Multimed Tools Appl (2014) 72:1387–1410
DOI 10.1007/s11042-013-1454-4

Robust patchwork-based watermarking method for stereo audio signals

Iynkaran Natgunanathan · Yong Xiang · Yue Rong · Dezhong Peng

Published online: 16 April 2013

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.

I. Natgunanathan · Y. Xiang (✉)
School of Information Technology, Deakin University,
Burwood Campus, Melbourne, VIC 3125, Australia
e-mail: yxiang@deakin.edu.au

I. Natgunanathan
e-mail: iynkaran.n@research.deakin.edu.au

Y. Rong
Department of Electrical and Computer Engineering,
Curtin University, Bentley, WA 6102, Australia
e-mail: y.rong@curtin.edu.au

D. Peng
Machine Intelligence Laboratory, College of Computer Science,
Sichuan University, Chengdu 610065, China
e-mail: pengdz@scu.edu.cn

Keywords Audio watermarking · Patchwork · Stereo audio signal · Discrete Fourier transform

1 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 [8, 10, 13–15, 22–25, 27, 29, 34, 41, 44, 45, 47], 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 [22–25, 29, 34, 41] and video [8], we limit our attention to audio watermarking in this paper.

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 [21, 26]. 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 [15]. Over the past decade, many audio watermarking methods have been developed by using different techniques such as spread-spectrum [17, 28, 38], support vector regression [16, 20, 39], low frequency modification [9, 21], transform domain [11], compressed domain [35], singular value decomposition [1, 5], echo-hiding [7, 10, 18, 44, 45], and patchwork [2, 15, 30, 47]. 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 [4] and then Arnold applied this technique to audio watermarking [2]. After that, the modified patchwork algorithm (MPA) was proposed by Yeo and Kim to improve watermarking performance [47]. 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 [15]. Kalantari et al. proposed a multiplicative patchwork method in [15] to deal with this problem. In [15], 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, Kalantari et al. [15] does not provide an answer to this question. While an approach was proposed in [36] to estimate the indices of selected image frames, it cannot be directly applied nor simply modified to identify the watermarked segments encountered in [15]. Recently, Natgunanathan et al. have proposed an audio watermarking method based on the patchwork concept [30]. In [30], one audio segment is divided into two subsegments and two sets of patches are chosen from both subsegments. Watermarks are only embedded into the selected patches which are decided based on a criterion. This method needs long audio segments to obtain a satisfactory detection rate. Its performance deteriorates considerably with the increase of embedding rate.

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 [6, 12, 19, 37]. In [37], 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 [12] is non-blind and the method in [19] is not secure. In [6], 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. Since human auditory system is insensitive to high frequency components and very low frequency components of audio signals, these frequency regions are not suitable for watermark embedding as watermarks embedded in these frequency regions can be easily removed by some intentional or unintentional attacks such as compression attack. As a result, the proposed method is designed in the discrete Fourier transform (DFT) domain to ensure that watermarks will not be embedded into these frequency regions. 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 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 [15, 19, 30]. The remainder of this paper is organized as follows. The new method is presented in Section 2 and its robustness against conventional attacks is analyzed in Section 3. The experimental results are shown in Section 4. Finally, Section 5 concludes the paper.

2 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.

2.1 Watermark embedding

2.1.1 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 mth 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

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

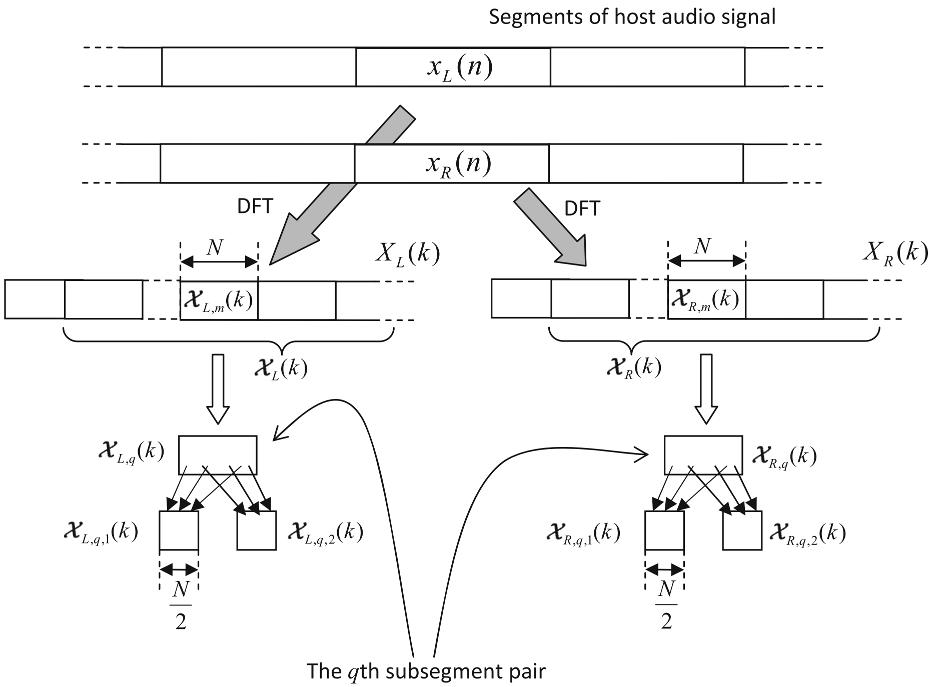

Fig. 1 Illustration of host audio signal segmentation and DFT coefficients classification

From (1), it follows

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

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

where $m = 1, 2, \ldots, M$. Here $\mathcal{X}_{L,m}(k)$ and $\mathcal{X}_{R,m}(k)$ form the mth 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

$$\bar{\mathcal{X}}_{L,m} = \frac{1}{N}\sum_{l=1}^{N} |a_{m,l}|, \tag{4}$$

$$\bar{\mathcal{X}}_{R,m} = \frac{1}{N}\sum_{l=1}^{N} |b_{m,l}|. \tag{5}$$

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

$$\min\left\{\bar{\mathcal{X}}_{L,m},\ \bar{\mathcal{X}}_{R,m}\right\} \geq \sigma \tag{6}$$

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 (6), 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 (6) and are selected for watermark embedding, where $Q \leq M$.

2.1.2 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 qth subsegment pair, into two pairs of groups. Let

$$p(n) = \{p_1, p_2, \ldots, p_N\} \tag{7}$$

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

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 (7), the first pair of groups corresponding to the qth subsegment pair consisting of $\mathcal{X}_{L,q}(k)$ and $\mathcal{X}_{R,q}(k)$ can be obtained by

$$\mathcal{X}_{L,q,1}(k) = \left\{a_{q,p_1},\ a_{q,p_2},\ \ldots,\ a_{q,p_{N/2}}\right\} \tag{8}$$

$$\mathcal{X}_{R,q,1}(k) = \left\{b_{q,p_1},\ b_{q,p_2},\ \ldots,\ b_{q,p_{N/2}}\right\} \tag{9}$$

where $q = 1, 2, \ldots, Q$. Similarly, the second pair of groups associated with the qth subsegment pair can be constructed by using the last $N/2$ elements of $p(n)$ as

$$\mathcal{X}_{L,q,2}(k) = \left\{a_{q,p_{N/2+1}},\ a_{q,p_{N/2+2}},\ \ldots,\ a_{q,p_N}\right\} \tag{10}$$

$$\mathcal{X}_{R,q,2}(k) = \left\{b_{q,p_{N/2+1}},\ b_{q,p_{N/2+2}},\ \ldots,\ b_{q,p_N}\right\} \tag{11}$$

where $q = 1, 2, \ldots, Q$. 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 $\mathcal{X}_{L,q,i}(k)$ and $\mathcal{X}_{R,q,i}(k)$, where $i = 1, 2$ and $q = 1, 2, \ldots, Q$.

2.1.3 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 $\mathcal{X}_{L,q,1}(k)$, $\mathcal{X}_{R,q,1}(k)$, $\mathcal{X}_{L,q,2}(k)$ and $\mathcal{X}_{R,q,2}(k)$, respectively. We insert a digital watermark into the

qth subsegment pair by modifying the DFT coefficients in the two pairs of groups as follows:

$$\begin{cases} \mathcal{Y}_{L,q,1}(k) = \alpha_{L,q,1} \times \mathcal{X}_{L,q,1}(k) \\ \mathcal{Y}_{R,q,1}(k) = \alpha_{R,q,1} \times \mathcal{X}_{R,q,1}(k) \end{cases} \tag{12}$$

and

$$\begin{cases} \mathcal{Y}_{L,q,2}(k) = \alpha_{L,q,2} \times \mathcal{X}_{L,q,2}(k) \\ \mathcal{Y}_{R,q,2}(k) = \alpha_{R,q,2} \times \mathcal{X}_{R,q,2}(k) \end{cases}. \tag{13}$$

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

$$\alpha_{\max} = 1 + \gamma \tag{14}$$

$$\alpha_{\min} = 1 - \gamma \tag{15}$$

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

$$\begin{cases} \alpha_{L,q,1} > 1 \\ \alpha_{R,q,1} < 1 \end{cases} \text{ and } \begin{cases} \alpha_{L,q,2} < 1 \\ \alpha_{R,q,2} > 1 \end{cases}. \tag{16}$$

Similarly, to embed watermark bit "1", we set

$$\begin{cases} \alpha_{L,q,1} < 1 \\ \alpha_{R,q,1} > 1 \end{cases} \text{ and } \begin{cases} \alpha_{L,q,2} > 1 \\ \alpha_{R,q,2} < 1 \end{cases}. \tag{17}$$

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 $\alpha_{\min}$ and $\alpha_{\max}$ as possible) to ensure satisfactory robustness.

Furthermore, we require that the modifications in (12) and (13) 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

$$\bar{\mathcal{Y}}_{L,q} = \bar{\mathcal{X}}_{L,q} \text{ or } \frac{\bar{\mathcal{Y}}_{L,q,1} + \bar{\mathcal{Y}}_{L,q,2}}{2} = \frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{2} \tag{18}$$

and

$$\bar{\mathcal{Y}}_{R,q} = \bar{\mathcal{X}}_{R,q} \text{ or } \frac{\bar{\mathcal{Y}}_{R,q,1} + \bar{\mathcal{Y}}_{R,q,2}}{2} = \frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{2}. \tag{19}$$

Ensuring (18) and (19) in watermark embedding is important as they can be exploited to identify the watermarked subsegment pairs at the decoding end.

Now the challenging task is how to find the positive real constants $\alpha_{L,q,1}$, $\alpha_{R,q,1}$, $\alpha_{L,q,2}$ and $\alpha_{R,q,2}$ such that (16), (18) and (19) are ensured when the watermark bit

"0" is embedded, and (17), (18) and (19) are satisfied when the watermark bit "1" is embedded. To proceed, we first define

$$f(\theta_1, \theta_2, \theta_3) = -\left(\frac{\theta_1}{\theta_2}\right)\theta_3 + \frac{\theta_1 + \theta_2}{\theta_2}. \tag{20}$$

Then we propose to insert watermark bit "0" into the subsegment pair by using the following α values in (12) and (13):

$$\alpha_{L,q,1},\ \alpha_{L,q,2}: \begin{cases} \text{if } \bar{\mathcal{X}}_{L,q,1} < \bar{\mathcal{X}}_{L,q,2} \\ \quad \alpha_{L,q,1} = \alpha_{\max} = 1+\gamma \\ \quad \alpha_{L,q,2} = f\left(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1+\gamma)\right) \\ \text{Otherwise} \\ \quad \alpha_{L,q,1} = f\left(\bar{\mathcal{X}}_{L,q,2}, \bar{\mathcal{X}}_{L,q,1}, (1-\gamma)\right) \\ \quad \alpha_{L,q,2} = \alpha_{\min} = 1-\gamma \end{cases} \tag{21}$$

and

$$\alpha_{R,q,1},\ \alpha_{R,q,2}: \begin{cases} \text{if } \bar{\mathcal{X}}_{R,q,1} < \bar{\mathcal{X}}_{R,q,2} \\ \quad \alpha_{R,q,1} = \alpha_{\min} = 1-\gamma \\ \quad \alpha_{R,q,2} = f\left(\bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2}, (1-\gamma)\right) \\ \text{Otherwise} \\ \quad \alpha_{R,q,1} = f\left(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1+\gamma)\right) \\ \quad \alpha_{R,q,2} = \alpha_{\max} = 1+\gamma \end{cases}. \tag{22}$$

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

$$\alpha_{L,q,1},\ \alpha_{L,q,2}: \begin{cases} \text{if } \bar{\mathcal{X}}_{L,q,1} < \bar{\mathcal{X}}_{L,q,2} \\ \quad \alpha_{L,q,1} = \alpha_{\min} = 1-\gamma \\ \quad \alpha_{L,q,2} = f\left(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1-\gamma)\right) \\ \text{Otherwise} \\ \quad \alpha_{L,q,1} = f\left(\bar{\mathcal{X}}_{L,q,2}, \bar{\mathcal{X}}_{L,q,1}, (1+\gamma)\right) \\ \quad \alpha_{L,q,2} = \alpha_{\max} = 1+\gamma \end{cases} \tag{23}$$

and

$$\alpha_{R,q,1},\ \alpha_{R,q,2}: \begin{cases} \text{if } \bar{\mathcal{X}}_{R,q,1} < \bar{\mathcal{X}}_{R,q,2} \\ \quad \alpha_{R,q,1} = \alpha_{\max} = 1+\gamma \\ \quad \alpha_{R,q,2} = f\left(\bar{\mathcal{X}}_{R,q,1}, \bar{\mathcal{X}}_{R,q,2}, (1+\gamma)\right) \\ \text{Otherwise} \\ \quad \alpha_{R,q,1} = f\left(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1-\gamma)\right) \\ \quad \alpha_{R,q,2} = \alpha_{\min} = 1-\gamma \end{cases}. \tag{24}$$

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 (21)–(24) that there exist eight sets of α values. It can be shown that these sets of α values either satisfy (16) or (17), depending on the watermark bit to be embedded. Furthermore, they also satisfy (18) and (19). 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 (21) and (22) that $\alpha_{L,q,1} = 1 + \gamma$, $\alpha_{L,q,2} = f\left(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1+\gamma)\right)$, $\alpha_{R,q,1} = f\left(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1+\gamma)\right)$ and $\alpha_{R,q,2} = 1 + \gamma$. Considering (20), it is easy to verify that this set of α values satisfy (16), (18) and (19).

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 (23) and (24) that the set of α values are $\alpha_{L,q,1} = 1 - \gamma$, $\alpha_{L,q,2} = f\left(\bar{\mathcal{X}}_{L,q,1}, \bar{\mathcal{X}}_{L,q,2}, (1-\gamma)\right)$, $\alpha_{R,q,1} = f\left(\bar{\mathcal{X}}_{R,q,2}, \bar{\mathcal{X}}_{R,q,1}, (1-\gamma)\right)$ and $\alpha_{R,q,2} = 1 - \gamma$, which satisfy (17)–(19).

Based on the watermark embedding scheme in (12) and (13), 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.

In summary, in the proposed watermarking method, the watermarks containing copyright information are added to the audio signal by modifying its frequency components. Only the frequency components in the range between $f_{\min}$ and $f_{\max}$ are modified. In other words, the watermarks are embedded into certain frequency range of the audio signal. This makes the proposed method robust against high-pass filtering, low-pass filtering and compression attacks. As will be shown in Section 3, the new method is also robust to many other common attacks such as noise, re-quantization, re-sampling, and amplitude attacks. All these make the proposed method more useful in practical applications.

2.2 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.

2.2.1 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 $\mathcal{Y}_{R,m}(k)$, $m = 1, 2, \ldots, M$. 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 (6) 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 Section 2.1, the proposed watermark embedding scheme ensures (18) and (19), i.e.,

$$\bar{\mathcal{Y}}_{L,m} = \bar{\mathcal{X}}_{L,m} \text{ and} \tag{25}$$

$$\bar{\mathcal{Y}}_{R,m} = \bar{\mathcal{X}}_{R,m}. \tag{26}$$

Based on this property, a criterion similar to (6) can be proposed to examine whether a subsegment pair is watermarked or not. Specifically, if

$$\min\left\{\bar{\mathcal{Y}}_{L,m},\ \bar{\mathcal{Y}}_{R,m}\right\} \geq \sigma \tag{27}$$

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 (27). 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 (27), the concerned segment pair does not contain a watermark and watermark extraction should not be conducted.

2.2.2 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 $q = 1, 2, \ldots, Q$. 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 qth subsegment pair $\mathcal{Y}_{L,q}(k)$ and $\mathcal{Y}_{R,q}(k)$, we define

$$\bar{\mathcal{Y}}'_{R,q,1} = \bar{\mathcal{Y}}_{R,q,1} + \left(\bar{\mathcal{Y}}_{L,q} - \bar{\mathcal{Y}}_{R,q}\right) \tag{28}$$

$$\bar{\mathcal{Y}}'_{R,q,2} = \bar{\mathcal{Y}}_{R,q,2} + \left(\bar{\mathcal{Y}}_{L,q} - \bar{\mathcal{Y}}_{R,q}\right) \tag{29}$$

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 qth subsegment pair is "0".
- Otherwise, the watermark bit embedded in the qth 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 (Section 2.1), 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

$$\alpha_{L,q,1} = 1+\gamma \tag{30}$$

$$\alpha_{L,q,2} = -\left(\frac{\bar{\mathcal{X}}_{L,q,1}}{\bar{\mathcal{X}}_{L,q,2}}\right)(1+\gamma) + \frac{\bar{\mathcal{X}}_{L,q,1}+\bar{\mathcal{X}}_{L,q,2}}{\bar{\mathcal{X}}_{L,q,2}} \tag{31}$$

$$\alpha_{R,q,1} = -\left(\frac{\bar{\mathcal{X}}_{R,q,2}}{\bar{\mathcal{X}}_{R,q,1}}\right)(1+\gamma) + \frac{\bar{\mathcal{X}}_{R,q,1}+\bar{\mathcal{X}}_{R,q,2}}{\bar{\mathcal{X}}_{R,q,1}} \tag{32}$$

$$\alpha_{R,q,2} = 1+\gamma \tag{33}$$

where $\alpha_{L,q,2}$ and $\alpha_{R,q,1}$ result from (20). Based on this set of α values, it follows from (12) and (13) that

$$\bar{\mathcal{Y}}_{L,q,1} = (1+\gamma)\bar{\mathcal{X}}_{L,q,1} \tag{34}$$

$$\bar{\mathcal{Y}}_{L,q,2} = -\bar{\mathcal{X}}_{L,q,1}(1+\gamma) + \bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2} \tag{35}$$

$$\bar{\mathcal{Y}}_{R,q,1} = -\bar{\mathcal{X}}_{R,q,2}(1+\gamma) + \bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2} \tag{36}$$

$$\bar{\mathcal{Y}}_{R,q,2} = (1+\gamma)\bar{\mathcal{X}}_{R,q,2}. \tag{37}$$

Then, from (28) and (29), it yields

$$\begin{aligned}\bar{\mathcal{Y}}'_{R,q,1} = &-\bar{\mathcal{X}}_{R,q,2}(1+\gamma) + \bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}\\ &+\left(\frac{\bar{\mathcal{X}}_{L,q,1}+\bar{\mathcal{X}}_{L,q,2}}{2}\right)-\left(\frac{\bar{\mathcal{X}}_{R,q,1}+\bar{\mathcal{X}}_{R,q,2}}{2}\right)\end{aligned} \tag{38}$$

$$\begin{aligned}\bar{\mathcal{Y}}'_{R,q,2} = &(1+\gamma)\bar{\mathcal{X}}_{R,q,2}\\ &+\left(\frac{\bar{\mathcal{X}}_{L,q,1}+\bar{\mathcal{X}}_{L,q,2}}{2}\right)-\left(\frac{\bar{\mathcal{X}}_{R,q,1}+\bar{\mathcal{X}}_{R,q,2}}{2}\right).\end{aligned} \tag{39}$$

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

$$\begin{aligned}\bar{\mathcal{Y}}_{L,q,1} - \bar{\mathcal{Y}}'_{R,q,1} &= (1+\gamma)\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}(1+\gamma) - \bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}\\ &\quad - \frac{\bar{\mathcal{X}}_{L,q,1}+\bar{\mathcal{X}}_{L,q,2}}{2} + \frac{\bar{\mathcal{X}}_{R,q,1}+\bar{\mathcal{X}}_{R,q,2}}{2}\end{aligned} \tag{40}$$

$$\begin{aligned}&=\left(\frac{\bar{\mathcal{X}}_{L,q,1}-\bar{\mathcal{X}}_{L,q,2}}{2} - \frac{\bar{\mathcal{X}}_{R,q,1}-\bar{\mathcal{X}}_{R,q,2}}{2}\right)\\ &\quad+\gamma\left(\bar{\mathcal{X}}_{L,q,1}+\bar{\mathcal{X}}_{R,q,2}\right).\end{aligned} \tag{41}$$

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 (40), this implies $\bar{\mathcal{Y}}_{L,q,1} - \bar{\mathcal{Y}}'_{R,q,1} \approx \gamma \left(\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}\right) > 0$.

Similarly, we have

$$\begin{aligned}\bar{\mathcal{Y}}_{L,q,2} - \bar{\mathcal{Y}}'_{R,q,2} &= -\bar{\mathcal{X}}_{L,q,1}(1+\gamma) + \bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2} - (1+\gamma)\bar{\mathcal{X}}_{R,q,2} \\ &\quad - \frac{\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{L,q,2}}{2} + \frac{\bar{\mathcal{X}}_{R,q,1} + \bar{\mathcal{X}}_{R,q,2}}{2} \qquad (42) \\ &= \left(\frac{\bar{\mathcal{X}}_{L,q,2} - \bar{\mathcal{X}}_{L,q,1}}{2} - \frac{\bar{\mathcal{X}}_{R,q,2} - \bar{\mathcal{X}}_{R,q,1}}{2}\right) \\ &\quad - \gamma \left(\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}\right) \qquad (43) \\ &\approx -\gamma \left(\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}\right) \\ &< 0. \qquad (44)\end{aligned}$$

Since $\bar{\mathcal{Y}}_{L,q,1} > \bar{\mathcal{Y}}'_{R,q,1}$ and $\bar{\mathcal{Y}}_{L,q,2} < \bar{\mathcal{Y}}'_{R,q,2}$ 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

$$\begin{aligned}\bar{\mathcal{Y}}_{L,q,1} - \bar{\mathcal{Y}}'_{R,q,1} &= \left(\frac{\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}}{2} - \frac{\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}}{2}\right) \\ &\quad - \gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}) \qquad (45) \\ &< 0\end{aligned}$$

and

$$\begin{aligned}\bar{\mathcal{Y}}_{L,q,2} - \bar{\mathcal{Y}}'_{R,q,2} &= \left(\frac{\bar{\mathcal{X}}_{L,q,2} - \bar{\mathcal{X}}_{L,q,1}}{2} - \frac{\bar{\mathcal{X}}_{R,q,2} - \bar{\mathcal{X}}_{R,q,1}}{2}\right) \\ &\quad + \gamma (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2}) \qquad (46) \\ &> 0.\end{aligned}$$

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 (21) and (22), or (23) and (24) can be extracted from the qth 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 (40) and (42), we can see that the correct extraction of watermarks is determined by the relationship between $\left|\left(\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}\right) - \left(\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}\right)\right|$ and $(\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2})$. More precisely, the smaller the value of $R_1 = \left|\left(\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}\right) - \left(\bar{\mathcal{X}}_{R,q,1} - \bar{\mathcal{X}}_{R,q,2}\right)\right| / (\bar{\mathcal{X}}_{L,q,1} + \bar{\mathcal{X}}_{R,q,2})$, 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 $R_2 = \left|\left(\bar{\mathcal{X}}_{L,q,1} - \bar{\mathcal{X}}_{L,q,2}\right)\right| / \bar{\mathcal{X}}_{L,q,1}$. We used 1,000 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.

3 Analysis of robustness against 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 [37]. f_t is 20 kHz in general but could reduce to about 16 kHz 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 $\bar{\mathcal{Y}}_{L,q,i}$ and $\bar{\mathcal{Y}}'_{R,q,i}$. 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.

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.

Remark 2 Like other patchwork-based audio watermarking methods [2, 15, 47], the proposed method assumes that the encoder and decoder are synchronized. It can be combined with the synchronization mechanisms in [40, 42] to cope with de-synchronization attacks such as fractional delay and cropping.

Remark 3 Although the decoding process can be viewed as a blind signal processing problem, the existing blind methods, such as those in [32, 43, 46], cannot solve this problem as they require certain conditions on the host signal and the mixing system.

4 Experimental results

In this section, experimental results are provided to illustrate the performance of the proposed watermarking method. In the experiments, we used 160 stereo audio clips belonging to four different categories as host signals, which are as follows:

- *Western music (WM):* 40 clips containing pop, jazz, rock and roll music;
- *Eastern music (EM):* 40 clips containing Carnatic, Eastern classical, country and folk music.

- *Speeches (SP):* 40 clips containing male and female voices.
- *Others (OT):* 40 clips containing rain, bird, animal and drum sounds.

All these stereo audio clips have a duration of 60 sec, which sum to a total stereo audio length of 160 min. They are sampled at the rate of 44.1 kHz, quantized with 16 bits, and then segmented. Each stereo audio segment contains 4,410 samples and each sample includes two values corresponding to the left and right sound channels, respectively. This segment length is chosen empirically to ensure satisfactory robustness and embedding rate. It should be noted that higher embedding rate can be obtained by reducing the segment length at the expense of lower robustness. Other parameters used in the experiments are $f_{\min} = 20$ Hz, $f_{\max} = 10$ 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 [33], as used in [3, 15], to asses 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:

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

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 → 8 bits re-quantization) [7, 18].
- RS *attack*: The watermarked signals are down-sampled to 22.05 kHz and 16 kHz, respectively, and then up-sampled back to 44.1 kHz (i.e., 44.1 kHz → 22.05 kHz → 44.1 kHz re-sampling and 44.1 kHz → 16 kHz → 44.1 kHz re-sampling).
- *Noise attack*: Random noise is added to the watermarked signals, where the ratio of the watermarked signal to noise is 20 dB.
- *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 50 Hz and 100 Hz are applied to the watermarked signal.
- LPF *attack*: Low-pass filters with cutoff frequencies 12 kHz and 8 kHz are applied to the watermarked signal.

Firstly, we compare the proposed method with the patchwork methods in [15, 30]. Specifically, we compare the robustness of these methods under the same perceptual quality with ODG = −0.3 and the same embedding rate of 10bps. Since the methods in [15, 30] are designed for mono audio signals, we only used one channel of the stereo audio signals for these methods. The ODG value of −0.3 ensures that the watermarked signals by both methods have high imperceptibility. Table 1 shows the

Table 1 Detection rates of the proposed method, the method in [15] and the method in [30], where ODG = − 0.3 for all three methods

Attacks	Host signals	DR (%)		
		Method in [15]	Method in [30]	Proposed method
Closed-loop	WM	100	100	100
	EM	100	100	100
	SP	100	100	100
	OT	100	100	100
Re-quantization	WM	99.92	99.84	99.99
	EM	99.54	99.26	99.96
	SP	99.57	99.48	99.89
	OT	99.47	98.84	99.80
RS (44/22/44)	WM	98.03	90.07	100
	EM	94.36	89.12	100
	SP	99.12	96.52	100
	OT	98.27	90.04	100
RS (44/16/44)	WM	98.63	89.63	99.98
	EM	93.46	88.96	99.98
	SP	98.81	96.28	99.97
	OT	97.77	89.65	99.91
Noise	WM	98.98	97.52	99.62
	EM	97.80	95.88	99.80
	SP	95.21	94.85	99.56
	OT	98.70	94.36	98.97
Amplitude (1.2)	WM	100	100	100
	EM	100	100	100
	SP	100	100	100
	OT	100	100	100
Amplitude (1.8)	WM	100	100	100
	EM	100	100	100
	SP	100	100	100
	OT	100	100	100
MP3 (128 kbps)	WM	99.99	100	100
	EM	99.96	100	100
	SP	99.95	100	100
	OT	99.87	100	100
AAC (128 kbps)	WM	99.98	100	100
	EM	99.98	99.98	100
	SP	99.99	100	100
	OT	99.91	100	100
HPF (50 Hz)	WM	100	100	100
	EM	99.97	100	100
	SP	100	100	100
	OT	100	100	100
HPF (100 Hz)	WM	100	100	100
	EM	99.57	100	100
	SP	99.83	100	100
	OT	99.96	100	100

Table 1 (continued)

Attacks	Host signals	DR (%)		
		Method in [15]	Method in [30]	Proposed method
LPF (12 kHz)	WM	100	99.99	100
	EM	100	100	100
	SP	99.87	99.73	100
	OT	100	99.35	100
LPF (8 kHz)	WM	99.96	99.53	99.98
	EM	99.94	99.08	100
	SP	99.84	99.80	100
	OT	99.87	99.16	99.92

DRs of these methods under the above mentioned common attacks. Here, the bit rate of 128 kbps is used for MP3 and AAC attacks. One can see that while all the methods are robust to most of the attacks, the proposed method consistently outperforms the methods in [15, 30]. This result is not surprising. As we have discussed in Section 3, the new method is robust to common attacks. One can also see from Table 1 that there is no clear relationship between the performance of the proposed method and the audio clip categories. For example, the proposed method works better for EM audio clips than for WM audio clips under noise attack but the reverse result is obtained under re-quantization attack. It is important to point out that the method in [15] 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 [15]. However, to our best knowledge, identifying the watermarked segments encountered in [15] 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 [19], 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. The embedding rate of 10bps is utilized for both methods. It can be seen from Table 2 that the proposed method performs better than the method in [19]. Similar to the previous table in Table 2 there is no noticeable performance variation pattern across music categories. 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 [19] cannot be improved by altering any parameter, this method is not suitable for practical applications.

Next, we further evaluate robustness of the proposed method against MP3 and AAC attacks under different bit rates: 64 kbps, 96 kbps, 128 kbps, and 160 kbps. 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 128 kbps or higher, the embedded watermarks are extracted without any error.

Table 2 Detection rates of the proposed method and the method in [19], where ODG = − 1.5 for both methods

Attacks	Host signals	DR (%)	
		Method in [19]	Proposed method
Closed-loop	WM	100	100
	EM	100	100
	SP	100	100
	OT	100	100
Re-quantization	WM	99.76	100
	EM	99.32	100
	SP	99.85	100
	OT	99.64	100
RS (44/22/44)	WM	99.72	100
	EM	97.81	100
	SP	99.92	100
	OT	99.69	100
RS (44/16/44)	WM	99.66	100
	EM	97.31	100
	SP	99.88	100
	OT	99.51	100
Noise	WM	99.59	100
	EM	98.33	100
	SP	98.70	100
	OT	99.42	100
Amplitude (1.2)	WM	100	100
	EM	100	100
	SP	100	100
	OT	100	100
Amplitude (1.8)	WM	100	100
	EM	100	100
	SP	100	100
	OT	100	100
MP3 (128 kbps)	WM	99.75	100
	EM	98.72	100
	SP	100	100
	OT	99.74	100
AAC (128 kbps)	WM	99.70	100
	EM	98.74	100
	SP	99.91	100
	OT	99.65	100
HPF (50 Hz)	WM	99.83	100
	EM	98.81	100
	SP	100	100
	OT	99.95	100
HPF (100 Hz)	WM	99.78	100
	EM	98.57	100
	SP	100	100
	OT	99.72	100

Finally, we compare the computational efficiency of the proposed method with the methods in [15, 19, 30]. We measure the computational efficiency in terms of the running time taken to embed one watermark bit. We carried out the simulation

Table 2 (continued)

Attacks	Host signals	DR (%)	
		Method in [19]	Proposed method
LPF (12 kHz)	WM	99.74	100
	EM	98.68	100
	SP	100	100
	OT	99.76	100
LPF (8 kHz)	WM	99.68	100
	EM	97.46	100
	SP	100	100
	OT	99.25	100

using a notebook computer with Microsoft Windows 7 (64-bit) operating system and MATLAB. Other specifications of the computer include 4 GB RAM and 2.30 GHz Intel Core i7-3610QM CPU. In the simulation, 10000 randomly selected audio segments are used. It can be seen from the Table 3 that the proposed method takes lesser time than the other two methods. So the proposed method is more efficient in computation than the methods in [15, 19, 30].

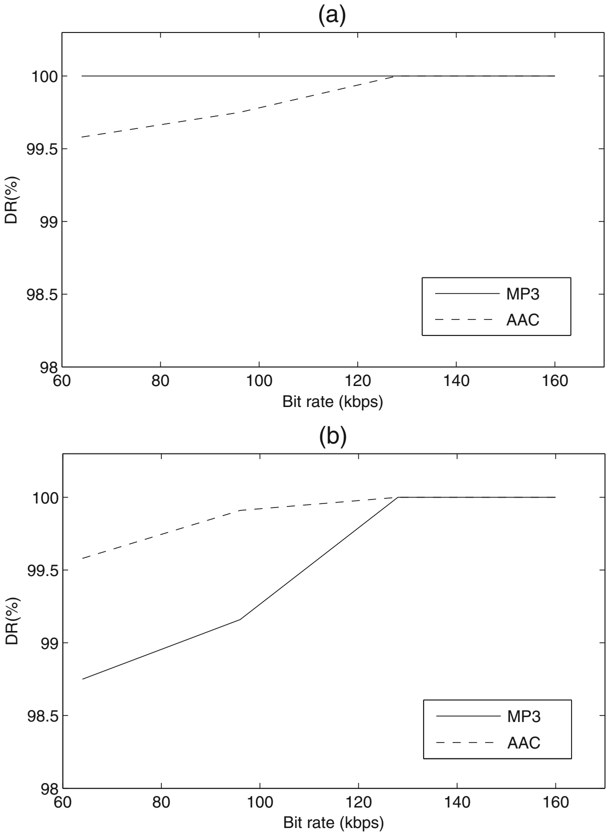

Fig. 2 DRs under MP3 and AAC attacks verses bit rate. **a** Western music **b** Eastern music

Table 3 Running time taken to embed one watermark bit

Methods	Time taken (s)
Method in [15]	3.6849
Method in [30]	0.8324
Method in [19]	0.0129
Proposed method	0.0084

5 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.

Acknowledgements This work was supported in part by the Australian Research Council under grants DP1095498 and DP110102076, the National Basic Research Program of China (973 Program) under grant 2011CB302201, and the National Natural Science Foundation of China under grant 61172180.

References

1. Al-Nuaimy W, El-Bendary MAM, Shafik A, Shawki F, Abou-El-azm AE, El-Fishawy NA, Elhalafawy SM, Diab SM, Sallam BM, El-Samie FEA, Kazemian HB (2011) An SVD audio watermarking approach using chaotic encrypted images. Digital Signal Process 21(6):764–779
2. Arnold M (2000) Audio watermarking: features, applications and algorithm. In: IEEE International conference on multimedia expo 2000, pp 1013–1016
3. Baras C, Moreau N, Dymarski P (2006) Controlling the inaudibility and maximazing the robustness in an audio annotation watermarking system. IEEE Trans Audio Speech Language Process 14(5):1772–1782
4. Bender W, Gruhl D, Morimoto N, Lu A (1996) Techniques for data hiding. IBM Syst J 35(3–4):313–336
5. Bhat KV, Sengupta I, Das A (2010) An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. Digital Signal Process 20(6):1547–1558
6. Cao W, Yan Y, Li S (2009) Bit replacement audio watermarking using stereo signals. In: Int. conf. new trends in information and service, pp 603–606
7. Chen OT-C, Wu W-C (2008) Highly robust, secure, and perceptual-quality echo hiding scheme. IEEE Trans Audio Speech Language Process 16(3):629–638
8. El'Arbi M, Koubaa M, Charfeddine M, Ben Amar C (2011) A dynamic video watermarking algorithm in fast motion areas in the wavelet domain. Multimed Tools Appl 55(3):579–600
9. Erçelebi E, Batakçi L (2009) Audio watermarking scheme based on embedding strategy in low frequency components with a binary image. Digital Signal Process 19(2):265–277

10. Erfani Y, Siahpoush S (2009) Robust audio watermarking using improved TS echo hiding. Digital Signal Process 19(5):809–814
11. Fallahpour M, Megias D (2011) High capacity audio watermarking using the high frequency band of the wavelet domain. Multimed Tools Appl 52(2–3):485–498
12. Foo SW (2008) Audio-watermarking with stereo signals. In: TENCON 2008–2008 IEEE region 10 conf., pp 1–4
13. Huang CH, Chuang SC, Huang YL, Wu JL (2009) Unseen visible watermarking: a novel methodology for auxiliary information delivery via visual contents. IEEE Trans Inf Forensics Security 4(2):193–206
14. Huang H, Yang C, Hsu W (2009) A video watermarking technique based on pseudo-3-D DCT and quantization index modulation. IEEE Trans Inf Forensics Security 5(4):625–637
15. Kalantari NK, Akhaee MA, Ahadi SM, Amindavar H (2009) Robust multiplicative patchwork method for audio watermarking. IEEE Trans Audio Speech Language Process 17(6):1133–1141
16. Kirbiz S, Gunsel B (2006) Robust audio watermark decoding by supervised learning. In: IEEE International conference on acoustics, speech and signal processing, pp 761–764
17. Kirovski D, Malvar H (2001) Robust spread-spectrum audio watermarking. In: IEEE International conference on acoustics, speech and signal processing, pp 1345–1348
18. Ko B-S, Nishimura R, Suzuki Y (2005) Time-spread echo method for digital audio watermarking. IEEE Trans Multimed 7(2):212–221
19. Kondo K, Nakagawa K (2008) A digital watermark for stereo audio signal using variable interchannel delay in high frequency bands. In: Int. conf. intelligent information hiding and multimedia signal process, pp 624–627
20. Lakshmi D, Ganesh R, Marni R, Prakash R, Arulmozhivarman P (2008) SVM based effective watermarking scheme for embedding binary logo and audio signals in images. In: TENCON 2008—2008 IEEE Region 10 conference, pp 1–5
21. Lie W-N, Chang L-C (2006) Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. IEEE Trans Multimed 8(1):46–59
22. Lin WH, Horng SJ, Kao TW, Fan P, Lee CL, Pan Y (2008) An efficient watermarking method based on significant difference of wavelet coefficient quantization. IEEE Trans Multimed 10(5):746–757
23. Lin WH, Horng SJ, Kao TW, Chen RJ, Chen YH, Lee CL, Terano T (2009) Image copyright protection with forward error correction. Expert Syst Appl 36(9):11888–11894
24. Lin WH, Wang YR, Horng SJ (2009) A wavelet-tree-based watermarking method using distance vector of binary cluster. Expert Syst Appl 36(6):9869–9878
25. Lin WH, Wang YR, Horng SJ, Pan Y (2009) A blind watermarking method using maximum wavelet coefficient quantization. Expert Syst Appl 36(9):11509–11516
26. Liu CS (2005) Multimedia security: steganography and digital watermarking techniques for protection of intellectual property. Idea Group Publishing, Hershey, PA
27. Luo L, Chen Z, Chen M, Zeng X, Xiong Z (2010) Reversible image watermarking using interpolation technique. IEEE Trans Inf Forensics Security 5(1):187-Ű193
28. Malvar HS, Florencio D (2003) Improved spread spectrum: a new modulation technique for robust watermarking. IEEE Trans Signal Process 52(4):898–905
29. Mohammad AA (2012) A new digital image watermarking scheme based on Schur decomposition. Multimed Tools Appl 59(3):851–883
30. Natgunanathan I, Xiang Y, Rong Y, Zhou W, Guo S (2012) Robust patchwork-based embedding and decoding scheme for digital audio watermarking. IEEE Trans Audio Speech Language Process 20(8):2232–2239
31. Oppenheim AV, Schaffer RW (1998) Discrete-time signal processing. Prentice-Hall, Upper Saddle River, NJ
32. Peng D, Xiang Y (2010) Underdetermined blind separation of non-sparse sources using spatial time-frequency distributions. Digital Signal Process 20(2): 581–596
33. Rec. B.S. 1387 (2001) Methods for objective measurements of perceived audio quality. Rec. B.S. 1387, Int. Telecomm. Union, Geneva, Switzerland
34. Rosiyadi D, Horng SJ, Fan P, Wang X, Khan M, Pan Y (2012) An efficient copyright protection scheme for e-government document images. IEEE Multimedia 19(3):62–73
35. Singh J, Garg P, De AN (2012) Audio watermarking based on quantization index modulation using combined perceptual masking. Multimed Tools Appl 59(3):921–939
36. Solanki K, Jacobsen N, Madhow U, Manjunath BS, Chandrasekaran S (2004) Robust image-adaptive data hiding based on erasure and error correction. IEEE Trans Image Process 13(12):1627–1639

37. Takahashi A, Nishimura R, Suzuki Y (2005) Multiple watermarks for stereo audio signals using phase-modulation techniques. IEEE Trans Signal Process 53(2):806–815
38. Valizadeh A, Wang ZJ (2011) Correlation-and-bit-aware spread spectrum embedding for data hiding. IEEE Trans Inf Forensics Security 6(2):267–282
39. Wang X, Qi W, Niu P (2007) A new adaptive digital audio watermarking based on support vector regression. IEEE Trans Audio Speech Language Process 15(8):2270–2277
40. Wang XY, Zhao H (2006) A novel synchronization invariant digital watermarking scheme based on DWT and DCT. IEEE Trans Signal Process 54(12):4835–4840
41. Wei L, Wei S, Hongtao H (2012) Novel robust image watermarking based on subsampling and DWT. Multimed Tools Appl 60(1):31–46
42. Wu SQ, Huang JW, Shi YQ (2005) Efficiently self-synchronized audio watermarking for assured audio data transmission. IEEE Trans Broadcast 51(1):69–76
43. Xiang Y (2008) Blind source separation based on constant modulus criterion and signal mutual information. Comput Electric Eng 34(5):416–422
44. Xiang Y, Peng D, Natgunanathan I, Zhou W (2011) Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo based audio watermarking. IEEE Trans Multimed 13(1):2–13
45. Xiang Y, Natgunanathan I, Peng D, Zhou W, Yu S (2012) A dual-channel time-spread echo method for audio watermarking. IEEE Trans Inf Forensics Security 7(2):383–392
46. Xiang Y, Peng D, Xiang Y, Guo S (2013) Novel Z-domain precoding method for blind separation of spatially correlated signals. IEEE Trans Neural Netw Learn Syst 24(1):94–105
47. Yeo IK, Kim HJ (2003) Modified patchwork algorithm: a novel audio watermarking scheme. IEEE Trans Speech Audio Process 11(4):381–386

Iynkaran Natgunanathan received the BSc Eng. (Hons) degree in electronics and telecommunication engineering from the University of Moratuwa, Katubedda, Moratuwa, Sri Lanka, in 2007. In 2012, he received the PhD degree from Deakin University, Australia.

From 2006 to 2008, he was a Software Engineer at Millennium Information technology (Pvt) Ltd, Malabe, Sri Lanka. Currently, he is a research associate at the School of Information Technology, Deakin University. His research interests include digital watermarking, psychoacoustics, audio/speech/image processing, telecommunication and robotics.

Yong Xiang received the BE and ME degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1983 and 1989, respectively. In 2003, he received the PhD degree from The University of Melbourne, Melbourne, Australia.

He was with the Southwest Institute of Electronic Equipment of China, Chengdu, from 1983 to 1986. In 1989, he joined the University of Electronic Science and Technology of China, where he was a Lecturer from 1989 to 1992 and an Associate Professor from 1992 to 1997. He was a Senior Communications Engineer with Bandspeed Inc., Melbourne, Australia, from 2000 to 2002. He is currently an Associate Professor with the School of Information Technology at Deakin University, Australia. His research interests include blind signal/system estimation, communication signal processing, biomedical signal processing, information and network security, and speech and image processing. He is a senior member of the IEEE.

Yue Rong received the BE degree from Shanghai Jiao Tong University, China, the MSc degree from the University of Duisburg-Essen, Duisburg, Germany, and the PhD degree (summa cum laude) from Darmstadt University of Technology, Darmstadt, Germany, all in electrical engineering, in 1999, 2002, and 2005, respectively.

From April 2001 to October 2001, he was a research assistant at the Fraunhofer Institute of Microelectronic Circuits and Systems, Duisburg, Germany. From October 2001 to March 2002, he was with Nokia Ltd., Bochum, Germany. From November 2002 to March 2005, he was a Research Associate at the Department of Communication Systems in the University of Duisburg-Essen. From April 2005 to January 2006, he was with the Institute of Telecommunications at

Darmstadt University of Technology, as a Research Associate. From February 2006 to November 2007, he was a Postdoctoral Researcher with the Department of Electrical Engineering, University of California, Riverside. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University of Technology, Perth, Australia, where he is now a Senior Lecturer. His research interests include signal processing for communications, wireless communications, wireless networks, applications of linear algebra and optimization methods, and statistical and array signal processing.

Dr. Rong received the Best Paper Award at the Third International Conference on Wireless Communications and Signal Processing, Nanjing, China, 2011, the Best Paper Award at the 16th Asia-Pacific Conference on Communications, Auckland, New Zealand, 2010, the 2010 Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University, the 2004 Chinese Government Award for Outstanding Self-Financed Students Abroad (China), and the 2001–2002 DAAD/ABB Graduate Sponsoring Asia Fellowship (Germany). He has co-authored more than 70 referred IEEE journal and conference papers. He is an Editor of IEEE Wireless Communications Letters, a Guest Editor of the IEEE JSAC special issue on Theories and Methods for Advanced Wireless Relays to be published in 2012, a Guest Editor of the EURASIP JASP Special Issue on Signal Processing Methods for Diversity and Its Applications to be published in 2012, and has served as a TPC member for IEEE ICC, IEEE ICCC, WCSP, IWCMC, and ChinaCom.

Dezhong Peng received the BS degree in applied mathematics, the MS and PhD degrees in computer software and theory from the University of Electronic Science and Technology of China, Chengdu, China, in 1998, 2001, 2006, respectively.

From 2001 to 2007, he was with the University of Electronic Science and Technology of China as an assistant lecturer and a lecturer. He was a Postdoctoral Research Fellow at the School of Engineering, Deakin University, Australia from 2007 to 2009. Currently, he is a Professor at the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China. His research interests include blind signal processing and neural networks. He is a member of the IEEE.