26] is non-blind and the method in [27] is not secure. In [28], Cao *et al.* utilize a bit replacement technique to hide watermarks but this method is not robust to some conventional attacks such as noise addition attack and compression attack.

In this paper, we propose a patchwork-based method for stereo audio watermarking. The proposed watermarking method makes use of the similarity of the left and right sound channels of a stereo audio signal and is implemented in frequency domain. First, the host stereo audio signal is segmented. For a given segment of the host signal, we apply the discrete Fourier transform (DFT) to its sound channels to obtain two sets of DFT coefficients. After discarding those DFT coefficients associated with frequencies that are vulnerable to compression-type attacks or are not audible, the remaining DFT coefficients are divided into multiple subsegment pairs. The two subsegments in each subsegment pair correspond to the left and right sound channels, respectively. A criterion is used to select those subsegment pairs suitable for embedding watermarks. This step is essential to ensuring that the watermarked signal is of high perceptual quality. For each subsegment in a subsegment pair, the corresponding DFT coefficients are classified into two groups based on a pseudonoise (PN) sequence, which serves as a secret key in the decoding process. Thus we can get two pairs of groups from one subsegment pair. Then a watermark is embedded into the subsegment pair by modifying the DFT coefficients in the two pairs of groups. The same watermark is also embedded into other subsegment pairs to enhance robustness. A special feature of our watermark embedding approach is that the mean of moduli of a watermarked subsegment is the same as that of its host counterpart. This feature can be employed to identify the watermarked subsegment pairs at the decoding side. After that, the watermark can be easily extracted from each watermarked audio segment by using the secret key, without resort to the host audio signal.

The proposed method is superior to the existing patchwork audio watermarking methods as it does not need to know whether a segment from the watermarked signal contains a watermark or not and is more robust to conventional attacks. Experimental results show the effectiveness of our method, in comparison with the methods in [1] and [27]. The remainder of this paper is organized as follows. The new method is presented in Section II and its robustness against conventional attacks is analyzed in Section III. The experimental results are shown in Section IV. Finally, Section V concludes the paper.

II. PROPOSED METHOD

In this section, we present the new patchwork-based stereo audio watermarking method. It utilizes the multiplicative patchwork concept but explicitly exploits the similarity existed in the stereo sound channels to achieve desired performance.

A. Watermark Embedding

1) *Segmentation of host audio signal*: The segmentation of the host audio signal is shown in the upper part of Fig. 1. The host stereo audio signal is first divided into segments of equal length, where the segment length is chosen empirically. For each chosen segment, a digital watermark bit, which is either “1” or “0”, will be inserted into it. Clearly, a stereo audio segment includes two channel segments and we denote the left and right channel segments by $x_L(n)$ and $x_R(n)$, respectively. Let $X_L(k)$ and $X_R(k)$ be the DFTs of $x_L(n)$ and $x_R(n)$, respectively. Since human auditory system is insensitive to signals that are of high frequencies or very low frequencies, watermarks embedded in high frequency region or very low frequency region can be easily removed by some intentional or unintentional attacks such as compression attack. For this reason, we use a low to middle frequency region, say (f_{min}, f_{max}) , to embed watermarks. We denote the parts of $X_L(k)$ and $X_R(k)$ related to the frequency region (f_{min}, f_{max}) by $\mathcal{X}_L(k)$ and $\mathcal{X}_R(k)$, respectively.

An effective way of further enhancing robustness is to insert one watermark bit into a stereo audio segment multiple times. To implement this, we break up $\mathcal{X}_L(k)$ (resp. $\mathcal{X}_R(k)$) into M subsegments of length N , where N is an even number, and denote the m th subsegment by $\mathcal{X}_{L,m}(k)$ (resp. $\mathcal{X}_{R,m}(k)$). Assume that the length of $\mathcal{X}_L(k)$ and $\mathcal{X}_R(k)$ is $K = MN$ and define

$$\mathcal{X}_L(k) \triangleq \{a_1, a_2, \dots, a_K\}, \quad \mathcal{X}_R(k) \triangleq \{b_1, b_2, \dots, b_K\}. \quad (1)$$

From (1), it follows

$$\begin{aligned} \mathcal{X}_{L,m}(k) &\triangleq \{a_{m,1}, a_{m,2}, \dots, a_{m,N}\} \\ &= \{a_{(m-1)N+1}, a_{(m-1)N+2}, \dots, a_{mN}\} \end{aligned} \quad (2)$$

$$\begin{aligned} \mathcal{X}_{R,m}(k) &\triangleq \{b_{m,1}, b_{m,2}, \dots, b_{m,N}\} \\ &= \{b_{(m-1)N+1}, b_{(m-1)N+2}, \dots, b_{mN}\} \end{aligned} \quad (3)$$

where $m = 1, 2, \dots, M$. Here $\mathcal{X}_{L,m}(k)$ and $\mathcal{X}_{R,m}(k)$ form the m th subsegment pair, and M subsegment pairs can be constructed from one stereo audio segment.

Let $\bar{\mathcal{X}}_{L,m}$ and $\bar{\mathcal{X}}_{R,m}$ be the means of $|\mathcal{X}_{L,m}(k)|$ and $|\mathcal{X}_{R,m}(k)|$, respectively, where $|a|$ stands for the modulus of complex number a . It results from (2) and (3) that

$$\bar{\mathcal{X}}_{L,m} = \frac{1}{N} \sum_{l=1}^N |a_{m,l}|, \quad \bar{\mathcal{X}}_{R,m} = \frac{1}{N} \sum_{l=1}^N |b_{m,l}|.$$

To avoid audible distortions in the watermarked signal, watermark bits should not be implanted into the subsegment pairs that have little contents. A subsegment pair is selected to embed a watermark only if

$$\min \{\bar{\mathcal{X}}_{L,m}, \bar{\mathcal{X}}_{R,m}\} \geq \sigma \quad (4)$$

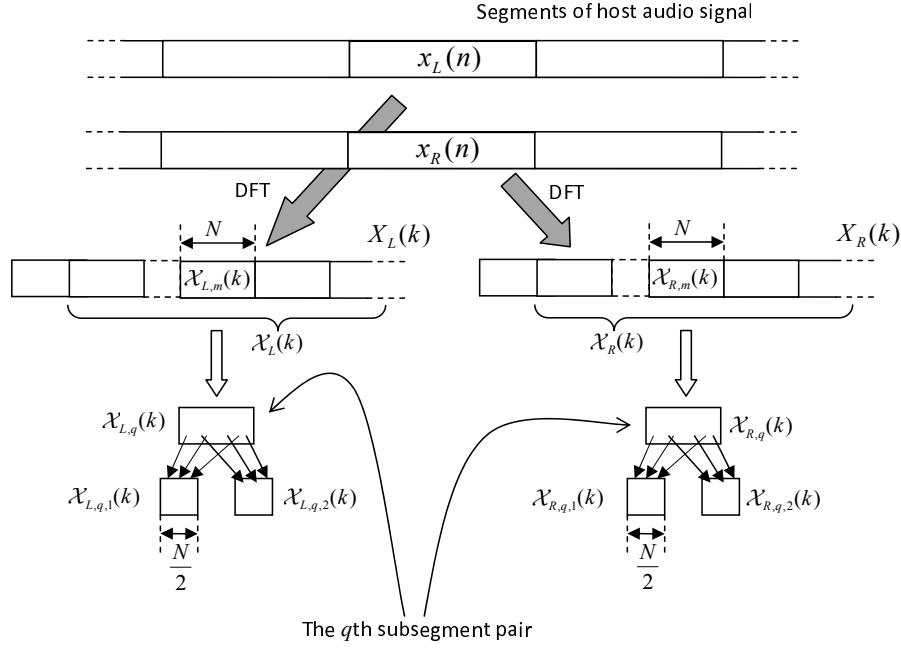


Fig. 1. Illustration of host audio signal segmentation and DFT coefficients classification.

where $1 \leq m \leq M$ and σ is a small positive threshold which can be chosen empirically. If all the M subsegment pairs in a stereo audio segment do not satisfy the selection criterion (4), this stereo audio segment will not be used to embed watermark. We assume without loss of generality that out of the M subsegment pairs, the first Q subsegment pairs satisfy condition (4) and are selected for watermark embedding, where $Q \leq M$.

2) *Classification of DFT coefficients*: Security is a key aspect that must be considered in the development of watermarking methods. In order to introduce security into our watermarking method, we use a PN sequence to classify the DFT coefficients in a selected subsegment pair, say the q th subsegment pair, into two pairs of groups. Let

$$p(n) = \{p_1, p_2, \dots, p_N\} \quad (5)$$

be a randomly generated PN sequence of length N , where $p_i \in \{1, 2, \dots, N\}$, and $p_i \neq p_j$ if $i \neq j$. For example, if $N = 30$, a possible PN sequence could be $p(n) = \{14, 1, 30, \dots, 11, 7, 21\}$.

Then the first $N/2$ elements of $p(n)$ are employed to find the first pair of groups in a subsegment pair. Based on the symbols given in (2), (3) and (5), the first pair of groups corresponding to the q th subsegment pair consisting of $X_{L,q}(k)$ and $X_{R,q}(k)$ can be obtained by

$$\begin{aligned} X_{L,q,1}(k) &= \{a_{q,p_1}, a_{q,p_2}, \dots, a_{q,p_{N/2}}\} \\ X_{R,q,1}(k) &= \{b_{q,p_1}, b_{q,p_2}, \dots, b_{q,p_{N/2}}\} \end{aligned}$$

where $q = 1, 2, \dots, Q$. Similarly, the second pair of groups associated with the q th subsegment pair can be constructed by

using the last $N/2$ elements of $p(n)$ as

$$\begin{aligned} X_{L,q,2}(k) &= \{a_{q,p_{N/2+1}}, a_{q,p_{N/2+2}}, \dots, a_{q,p_N}\} \\ X_{R,q,2}(k) &= \{b_{q,p_{N/2+1}}, b_{q,p_{N/2+2}}, \dots, b_{q,p_N}\} \end{aligned}$$

where $q = 1, 2, \dots, Q$. The classification of DFT coefficients is illustrated in the lower part of Fig. 1.

Next we shall show how to insert a digital watermark into the segment pair by modifying the DFT coefficients in $X_{L,q,i}(k)$ and $X_{R,q,i}(k)$, where $i = 1, 2$ and $q = 1, 2, \dots, Q$.

3) *Insertion of watermark*: For a given q , let $\alpha_{L,q,1}$, $\alpha_{R,q,1}$, $\alpha_{L,q,2}$ and $\alpha_{R,q,2}$ be four positive real constants. Let $\mathcal{Y}_{L,q,1}(k)$, $\mathcal{Y}_{R,q,1}(k)$, $\mathcal{Y}_{L,q,2}(k)$ and $\mathcal{Y}_{R,q,2}(k)$ be the modified counterparts of $X_{L,q,1}(k)$, $X_{R,q,1}(k)$, $X_{L,q,2}(k)$ and $X_{R,q,2}(k)$, respectively. We insert a digital watermark into the q th subsegment pair by modifying the DFT coefficients in the two pairs of groups as follows:

$$\begin{cases} \mathcal{Y}_{L,q,1}(k) = \alpha_{L,q,1} \times X_{L,q,1}(k) \\ \mathcal{Y}_{R,q,1}(k) = \alpha_{R,q,1} \times X_{R,q,1}(k) \end{cases} \quad (6)$$

and

$$\begin{cases} \mathcal{Y}_{L,q,2}(k) = \alpha_{L,q,2} \times X_{L,q,2}(k) \\ \mathcal{Y}_{R,q,2}(k) = \alpha_{R,q,2} \times X_{R,q,2}(k) \end{cases} \quad (7)$$

Here $\alpha_{L,q,i}$ and $\alpha_{R,q,i}$, $i = 1, 2$ take values from the range $[\alpha_{min}, \alpha_{max}]$, where

$$\alpha_{max} = 1 + \gamma \quad (8)$$

$$\alpha_{min} = 1 - \gamma \quad (9)$$

with $\gamma \in (0, 1)$. To embed watermark bit "0", we set

$$\begin{cases} \alpha_{L,q,1} > 1 \\ \alpha_{R,q,1} < 1 \end{cases} \quad \text{and} \quad \begin{cases} \alpha_{L,q,2} < 1 \\ \alpha_{R,q,2} > 1 \end{cases} \quad (10)$$

Similarly, to embed watermark bit “1”, we set

$$\begin{cases} \alpha_{L,q,1} < 1 \\ \alpha_{R,q,1} > 1 \end{cases} \quad \text{and} \quad \begin{cases} \alpha_{L,q,2} > 1 \\ \alpha_{R,q,2} < 1 \end{cases}. \quad (11)$$

Clearly, $\alpha_{L,q,i}$ and $\alpha_{R,q,i}$, $i = 1, 2$ should be close to 1 to ensure that the watermarked signal has good perceptual quality. On the other hand, if they are too close to 1, the watermarked signal will be vulnerable to attacks. To consider both requirements, a typical value for γ can be chosen as $\gamma = 0.05$, which is small enough to guarantee imperceptibility. After the range $[\alpha_{min}, \alpha_{max}]$ is determined, the values of $\alpha_{L,q,1}$, $\alpha_{R,q,1}$, $\alpha_{L,q,2}$ and $\alpha_{R,q,2}$ should be as distant from 1 as possible (or as close to α_{min} and α_{max} as possible) to ensure satisfactory robustness.

Furthermore, we require that the modifications in (6) and (7) do not change the means of the moduli of every subsegment pair. Let $\mathcal{Y}_{L,q}(k)$ and $\mathcal{Y}_{R,q}(k)$ be the modified counterparts of $\mathcal{X}_{L,q}(k)$ and $\mathcal{X}_{R,q}(k)$ respectively, and $\bar{\mathcal{Y}}_{L,q}$ and $\bar{\mathcal{Y}}_{R,q}$ be the means of $|\mathcal{Y}_{L,q}(k)|$ and $|\mathcal{Y}_{R,q}(k)|$ respectively. We also denote the means of $|\mathcal{X}_{L,q,i}(k)|$ and $|\mathcal{X}_{R,q,i}(k)|$ by $\bar{\mathcal{X}}_{L,q,i}$ and $\bar{\mathcal{X}}_{R,q,i}$ respectively, and the means of $|\mathcal{Y}_{L,q,i}(k)|$ and $|\mathcal{Y}_{R,q,i}(k)|$ by $\bar{\mathcal{Y}}_{L,q,i}$ and $\bar{\mathcal{Y}}_{R,q,i}$ respectively, where $i = 1, 2$. This requirement means

$$\bar{\mathcal{Y}}_{L,q} = \bar{\mathcal{X}}_{L,q} \quad \text{or} \quad \frac{\bar{\mathcal{Y}}_{L,q,1} + \bar{\mathcal{Y}}_{L,q,2}}{2} = \frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{2} \quad (12)$$

and

$$\bar{\mathcal{Y}}_{R,q} = \bar{\mathcal{X}}_{R,q} \quad \text{or} \quad \frac{\bar{\mathcal{Y}}_{R,q,1} + \bar{\mathcal{Y}}_{R,q,2}}{2} = \frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{2}. \quad (13)$$

Ensuring (12) and (13) in watermark embedding is important as they can be exploited to identify the watermarked subsegment pairs at the decoding end. It is easy to verify that (12) and (13) hold if the following eight relations are satisfied:

- In the case of embedding watermark bit “0”

$$\begin{aligned} \bar{\mathcal{Y}}_{L,q,1} &= \bar{\mathcal{X}}_{L,q,1} + \gamma \cdot \min \{ \bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2} \} \\ \bar{\mathcal{Y}}_{L,q,2} &= \bar{\mathcal{X}}_{L,q,2} - \gamma \cdot \min \{ \bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2} \} \\ \bar{\mathcal{Y}}_{R,q,1} &= \bar{\mathcal{X}}_{R,q,1} - \gamma \cdot \min \{ \bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2} \} \\ \bar{\mathcal{Y}}_{R,q,2} &= \bar{\mathcal{X}}_{R,q,2} + \gamma \cdot \min \{ \bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2} \} \end{aligned} \quad (14)$$

- In the case of embedding watermark bit “1”

$$\begin{aligned} \bar{\mathcal{Y}}_{L,q,1} &= \bar{\mathcal{X}}_{L,q,1} - \gamma \cdot \min \{ \bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2} \} \\ \bar{\mathcal{Y}}_{L,q,2} &= \bar{\mathcal{X}}_{L,q,2} + \gamma \cdot \min \{ \bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2} \} \\ \bar{\mathcal{Y}}_{R,q,1} &= \bar{\mathcal{X}}_{R,q,1} + \gamma \cdot \min \{ \bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2} \} \\ \bar{\mathcal{Y}}_{R,q,2} &= \bar{\mathcal{X}}_{R,q,2} - \gamma \cdot \min \{ \bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2} \}. \end{aligned} \quad (15)$$

To satisfy the above relations in watermark embedding, we first define

$$f(\theta_1, \theta_2, \theta_3) = - \left(\frac{\theta_1}{\theta_2} \right) \theta_3 + \frac{\theta_1 + \theta_2}{\theta_2}. \quad (16)$$

Then we propose to insert watermark bit “0” into the subsegment pair by using the following α values in (6) and (7):

$$\alpha_{L,q,1}, \alpha_{L,q,2} : \begin{cases} \text{if } \bar{\mathcal{X}}_{L,q,1} < \bar{\mathcal{X}}_{L,q,2} \\ \alpha_{L,q,1} = \alpha_{max} = 1 + \gamma \\ \alpha_{L,q,2} = f(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1 + \gamma)) \\ \text{Otherwise} \\ \alpha_{L,q,1} = f(\bar{\mathcal{X}}_{L,q,2}, \bar{\mathcal{X}}_{L,q,1}, (1 - \gamma)) \\ \alpha_{L,q,2} = \alpha_{min} = 1 - \gamma \end{cases} \quad (17)$$

and

$$\alpha_{R,q,1}, \alpha_{R,q,2} : \begin{cases} \text{if } \bar{\mathcal{X}}_{R,q,1} < \bar{\mathcal{X}}_{R,q,2} \\ \alpha_{R,q,1} = \alpha_{min} = 1 - \gamma \\ \alpha_{R,q,2} = f(\bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2}, (1 - \gamma)) \\ \text{Otherwise} \\ \alpha_{R,q,1} = f(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1 + \gamma)) \\ \alpha_{R,q,2} = \alpha_{max} = 1 + \gamma \end{cases}. \quad (18)$$

If the watermark bit to be inserted is “1”, the α values below are employed:

$$\alpha_{L,q,1}, \alpha_{L,q,2} : \begin{cases} \text{if } \bar{\mathcal{X}}_{L,q,1} < \bar{\mathcal{X}}_{L,q,2} \\ \alpha_{L,q,1} = \alpha_{min} = 1 - \gamma \\ \alpha_{L,q,2} = f(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1 - \gamma)) \\ \text{Otherwise} \\ \alpha_{L,q,1} = f(\bar{\mathcal{X}}_{L,q,2}, \bar{\mathcal{X}}_{L,q,1}, (1 + \gamma)) \\ \alpha_{L,q,2} = \alpha_{max} = 1 + \gamma \end{cases} \quad (19)$$

and

$$\alpha_{R,q,1}, \alpha_{R,q,2} : \begin{cases} \text{if } \bar{\mathcal{X}}_{R,q,1} < \bar{\mathcal{X}}_{R,q,2} \\ \alpha_{R,q,1} = \alpha_{max} = 1 + \gamma \\ \alpha_{R,q,2} = f(\bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2}, (1 + \gamma)) \\ \text{Otherwise} \\ \alpha_{R,q,1} = f(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1 - \gamma)) \\ \alpha_{R,q,2} = \alpha_{min} = 1 - \gamma \end{cases}. \quad (20)$$

Depending on the watermark to be embedded (“0” or “1”), the relationship between $\bar{\mathcal{X}}_{L,q,1}$ and $\bar{\mathcal{X}}_{L,q,2}$, and the relationship between $\bar{\mathcal{X}}_{R,q,1}$ and $\bar{\mathcal{X}}_{R,q,2}$, it can be seen from (17)-(20) that there exist eight sets of α values. Since these α values satisfy the relations in (14) and (15), they also satisfy (12) and (13). For example, two of the eight sets of α values are as follows.

Case 1: Assume $\bar{\mathcal{X}}_{L,q,1} < \bar{\mathcal{X}}_{L,q,2}$, $\bar{\mathcal{X}}_{R,q,1} \geq \bar{\mathcal{X}}_{R,q,2}$, and the watermark to be embedded is “0”. It follows from (17) and (18) that $\alpha_{L,q,1} = 1 + \gamma$, $\alpha_{L,q,2} = f(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1 + \gamma))$, $\alpha_{R,q,1} = f(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1 + \gamma))$ and $\alpha_{R,q,2} = 1 + \gamma$. Considering (16), it is easy to verify that this set of α values satisfy (10), (12) and (13).

Case 2: Assume $\bar{\mathcal{X}}_{L,q,1} < \bar{\mathcal{X}}_{L,q,2}$, $\bar{\mathcal{X}}_{R,q,1} \geq \bar{\mathcal{X}}_{R,q,2}$, and the watermark to be embedded is “1”. It can be found from (19) and (20) that the set of α values are $\alpha_{L,q,1} = 1 - \gamma$, $\alpha_{L,q,2} = f(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1 - \gamma))$, $\alpha_{R,q,1} = f(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1 - \gamma))$ and $\alpha_{R,q,2} = 1 - \gamma$, which satisfy (11)-(13).

Based on the watermark embedding scheme in (6) and (7), one can insert the same watermark into all the Q subsegment pairs of a selected segment pair, i.e., one watermark bit is embedded into the selected segment pair Q times. Then the watermarked stereo audio segment can be obtained by applying the inverse discrete Fourier transform (IDFT) to the modified segment pair.

B. Watermark Decoding

This subsection presents a decoding scheme to extract watermarks from the watermarked stereo audio signal by utilizing the PN sequence $p(n)$ as a secret key. It is a blind decoding scheme as it does not rely on the host audio signal.

1) *Identification of watermarked subsegment pairs*: Similar to segmenting the host audio signal in the watermark embedding process, one can segment the watermarked stereo audio signal in the same manner to form the corresponding DFT-domain segment pairs and each of the segment pairs consists of M subsegment pairs. Given a segment pair, the M subsegment pairs are labelled as $\mathcal{Y}_{L,m}(k)$ and $\mathcal{Y}_{R,m}(k)$, $m = 1, 2, \dots, M$. Here, $\mathcal{Y}_{L,m}(k)$ and $\mathcal{Y}_{R,m}(k)$ are respectively the counterparts of $\mathcal{X}_{L,m}(k)$ and $\mathcal{X}_{R,m}(k)$ attained from the host audio signal.

Recall that in the watermark embedding process, if the selection criterion (4) does not hold, the subsegment pair $\mathcal{X}_{L,m}(k)$ and $\mathcal{X}_{R,m}(k)$ will not be used to embed watermark. Thus the watermarked stereo audio signal would have both watermarked and un-watermarked subsegment pairs in the DFT domain. Therefore, prior to watermark extraction from the subsegment pair $\mathcal{Y}_{L,m}(k)$ and $\mathcal{Y}_{R,m}(k)$, it is essential to find whether this subsegment pair contains a watermark bit or not.

As we mentioned in the subsection II-A, the proposed watermark embedding scheme ensures (12) and (13), i.e.,

$$\bar{\mathcal{Y}}_{L,m} = \bar{\mathcal{X}}_{L,m} \quad \text{and} \quad \bar{\mathcal{Y}}_{R,m} = \bar{\mathcal{X}}_{R,m}.$$

Based on this property, a criterion similar to (4) can be proposed to examine whether a subsegment pair is watermarked or not. Specifically, if

$$\min \{ \bar{\mathcal{Y}}_{L,m}, \bar{\mathcal{Y}}_{R,m} \} \geq \sigma \quad (21)$$

the subsegment pair $\mathcal{Y}_{L,m}(k)$ and $\mathcal{Y}_{R,m}(k)$ contains a watermark. In the absence of attacks, all the Q watermarked subsegment pairs in the given segment pair can be identified using Eq. (21). Then one can extract the embedded watermark by using the watermark extraction approach to be presented next. If none of the M subsegment pairs satisfies (21), the concerned segment pair does not contain a watermark and watermark extraction should not be conducted.

2) *Extraction of watermark*: Assume without loss of generality that the Q watermarked subsegment pairs are $\mathcal{Y}_{L,q}(k)$ and $\mathcal{Y}_{R,q}(k)$, where $q = 1, 2, \dots, Q$. Similar to classifying the DFT coefficients of $\mathcal{X}_{L,q}(k)$ and $\mathcal{X}_{R,q}(k)$ in the watermark embedding process, we can use the PN sequence $p(n)$ to classify the DFT coefficients of $\mathcal{Y}_{L,q}(k)$ and $\mathcal{Y}_{R,q}(k)$ in the same way to obtain two pairs of groups: $\mathcal{Y}_{L,q,1}(k)$ and $\mathcal{Y}_{R,q,1}(k)$ in one pair and $\mathcal{Y}_{L,q,2}(k)$ and $\mathcal{Y}_{R,q,2}(k)$ in the other pair. The means of their moduli are labelled as $\bar{\mathcal{Y}}_{L,q,1}$, $\bar{\mathcal{Y}}_{R,q,1}$, $\bar{\mathcal{Y}}_{L,q,2}$ and $\bar{\mathcal{Y}}_{R,q,2}$, respectively.

To extract the embedded watermark from the q th subsegment pair $\mathcal{Y}_{L,q}(k)$ and $\mathcal{Y}_{R,q}(k)$, we define

$$\bar{\mathcal{Y}}'_{R,q,1} = \bar{\mathcal{Y}}_{R,q,1} + (\bar{\mathcal{Y}}_{L,q} - \bar{\mathcal{Y}}_{R,q}) \quad (22)$$

$$\bar{\mathcal{Y}}'_{R,q,2} = \bar{\mathcal{Y}}_{R,q,2} + (\bar{\mathcal{Y}}_{L,q} - \bar{\mathcal{Y}}_{R,q}) \quad (23)$$

where $\bar{\mathcal{Y}}_{L,q}$ and $\bar{\mathcal{Y}}_{R,q}$ are the means of $|\mathcal{Y}_{L,q}(k)|$ and $|\mathcal{Y}_{R,q}(k)|$, respectively. Then the embedded watermark can be extracted using the following criterion:

- If $\bar{\mathcal{Y}}_{L,q,1} > \bar{\mathcal{Y}}'_{R,q,1}$ and $\bar{\mathcal{Y}}_{L,q,2} < \bar{\mathcal{Y}}'_{R,q,2}$, the watermark bit embedded in the q th subsegment pair is “0”.
- Otherwise, the watermark bit embedded in the q th subsegment pair is “1”.

Next we use two examples to illustrate the above watermark extraction criterion.

As we have mentioned in the watermark embedding process (subsection II-A), eight set of α values could be used for watermark embedding, depending on the watermark to be embedded and the values of $\bar{\mathcal{X}}_{L,q,1}$, $\bar{\mathcal{X}}_{L,q,2}$, $\bar{\mathcal{X}}_{R,q,1}$ and $\bar{\mathcal{X}}_{R,q,2}$. In the first example, we consider the watermark embedding Case 1. One can see that in Case 1, the α values used to embed watermark bit “0” are

$$\begin{aligned} \alpha_{L,q,1} &= 1 + \gamma \\ \alpha_{L,q,2} &= - \left(\frac{\bar{\mathcal{X}}_{L,q,1}}{\bar{\mathcal{X}}_{L,q,2}} \right) (1 + \gamma) + \frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{\bar{\mathcal{X}}_{L,q,2}} \\ \alpha_{R,q,1} &= - \left(\frac{\bar{\mathcal{X}}_{R,q,2}}{\bar{\mathcal{X}}_{R,q,1}} \right) (1 + \gamma) + \frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{\bar{\mathcal{X}}_{R,q,1}} \\ \alpha_{R,q,2} &= 1 + \gamma \end{aligned}$$

where $\alpha_{L,q,2}$ and $\alpha_{R,q,1}$ result from Eq. (16). Based on this set of α values, it follows from (6) and (7) that

$$\begin{aligned} \bar{\mathcal{Y}}_{L,q,1} &= (1 + \gamma)\bar{\mathcal{X}}_{L,q,1} \\ \bar{\mathcal{Y}}_{L,q,2} &= -\bar{\mathcal{X}}_{L,q,1}(1 + \gamma) + \bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2} \\ \bar{\mathcal{Y}}_{R,q,1} &= -\bar{\mathcal{X}}_{R,q,2}(1 + \gamma) + \bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2} \\ \bar{\mathcal{Y}}_{R,q,2} &= (1 + \gamma)\bar{\mathcal{X}}_{R,q,2}. \end{aligned}$$

Then, from (22) and (23), it yields

$$\begin{aligned} \bar{\mathcal{Y}}'_{R,q,1} &= -\bar{\mathcal{X}}_{R,q,2}(1 + \gamma) + \bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2} \\ &\quad + \left(\frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{2} \right) - \left(\frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{2} \right) \\ \bar{\mathcal{Y}}'_{R,q,2} &= (1 + \gamma)\bar{\mathcal{X}}_{R,q,2} \\ &\quad + \left(\frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{2} \right) - \left(\frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{2} \right). \end{aligned}$$

From the expressions of $\bar{\mathcal{Y}}_{L,q,i}$, $\bar{\mathcal{Y}}_{R,q,i}$ and $\bar{\mathcal{Y}}'_{R,q,i}$ ($i = 1, 2$), one can easily obtain

$$\begin{aligned} \bar{\mathcal{Y}}_{L,q,1} - \bar{\mathcal{Y}}'_{R,q,1} &= (1 + \gamma)\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}(1 + \gamma) - \bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2} \\ &\quad - \frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{2} + \frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{2} \\ &= \left(\frac{\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}}{2} - \frac{\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}}{2} \right) \\ &\quad + \gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}). \end{aligned} \quad (24)$$

Since the elements of $\mathcal{X}_{L,q,1}(k)$ and $\mathcal{X}_{L,q,2}(k)$ are randomly selected from $\mathcal{X}_{L,q}(k)$, the absolute value $|\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}|$ is generally small. Similarly, $|\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}|$ is also small. Furthermore, since the left and right channels of a stereo audio signal have similar characteristics, the values of $(\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2})$ and $(\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2})$ are usually very close. Due to these two reasons, $\left(\frac{\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}}{2} - \frac{\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}}{2} \right) \approx 0$. From (24), this implies $\bar{\mathcal{Y}}_{L,q,1} - \bar{\mathcal{Y}}'_{R,q,1} \approx \gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}) > 0$.

Similarly, we have

$$\begin{aligned}
& \bar{\mathcal{Y}}_{L,q,2} - \bar{\mathcal{Y}}'_{R,q,2} \\
&= -\bar{\mathcal{X}}_{L,q,1}(1+\gamma) + \bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2} - (1+\gamma)\bar{\mathcal{X}}_{R,q,2} \\
&\quad - \frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{2} + \frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{2} \\
&= \left(\frac{\bar{\mathcal{X}}_{L,q,2} - \bar{\mathcal{X}}_{L,q,1}}{2} - \frac{\bar{\mathcal{X}}_{R,q,2} - \bar{\mathcal{X}}_{R,q,1}}{2} \right) \\
&\quad - \gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}) \\
&\approx -\gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}) \\
&< 0.
\end{aligned} \tag{25}$$

Since $\bar{\mathcal{Y}}_{L,q,1} > \bar{\mathcal{Y}}'_{R,q,1}$ and $\bar{\mathcal{Y}}_{L,q,2} < \bar{\mathcal{Y}}'_{R,q,2}$ for this case, according to the proposed watermark extraction criterion, the watermark bit “0” is extracted.

In the second example, we consider the watermark embedding Case 2, where the embedded watermark bit is “1”. Following the procedure used in the first example, we can obtain

$$\begin{aligned}
& \bar{\mathcal{Y}}_{L,q,1} - \bar{\mathcal{Y}}'_{R,q,1} \\
&= \left(\frac{\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}}{2} - \frac{\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}}{2} \right) \\
&\quad - \gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}) \\
&< 0
\end{aligned}$$

and

$$\begin{aligned}
& \bar{\mathcal{Y}}_{L,q,2} - \bar{\mathcal{Y}}'_{R,q,2} \\
&= \left(\frac{\bar{\mathcal{X}}_{L,q,2} - \bar{\mathcal{X}}_{L,q,1}}{2} - \frac{\bar{\mathcal{X}}_{R,q,2} - \bar{\mathcal{X}}_{R,q,1}}{2} \right) \\
&\quad + \gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}) \\
&> 0.
\end{aligned}$$

Since $\bar{\mathcal{Y}}_{L,q,1} < \bar{\mathcal{Y}}'_{R,q,1}$ and $\bar{\mathcal{Y}}_{L,q,2} > \bar{\mathcal{Y}}'_{R,q,2}$, the watermark bit “1” is extracted.

In the same way, it can be verified that the watermark bit embedded by using any set of α values satisfying (17) and (18), or (19) and (20) can be extracted from the q th subsegment pair $\mathcal{Y}_{L,q}(k)$ and $\mathcal{Y}_{R,q}(k)$. Based on the watermark bits extracted from the Q subsegment pairs, the majority rule is then used to determine whether the embedded watermark in the given segment pair is “0” or “1”. Similarly, watermarks can be extracted from other watermarked segment pairs.

Remark 1: From (24) and (25), we can see that the correct extraction of watermarks is determined by the relationship between $|(\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}) - (\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2})|$ and $(\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2})$. More precisely, the smaller the value of $R_1 = |(\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}) - (\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2})| / (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2})$, the better watermark extraction performance. This value is usually small for stereo signals. However, if only one sound channel, say the left sound channel, is considered, the corresponding value becomes $R_2 = |(\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2})| / \bar{\mathcal{X}}_{L,q,1}$. We used 1000 subsegment pairs randomly selected from various stereo music genres to compute R_1 and R_2 , and we got $R_1 = 0.0107$ and $R_2 = 0.0855$. Clearly, R_1 is much smaller than R_2 . This means that using stereo signal can considerably improve the performance of watermark extraction.

III. ANALYSIS OF ROBUSTNESS AGAINST ATTACKS

An effective audio watermarking method must withstand, to a large extent, intentional and unintentional attacks. In this section, we briefly analyzed the robustness of the proposed watermarking method against some conventional attacks.

1) *Compression attack:* Compression attack is very common in practice and often occurs unintentionally, e.g., someone compresses audio data to reduce data size. Two typical compression algorithms are MPEG 1 Layer III (MP3) and MPEG 4 advanced audio coding (AAC). It is known that human auditory system is not sensitive to frequency components above certain frequency threshold f_t and is also insensitive to small-amplitude frequency components masked by large-amplitude neighbouring frequency components [25]. f_t is 20kHz in general but could reduce to about 16kHz for adults. Taking advantage of this fact, the compression algorithms remove high frequency components and small-amplitude frequency components from a signal.

The proposed watermarking method uses the frequency region (f_{min}, f_{max}) to embed watermarks. If f_{max} is properly chosen to satisfy $f_{max} < f_t$, discarding high frequency components will not affect watermark extraction. On the other hand, the new method extracts watermark from a subsegment pair by comparing $\bar{\mathcal{Y}}_{L,q,i}$ and $\bar{\mathcal{Y}}'_{R,q,i}$, which are two mean values associated with the left and right channels of the stereo audio signal respectively. Clearly, the small-amplitude frequency components have very limited contributions to $\bar{\mathcal{Y}}_{L,q,i}$ and $\bar{\mathcal{Y}}'_{R,q,i}$. Besides, both the left and right channels contain similar frequency contents. Thus, removing small-amplitude frequency components from both channels simultaneously has little impact on watermark extraction. Therefore, the proposed method is robust to compression attack.

2) *Noise and re-quantization attacks:* Re-quantization attack can be considered as a type of noise attack because re-quantization adds quantization noise to the watermarked signal. Both noise attack and re-quantization attack are additive to the signal in time domain and subsequently are additive in frequency domain as well. Furthermore, when a stereo audio signal is attacked by these attacks, both the left and right channels are affected in the same manner. Consequently, adding similar noise terms to $\bar{\mathcal{Y}}_{L,q,i}$ and $\bar{\mathcal{Y}}'_{R,q,i}$ does not significantly change $\bar{\mathcal{Y}}_{L,q,i} - \bar{\mathcal{Y}}'_{R,q,i}$.

3) *Re-sampling (RS) attack:* Under this attack, watermarked signals are down-sampled and then up-sampled (or the other way) back to its original sampling rate. Assume that the original sampling rate is f_0 and is reduced to f_d in the down-sampling process. To avoid frequency aliasing, frequency components greater than $f_d/2$ are removed from a signal [31]. This is similar to the low-pass filtering scenario in compression attack. If f_{max} is selected appropriately, the impact of re-sampling attack on watermark extraction is very small.

4) *Amplitude attack:* Under this attack, the amplitudes of watermarked stereo audio signals are scaled by a positive constant. Obviously, the mono audio signals in the left and right channels will be multiplied by the same constant to preserve perceptual quality after the attack. Subsequently, the DFT coefficients of the mono signal in the left channel and

those of the mono signal in the right channel are scaled by the same constant. Clearly, this will not alter the sign of the difference between $\bar{Y}_{L,q,i}$ and $\bar{Y}'_{R,q,i}$. As a result, amplitude attack does not have an effect on the performance of our watermarking method.

5) *Filtering attack* : This attack is either based on low-pass filtering or high-pass filtering, which removes perceptually insignificant portion of the frequency spectrum from the watermarked audio signal. In the proposed method, one watermark is embedded multiple times into an audio segment pair within the frequency range (f_{min}, f_{max}) . Since most part of this frequency range corresponds to the perceptually significant region, only very few subsegment pairs will be affected by a filtering attack. Furthermore, the watermark embedded in an audio segment pair is determined by using majority rule, so filtering attack has negligible effect on the robustness of the proposed method.

In addition to the above discussions, we would like to note that in watermark embedding, the proposed method inserts the same watermark bit into multiple subsegment pairs of a segment pair. If the subsegment pairs severely affected by attacks are minority, correct watermark extraction can still be achieved. This further enhance the robustness of our method.

IV. EXPERIMENTAL RESULTS

In this section, experimental results are provided to illustrate the performance of the proposed watermarking method. In the experiments, we used 50 stereo audio clips belonging to two different groups as host signals, which are as follows:

- *Western music (WM)*: 25 clips containing English pop, jazz, rock and roll music;
- *Eastern music (EM)*: 25 clips containing Eastern classical, country and folk music.

All these stereo audio clips have a duration of 10 seconds. They are sampled at the rate of 44.1kHz, quantized with 16 bits, and then segmented. Each stereo audio segment contains 4410 samples and each sample includes two values corresponding to the left and right sound channels, respectively. Other parameters used in the experiments are $f_{min} = 20\text{Hz}$, $f_{max} = 10\text{kHz}$, $M = 15$ and $\sigma = 0.1$.

A practically useful watermarking method should ensure that the watermarked signals have good perceptual quality and are robust to conventional attacks. We employ the Perceptual Evaluation of Audio Quality (PEAQ) algorithm [29], as used in [1] and [30], to assess the imperceptibility of the watermarked signals. The PEAQ algorithm compares the quality of the host (un-watermarked) signal with its watermarked counterpart and returns a parameter called Objective Difference Grade (ODG). The ODG value ranges from -4 to 0 , where the higher ODG value the better perceptual quality.

To measure the robustness, we define the detection rate (DR) as follows:

$$\text{DR} = \left(\frac{\text{Number of watermarks correctly extracted}}{\text{Number of watermarks embedded}} \right) \times 100\%.$$

The following common attacks are utilized in the evaluation of robustness:

- *Closed-loop attack*: The watermarks are extracted from the watermarked signals without any attacks.
- *Re-quantization attack*: Each sample of the watermarked signals is re-quantized to 8 bits (i.e., 16 bits \rightarrow 8 bits re-quantization) [20], [21].
- *RS attack*: The watermarked signals are down-sampled to 22.05kHz and 16kHz, respectively, and then up-sampled back to 44.1kHz (i.e., 44.1kHz \rightarrow 22.05kHz \rightarrow 44.1kHz re-sampling and 44.1kHz \rightarrow 16kHz \rightarrow 44.1kHz re-sampling).
- *Noise attack*: Random noise is added to the watermarked signals, where the ratio of the watermarked signal to noise is 20dB.
- *Amplitude attack*: The amplitudes of the watermarked signals are increased by 1.2 times and 1.8 times, respectively.
- *MP3 attack*: MPEG 1 Layer III compression is performed on the watermarked signals.
- *AAC attack*: MPEG 4 advanced audio coding based compression is performed on the watermarked signals.
- *HPF attack*: High-pass filters with cutoff frequencies 50Hz and 100Hz are applied to the watermarked signal.
- *LPF attack*: Low-pass filters with cutoff frequencies 12kHz and 8kHz are applied to the watermarked signal.

Firstly, we compare the proposed method with the patchwork method in [1]. Specifically, we compare the robustness of both methods under the same perceptual quality with $\text{ODG} = -0.3$ and the same embedding rate of 10bps. The ODG value of -0.3 ensures that the watermarked signals by both methods have high imperceptibility. Table I shows the DRs of these methods under the above mentioned common attacks. Here, the bit rate of 128kbps is used for MP3 and AAC attacks. One can see that while both methods are very robust to all attacks, our method consistently outperforms the method in [1]. It is important to point out that the method in [1] requires additional information of which segments of the watermarked audio signal contain watermarks. In the experiment, we assume that this information is known at the decoding end for the method in [1]. However, to our best knowledge, identifying the watermarked segments encountered in [1] is still an open problem. So the usage of this watermarking method is restrictive in practice. In contrast, our method does not require any additional information to find the watermarked subsegments in the decoding process.

Secondly, we compare our method with the watermarking method in [27], which is proposed for stereo audio signals. The ODG value of this method is -1.5 . To do a fair comparison, we also use the ODG value of -1.5 for our method, which can be achieved by adjusting γ . It can be seen from Table II that the proposed method performs better than the method in [27]. We would like to note that since the ODG value of -1.5 is far below zero, the watermarked audio signals are of low perceptual quality, which is verified by our own listening test. When playing these watermarked audio signals, we can hear obvious watermark-induced noise. Since the perceptual quality of the watermarked signals by the method in [27] cannot be improved by altering any parameter, this method is not suitable for practical applications.

TABLE I

DETECTION RATES OF THE PROPOSED METHOD AND THE METHOD IN [1], WHERE ODG=-0.3 FOR BOTH METHODS

Attacks	Host signals	DR (%)	
		Method in [1]	Proposed method
Closed-loop	WM	100	100
	EM	100	100
Re-quantization	WM	99.12	99.96
	EM	99.72	99.92
RS (44/22/44)	WM	100	100
	EM	100	100
RS (44/16/44)	WM	99.72	99.96
	EM	98.72	100
Noise	WM	94.52	99.68
	EM	97.16	99.76
Amplitude (1.2)	WM	100	100
	EM	100	100
Amplitude (1.8)	WM	100	100
	EM	100	100
MP3 (128kbps)	WM	99.80	100
	EM	99.80	100
AAC (128kbps)	WM	99.64	100
	EM	97.00	100
HPF (50Hz)	WM	100	100
	EM	100	100
HPF (100Hz)	WM	100	100
	EM	100	100
LPF (12kHz)	WM	100	100
	EM	100	100
LPF (8kHz)	WM	99.95	99.96
	EM	100	100

TABLE II

DETECTION RATES OF THE PROPOSED METHOD AND THE METHOD IN [27], WHERE ODG=-1.5 FOR BOTH METHODS

Attacks	Host signals	DR (%)	
		Method in [27]	Proposed method
Closed-loop	WM	100	100
	EM	100	100
Re-quantization	WM	99.04	100
	EM	99.40	100
RS (44/22/44)	WM	99.68	100
	EM	99.84	100
RS (44/16/44)	WM	98.20	100
	EM	99.44	100
Noise	WM	97.36	100
	EM	98.24	100
Amplitude (1.2)	WM	100	100
	EM	100	100
Amplitude (1.8)	WM	100	100
	EM	100	100
MP3 (128kbps)	WM	99.32	100
	EM	99.56	100
AAC (128kbps)	WM	99.48	100
	EM	99.16	100
HPF (50Hz)	WM	100	100
	EM	100	100
HPF (100Hz)	WM	100	100
	EM	100	100
LPF (12kHz)	WM	100	100
	EM	99.96	100
LPF (8kHz)	WM	99.84	100
	EM	99.76	100

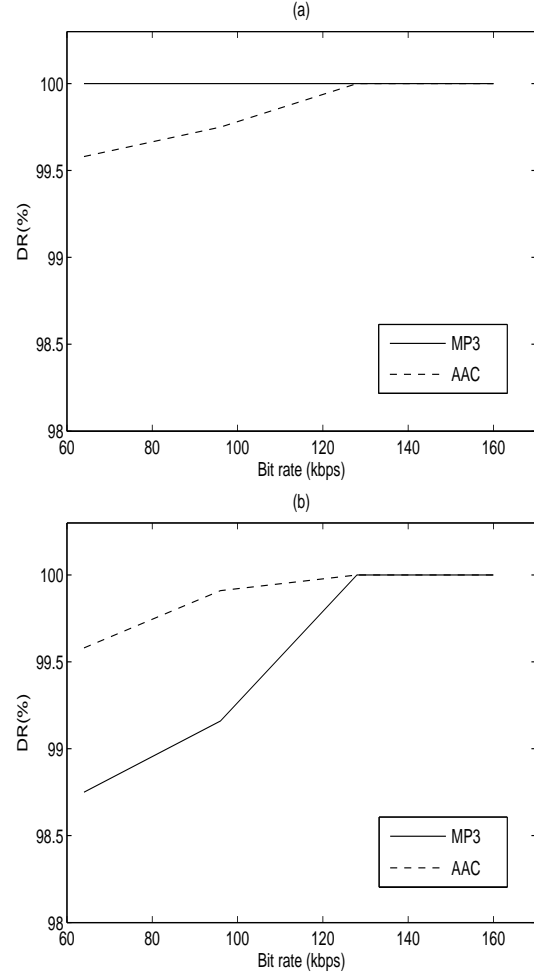


Fig. 2. DRs under MP3 and AAC attacks verses bit rate. (a) Western music (b) Eastern music

Finally, we evaluate robustness of the proposed method against MP3 and AAC attacks under different bit rates: 64kbps, 96kbps, 128kbps, and 160kbps. These compression bit rates are widely used in real world applications. As shown in Fig. 2, satisfactory detection rates have been achieved under both compression attacks, at different bit rates, and for both western music and eastern music. Also, as expected, the detection rates improve with the increase of bit rate. When the bit rate is 128kbps or higher, the embedded watermarks are extracted without any error.

V. CONCLUSION

In this paper, a robust patchwork-based audio watermarking method is developed for stereo audio signals, which hides watermarks into the two sound channels of the host audio signal in frequency domain. The watermarks are embedded in such a way that only certain frequency region is used for watermarking, each watermark bit is inserted into multiple DFT subsegment pairs, and watermark embedding does not change the mean of moduli of a subsegment. In the decoding process, the special features of the watermark embedding scheme and the similarity of the two sound channels are

exploited to identify the watermarked subsegment pairs and then to extract the embedded watermarks. Its robustness is analyzed in theory and verified by experimental results. The new method is also secure, has high imperceptibility, and does not need the host audio signal for watermark decoding. Compared with existing patchwork watermarking methods, our method does not require information of which segments of the watermarked audio signal enclose watermarks and is more robust to conventional attacks.

REFERENCES

- [1] N.K. Kalantari, M.A. Akhaee, S.M. Ahadi, and H. Amindavar, "Robust multiplicative patchwork method for audio watermarking," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 17, no. 6, pp. 1133-1141, Aug. 2009.
- [2] L. Luo, Z. Chen, M. Chen, X. Zeng, and Z. Xiong, "Reversible image watermarking using interpolation technique," *IEEE Trans. Information Forensics and Security*, vol. 5, no. 1, pp. 187-193, Mar. 2010.
- [3] C.H. Huang, S.C. Chuang, Y.L. Huang, and J.L. Wu, "Unseen visible watermarking: A novel methodology for auxiliary information delivery via visual contents," *IEEE Trans. Information Forensics and Security*, vol. 4, no. 2, pp. 193-206, Jun. 2009.
- [4] H. Huang, C. Yang, and W. Hsu, "A video watermarking technique based on pseudo-3-D DCT and quantization index modulation," *IEEE Trans. Information Forensics and Security*, vol. 5, no. 4, pp. 625-637, Dec. 2010.
- [5] Y. Erfani and S. Siahpoush, "Robust audio watermarking using improved TS echo hiding," *Digital Signal Processing*, vol. 19, pp. 809-814, 2009.
- [6] H. Wang, R. Nishimura, Y. Suzuki, and L. Miao, "Fuzzy self-adaptive digital audio watermarking based on time-spread echo hiding," *Applied Acoustics*, vol. 69, pp. 868-874, 2008.
- [7] Y. Xiang, D. Peng, I. Natgunanathan, and W. Zhou, "Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo based audio watermarking," *IEEE Trans. Multimedia*, vol. 13, no. 1, pp. 2-13, Feb. 2011.
- [8] I. K. Yeo and H. J. Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 4, pp. 381-386, Jul. 2003.
- [9] Y. H. Lin and J. L. Wu, "A digital blind watermarking for depth-image-based rendering 3D images," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 602-611, Jun. 2011.
- [10] A. K. Parthasarathy and S. Kak, "An improved method of content based image watermarking," *IEEE Trans. Broadcasting*, vol. 53, no. 2, pp. 468-479, Jun. 2007.
- [11] L. Wang, H. Ling, F. Zou, and Z. Lu, "Real-time compressed-domain video watermarking resistance to geometric distortions," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 70-79, Jan. 2012.
- [12] C. S. Liu, *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property*. Hershey, PA: Idea Group Publishing, 2005.
- [13] W.-N. Lie and L.-C. Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 46-59, Feb. 2006.
- [14] D. Kirovski and H. Malvar, "Robust spread-spectrum audio watermarking," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2001, pp. 1345-1348.
- [15] A. Valizadeh and Z. J. Wang, "Correlation-and-bit-aware spread spectrum embedding for data hiding," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 2, pp. 267-282, Jun. 2011.
- [16] H. S. Malvar and D. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 898-905, Mar. 2003.
- [17] S. Kirbiz, and B. Günsel, "Robust audio watermark decoding by supervised learning," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, 2006.
- [18] X. Wang, W. Qi, and P. Niu, "A new adaptive digital audio watermarking based on support vector regression," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 15, no. 8, pp. 2270-2277, Nov. 2007.
- [19] D. Lakshmi, R. Ganesh, R. Marni, R. Prakash, and P. Arulmozhivarman, "SVM based effective watermarking scheme for embedding binary logo and audio signals in images," in *Proc. TENCON 2008 - 2008 IEEE Region 10 Conference.*, 2008, pp. 1-5.
- [20] B.-S. Ko, R. Nishimura, and Y. Suzuki, "Time-spread echo method for digital audio watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 212-221, Apr. 2005.
- [21] O. T.-C. Chen and W.-C. Wu, "Highly robust, secure, and perceptual-quality echo hiding scheme," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 16, no. 3, pp. 629-638, Mar. 2008.
- [22] M. Arnold, "Audio watermarking: Features, applications and algorithm," in *Proc. IEEE Int. Conf. Multimedia Expo 2000*, 2000, vol. 2, pp. 1013-1016.
- [23] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Syst. J.*, vol. 35, no. 3&4, 1996.
- [24] k. Solanki, N. Jacobsen, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Robust image-adaptive data hiding based on erasure and error correction," *IEEE Trans. Image Process.*, vol. 13, no. 12, pp. 1627-1639, Dec. 2004.
- [25] A. Takahashi, R. Nishimura, and Y. Suzuki, "Multiple watermarks for stereo audio signals using phase-modulation techniques," *IEEE Trans. Signal Process.* vol. 53, no. 2, pp. 806-815, Feb. 2005.
- [26] S. W. Foo, "Audio-watermarking with stereo signals," in *Proc. TENCON 2008 - 2008 IEEE region 10 conf.*, 2008, pp. 1-4.
- [27] K. Kondo and K. Nakagawa, "A digital watermark for stereo audio signal using variable interchannel delay in high frequency bands," in *Proc. Int. Conf. Intelligent Information Hiding and Multimedia Signal Process.*, 2008, pp. 624-627.
- [28] W. Cao, Y. Yan, and S. Li, "Bit replacement audio watermarking using stereo signals," in *Proc. Int. Conf. New Trends in Information and Service*, 2009, pp. 603-606.
- [29] *Rec.B. S. 1387: Methods for Objective Measurements of Perceived Audio Quality*, Rec. B.S. 1387, Int. Telecomm. Union, Geneva, Switzerland, 2001.
- [30] C. Baras, N. Moreau, and P. Dymarski, "Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 5, pp. 1772-1782, Sept. 2006.
- [31] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1998.