Signal Processing 91 (2011) 1693–1708

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# A robust audio watermarking scheme based on reduced singular value decomposition and distortion removal

Jian Wang*, Ron Healy, Joe Timoney

*Department of Computer Science, NUI Maynooth, County Kildare, Ireland*

ARTICLE INFO

*Article history:*
Received 26 July 2010
Received in revised form
13 January 2011
Accepted 17 January 2011
Available online 4 February 2011

*Keywords:*
Audio watermarking
Singular value decomposition
MP3
Distortion removal
Psychoacoustics

ABSTRACT

This paper presents a blind audio watermarking algorithm based on the reduced singular value decomposition (RSVD). A new observation on one of the resulting unitary matrices is uncovered. The proposed scheme manipulates coefficients based on this observation in order to embed watermark bits. To preserve audio fidelity a threshold-based distortion control technique is applied and this is further supplemented by distortion suppression utilizing psychoacoustic principles. Test results on real music signals show that this watermarking scheme is in the range of imperceptibility for human hearing, is accurate and also robust against MP3 compression at various bit rates as well as other selected attacks. The data payload is comparatively high compared to existing audio watermarking schemes.



## 1. Introduction

As audio, video and other works have become more widely available in digital form in recent years, the ease with which perfect copies can be made has increased the impact of unauthorized copying to such an extent that it is a major concern of the entertainment industry. The Recorded Industry Association of America (RIAA) states that the cost to the music industry alone is in the region of $12.5 billion annually [5]. It is obvious therefore that the issue undermines the music and film publishing industries. Concerns over protecting copyright have triggered significant research in, amongst other techniques, attempts to hide copyright messages in digital media. Examples of various techniques that have been developed are described in [1,11].

Digital watermarking is a technique whereby data is hidden in a cover or host signal as a form of steganography. Depending on whether the decoding phase of a watermarking scheme requires access to the original signal and/or the actual watermark in order to successfully decode the embedded data, the scheme may be described as 'informed' or 'blind'. Informed decoding requires the original cover or the actual watermark. This presents many limitations for the scheme [4]. Blind decoding [12] does not require any information except private key in order to decode the watermark from the candidate signal. A watermarking scheme that provides for blind decoding is more difficult to achieve but offers a wider variety of potential applications of the scheme.

Depending on the watermark application and purpose, different requirements arise resulting in various design issues. Imperceptibility, or perceptual transparency, of the watermark is a general requirement independent of the application domain [11]. Artefacts introduced through a watermarking process are not only annoying and undesirable but may also reduce or destroy the commercial value of the watermarked data. Robustness, on the other hand, has to be considered in many application domains to prevent potential removal of watermark information by deliberate or accidental attacks on the watermarked signal. For some applications, the accidental attacks that would be

* Corresponding author.
*E-mail address:* jwang@cs.nuim.ie (J. Wang).

0165-1684/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2011.01.014

expected to be applied to watermarked signals would be known. As far as music is concerned, one of the most common attacks is compression. One very common technique is MP3 compression and surviving MP3 compression has therefore become an essential requirement for a watermarking scheme in the music domain [13]. This is due to the fact that MPEG-1 Layer III (MP3) and its variants have become the *de facto* standard for transmission and storage of compressed audio for both World Wide Web (WWW) and portable media applications [13]. The difficulty when designing a watermark that can withstand MP3 compression is related to the fact that the MP3 algorithm nonlinearly modifies the spectrum based on a psychoacoustic model and the expected bit rate. Thus, any watermark embedded among the audio's spectral components risks being lost after compression. Furthermore, the audio watermarking scheme should be robust against some other common attacks such as filtering [2,4,11].

Most audio watermarking algorithms use either a time domain [6,7] or frequency domain[8–10] masking property to embed a watermark [3,4]. Our previous work has considered a watermarking scheme utilizing the CSPE algorithm to encode the cover signal [60]. Due to the high precision of CSPE in estimating frequency components and thus through manipulating the identified components by CSPE, this scheme [60] achieved an almost 100% precision and extremely high perceptual transparency. However, the approach was found to be unsuitable for audio files that were subject to compression, as the watermark was sometimes lost during the process.

Recently, the singular value decomposition (SVD) has been used extensively as an effective technique in digital watermarking [14–18,22–34]. Most existing SVD based watermarking techniques are applied in images [16–18, 22–34]. SVD based audio watermarking techniques, also exist but fewer, some of which can be found in [14,15]. All these SVD based watermarking algorithms can be categorized into two groups.

The first group of SVD based watermarking algorithms are 'informed' [32–34], requiring access to the original signal or the watermark in order to successfully decode the embedded watermark. The scheme proposed in [32] can be used as a typical example to illustrate the idea used for this group. In the embedding stage, a cover image $A$ will be decomposed into three matrices: $U$, $S$ and $V$ by SVD. The watermark information $W$ then will be linearly added to $S$, resulting in a new matrix $S'$, then inverse SVD will be applied on $S'$, $U$ and $V$ to get a watermarked image $A_w$. In the extraction stage, $U^T$ and $V$ will be used to multiply with $A_w$ to get $S'$. Then the watermark $W$ can be extracted out by a linear subtraction between $S'$ and $S$.

The second group SVD based schemes are 'blind' and embedding watermark information by manipulating the coefficients in the SVD decomposed matrices, such as $U$ or $S$. These schemes are based on some observations proposed in [14–18]. Particularly, a digital image watermarking algorithm presented in [16] is based on two observations. Firstly, column-wise modification of the elements of the $U$ matrix will cause less visible distortion than modifying the elements row-wise. Secondly, row-wise modification of the elements of $V^T$ will cause less visible distortion than modifying the elements column-wise. Another particular SVD based digital image watermarking algorithm proposed in [17], is formulated on the observation that the elements in the first column of $U$ and $V$ can be modified without significantly impacting signal integrity, allowing them to survive common attacks including compression. In [14,15], the proposed audio watermarking algorithm is based on the observation that changing $S$ slightly does not affect the quality of the signal much and that the singular values in $S$ are consistent under common signal processing operations. In [18], an image watermarking algorithm based on tuning the coefficients in $U$ was proposed.

Developing robust watermarking techniques for digital audio signals is relatively difficult compared to watermarking digital images. This is due to the high sensitivity of the human ear over a large dynamic range in comparison to the human eye [4,21]. Therefore, alterations have to be made to the SVD based image watermarking approach before it can be applied to audio. Compared with manipulating $U$, there are three deficiencies of manipulating $S$ [14,15]. The first is that the modification of the largest coefficients in $S$ [14] would cause a greater degradation [18] and it can then be easily detected and/or destroyed as the embedding position is public. The second is that embedding watermark information based on the modification of the less significant coefficients in $S$ risks their loss after compression, which is based on removing less significant singular coefficients in $S$ [19,20]. Finally, only a small number of singular elements in $S$ are available to manipulate, compared with $U$, which makes it hard to achieve a high data payload [18]. No SVD based audio scheme, which achieves watermarking through manipulating $U$ to embed information, has been proposed yet.

A new observation that the peaks in the second column of $U$ are consistent after different attacks is proposed in this paper. Compared with manipulating the first column of $U$, modification of the second column introduces much less audible distortion. The proposed watermarking algorithm is based on this new observation. We investigated embedding on other columns and found that the robustness was adversely affected for no apparent gain. As mentioned above, the human auditory system is more sensitive than the human visual system. We therefore need to resolve any specific perceptual issues that arise. Specifically, distortion control is used to control the audible distortions and determine the strength of peaks created to represent watermark information in the embedding stage. Therefore, this scheme is adaptive to specific signal characteristics. Furthermore, psychoacoustic techniques are utilized to further suppress any resulting audible distortion. To the best of our knowledge, control of perceptual distortion has not been used in any SVD based audio watermarking schemes previously. The following sections details how these are achieved.

## 2. SVD

The SVD is a well-known numerical analysis tool used on matrices. If $A$ is an arbitrary $m$-by-$n$ matrix, with the full SVD, it can be decomposed as

$$A = USV^T \tag{1}$$

where $U$ is a $m$-by-$m$ unitary matrix, $S$ is a $m$-by-$n$ diagonal matrix with non-negative elements and $V$ is a $n$-by-$n$ unitary matrix. The reduced singular value decomposition (RSVD) is a more compact representation and, compared with the full SVD, this version is more computationally efficient when $n \ll m$ or the rank of matrix $r \ll n$ [31]. With RSVD, $A$ can be decomposed as Eq. (2) where $U$ and $V$ are $p \times r$ and $r \times q$ unitary matrices, respectively, and $S$ is a $r \times r$ diagonal matrix with positive elements

$$A = USV^T = \begin{bmatrix} u_{1,1} & \cdots & u_{1,r} \\ \vdots & \ddots & \vdots \\ u_{m,1} & \cdots & u_{m,r} \end{bmatrix} \begin{bmatrix} s_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_{r,r} \end{bmatrix} \begin{bmatrix} v_{1,1} & \cdots & v_{1,r} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,r} \end{bmatrix}^{\mathrm{T}}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{r} u_{i,k} * s_{k,k} * v_{k,j} \quad (2)$$

## 3. Methodology

### 3.1. Watermark embedding

The basic steps involved in embedding watermark using RSVD include:

(a) Decompose the signal into frames and organize each frame′s spectral magnitudes into matrix $A$.
(b) Apply RSVD to $A$ by Eq. (2) and thus $U$ is obtained.
(c) Embedding watermark information in $U$ by utilizing a spectrum distortion control and minimization process.
(d) Perceptual distortion suppression.
(e) Reconstructing the time-domain watermarked signal by using a specifically designed overlap-add procedure.

All these steps will be detailed in the following sections.

#### 3.1.1. Organization of matrix A

Before applying RSVD to an audio signal, the issue of how to organize the data into matrix form has to be addressed. The audio signal should firstly be split into frames whose length is denoted as $L$. The magnitude spectrum of the signal in each frame can then be computed. This spectrum will contain $L/2$ frequency bins below the Nyquist frequency. In this paper, the RSVD input matrix is organized by putting all the spectral components from one frame into a single matrix. For example, with a frame length of 1024 we use only the first 512 magnitude values and transform them into a $64 \times 8$ matrix $A$. The magnitudes of the first 64 frequency bins will be put into the first column, the next 64 bin magnitudes into the second column and so on. The values in the first column will normally, but not always, be more significant than those in subsequent columns. Then RSVD will be applied to $A$ and thus derives three matrices $U$, $S$ and $V$.

#### 3.1.2. Using the second column of U to embed watermark

As in Eq. (2), $s_{1,1}$ is the largest value in $S$ and, when considering the product, it is only multiplied with the column elements $u_{1\ldots m,1}$, which results in the most significant peaks within the spectrum $A$. Even a minor alteration of any element in $u_{1\ldots m,1}$ may therefore result in significant distortion when the time signal is resynthesised. To overcome this obstacle the following process was devised. Similarly to the point made about $s_{1,1}$ in (2), so $s_{2,2}$ is the second largest value in the $S$ matrix and, when expanding the product, it is multiplied with $u_{1\ldots m,2}$, which is more significant than the coefficients in all other columns except $u_{1\ldots m,1}$. We found that the peaks in $u_{1\ldots m,2}$ are almost always retained after compression. Modifying the elements of $u_{1\ldots m,2}$ rather than those in $u_{1\ldots m,1}$ will therefore result in less distortion in the reconstructed signal as $s_{2,2}$ is generally much smaller than $s_{1,1}$. Thus, we choose to create a local peak in $u_{1\ldots m,2}$ to represent the watermark data. The criterion of embedding a watermark is detailed as below.

Assume we have a binary stream representing the watermark information. At the embedding stage, if the bit to be embedded is a '1', we increase the value of the element $u_{t,2}$, where $t$ is a user-defined key value. The surrounding elements at $u_{t-1,2}$ and $u_{t+1,2}$ are reduced to a negligible low value QUOTE such that $u'_{t,2}$ is a peak value, where $u'_{t,2}$ represents the modified $u_{t,2}$.

Similarly, if the bit to be embedded is a '0', we increase the value of the element $u_{t+1,2}$ and the surrounding elements at $u_{t,2}$ and $u_{t+2,2}$ are reduced to the same negligible low value $T_m$, such that $u'_{t+1,2}$ is a peak value. At the extraction stage, we compare $u'_{t,2}$ and $u'_{t+1,2}$. If $u'_{t,2} > u'_{t+1,2}$, then the bit is decoded as '1', otherwise it is a '0'.

This embedding criterion also can be formulated by

$$\begin{cases} bit = 1 \xrightarrow{\text{yields}} u'_{t,2} > u'_{t-1,2} \text{ and } u'_{t,2} > u'_{t+1,2}, \text{ where } u'_{t-1,2} = T_m \text{ and } u'_{t+1,2} = T_m \\ bit = 0 \xrightarrow{\text{yields}} u'_{t+1,2} > u'_{t,2} \text{ and } u'_{t+1,2} > u'_{t+2,2}, \text{ where } u'_{t,2} = T_m \text{ and } u'_{t+2,2} = T_m \end{cases} \quad (3)$$

We use Fig. 1 as an example to demonstrate the robustness of the created peaks in the second column of $U$ after MP3 compression at 64 kbps. In this example, we created 4 peaks around positions 11, 25, 46 and 61 of the second column of $U$. More specifically, we make $u_{11,2}$, $u_{25,2}$, $u_{46,2}$, and $u_{61,2}$ local peaks, and their respective neighboring elements ($u_{10,2}$, $u_{12,2}$), ($u_{24,2}$, $u_{26,2}$), ($u_{45,2}$, $u_{47,2}$), ($u_{60,2}$, $u_{62,2}$) equal to QUOTE . To ensure good visibility of the figure, we only display the relevant regions. The first panel in Fig. 1 shows the position ranges from 8 to 27 and the second panel shows the position ranges from 42 to 62. The third panel shows the comparison in magnitude between original signal and modified signal. From the first and second panel, we can see that all 4 created peaks have been retained after MP3 compression at 64 kbps, which are marked with arrows. It is worth noting from the third panel that only negligible distortion was introduced, which is the advantage of using the second column of $U$ instead of the first column $U$.

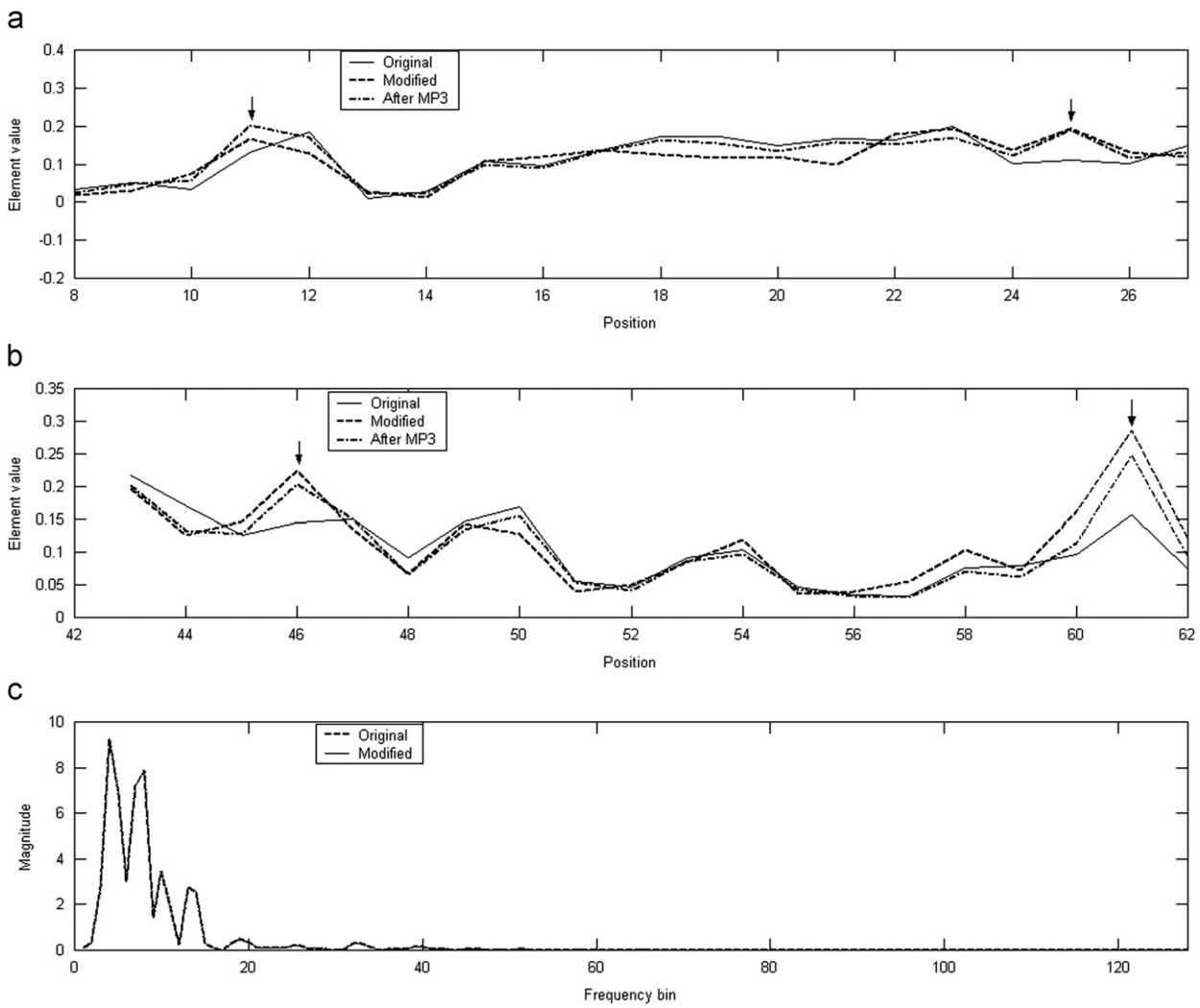


**Fig. 1.** (a) Robustness of the created peaks in second column of $U$ after 64 kbps MP3 compression (in the first region), (b) robustness of the created peaks in second column of $U$ after 64 kbps MP3 compression (in the second region) and (c) spectrum difference between the original signal and the modified signal.

### 3.1.3. Determine peak value with spectrum distortion control

The extent of the peak value introduced is an issue that requires careful evaluation. If it is big, it will be robust under attacks but may result in perceptual distortion in the reconstructed audio. Likewise, if it is small, it may not survive attacks but will not be perceived. Generally, a trade-off among the accuracy, robustness and perceptual transparency exists in digital watermarking [11]. We will address the issue of which value should be assigned to $u'_{t,2}$ if embedding a '1' bit.

The modification of any element of $u_{1\ldots m,2}$ must be tailored so that after the reconstructed audio is attacked this modification will persist. Take compression attack as an example. When any signal undergoes the compression process, not every spectral component with a large magnitude is guaranteed to survive it. This is because a component could be masked by other components of larger magnitude that exist in the same critical band. Only those components whose magnitude is strongest in their own critical band are likely to survive the process, even if they are comparatively low in the entire spectrum. Thus, it is best to set a value for the peak of $u'_{t,2}$ that will create a local maximum, within a particular range, in the spectrum of the reconstructed signal. The method of assigning a value to $u'_{t,2}$ will be illustrated as below.

By modifying elements in the second column of $U$ which indirectly alter the magnitudes of some components, distortion – whether audible or inaudible – will be introduced. We use the following equations to quantify the spectrum distortions:

$$dist = \sum_{i=t-3}^{i=t+3} \left|(A'(i,k) - A(i,k))\right| \tag{4}$$

where *dist* is the local absolute distortion in magnitude that has been introduced by modifying $u_{t,2}$, $k$ is the column index of Matrices $A$ and $A'$, the procedure for determining $k$

value will be detailed in Section 3.1.4

$$distr = \frac{\sum_{i=t-3}^{i=t+3} |A'(i,k) - A(i,k)|}{\sum_{i=t-3}^{i=t+3} A(i,k)} = \frac{dist}{\sum_{i=t-3}^{i=t+3} A(i,k)} \quad (5)$$

where *distr* is the ratio of *dist* to the total original surrounding spectrum magnitudes.

These two measures are used together to control distortion. If *dist* is large but is comparatively small, then this will not introduce audible distortion. Conversely, if *dist* is large and *dist* is comparatively small, this could introduce audible distortion. We note that, if *dist* is below a threshold, the distortion introduced is very hard to detect even though *distr* is comparatively large. This can be explained by the simultaneous masking principal: whenever two or more stimuli in the frequency domain are simultaneously presented to the auditory system, the presence of some spectral energy will mask the presence of other spectral energy. As a result, if the introduced distortion is low, compared to surrounding energy, it would be masked and thus inaudible.

Thus, an iterative algorithm applied to each frame of the signal, based on the above description to determine the peak value of $u'_{t,2}$ and control distortion is described as following:

(a) For each embedded position $t$ in each frame;
(b) Set $u'_{t,2}$=0.5, $u'_{t-1,2}$=0, $u'_{t+1,2}$=0 (in this scheme, we set QUOTE =0);
(c) Use the inverse RSVD to derive modified spectrum magnitudes;
(d) Calculate *dist*, *distr* using Eqs. (4) and (5);
(e) If $dist < \delta_1$ or $distr < \delta_2$, where $\delta_1$and $\delta_2$ are two threshold values, select next embedding position and go to step (a); else, $u'_{t,2} = u'_{t,2} * 0.8$, go to step (c).

The process used to determine the above two threshold values is described as following: setting $\delta_1$=20, $\delta_2$=1 initially, in the inner-loop $\delta_2$ is reduced by 0.1 every time; in the outer-loop $\delta_1$ is reduced by 1 every time. Then the watermark information was embedded and extracted in each loop, followed by calculation of the Objective Difference Grade (ODG) score and the Precision evaluation. Note that the ODG and the Precision details will be outlined in Sections 4.1.1 and 4.2, respectively, but that an ODG score close to zero indicates that a watermarked file is perceptually identical to its unwatermarked counterpart, and a precision close to 100% indicates that all the embedded watermark information is correctly recovered. By experimenting on 25 signals, we found that when $\delta_1$=5, $\delta_2$=0.3, it produces a comparatively satisfactory compromise between the ODG score and the Precision.

### 3.1.4. Minimize spectrum distortion

Following the assignment of peak value, we take a further step to minimize spectrum distortion without significantly affecting accuracy. Consider the embedding of a '1' bit: according to Eq. (6), all values in row $t$ of the reconstructed signal spectrum matrix $A'_{t,1\ldots q}$ are affected by the manipulation of $u_{t,2}$ and thus all of them will be distorted to a greater or lesser extent. An expression for the row $t$ of $A'$ is

$$A'_{t,j} = \sum_{i=1}^{r} u'_{t,i} * s_{i,i} * v_{i,j}, \quad \text{where } 1 \le j \le n \quad (6)$$

From (6), we can calculate $k$ by

$$k = \underset{j}{\operatorname{argmax}} \left( \frac{u'_{t,2} * s_{2,2} * v_{2,j}}{A'_{t,j}} \right), \quad \text{where } 1 \le j \le n \quad (7)$$

$k$ is the column index with which the ratio of $u'_{t,2} * s_{2,2} * v_{2,k}$ to $A'_{t,k}$ is the maximum. Thus we can say that $u'_{t,2}$ is more strongly correlated with $A'_{t,k}$. Then, we can eliminate the distortion to each component of $A'_{t,1\ldots n}$, except $A'_{t,k}$, by returning their magnitudes back to their original values given in $A_{t,1\ldots n}$. In doing so, the level of distortion in the reconstructed audio signal is reduced with minimal impact on the values of $u'_{t,2}$.

Here we will use a typical example to show the spectrum distortion been introduced after performing the steps stated in Section 3.1.3. As we embed more than one bit into each frame by altering more than one element of $U$, thus, we use the following equations to evaluate the total distortion introduced to each frame:

$$dist_f = \sum_{i=1}^{i=64} \sum_{k} |(A'(i,k) - A(i,k))| \quad (8)$$

where $dist_f$ QUOTE is the total distortion introduced in spectrum magnitudes after embedding the watermark in one frame

$$distr_f = \frac{\sum_{i=1}^{i=64} \sum_{k} |(A'(i,k) - A(i,k))|}{\sum_{i=1}^{i=64} \sum_{k} |A(i,k)|} = \frac{dist_f}{\sum_{i=1}^{i=64} \sum_{k} |A(i,k)|} \quad (9)$$

where $distr_f$ is the ratio of $dist_f$ to the total original spectrum magnitudes. $k$ is determined by Eq. (7). Note that for different embedding position $t$, $k$ can be different.

A typical music file was used to illustrate the distortion introduced by modifying the elements in the second column of $U$ as shown in Fig. 2. As a comparison, the distortion introduced by modifying the elements in the first column of $U$ is also shown in Fig. 2, to verify that modifying the second column of $U$ introduces less distortion than modifying the first column of $U$, as stated in Section 3.1.2.

In Fig. 2(a) and (b) panel, $x$-axis denotes the value of $dist_f$, and $y$-axis denotes the number of frames. In Fig. 2(c) and (d) panel, $x$-axis denotes the value of $distr_f$, and $y$-axis denotes the number of frames.

The lower that $dist_f$ and $distr_f$ are, the more perceptually transparent the corresponding frames are. From panel (a), we can see that most $dist_f$ are less than 5, while in panel (b), $dist_f$ distribution ranges from 0 to 100. Similarly, most $distr_f$ in panel (c) are less than 0.5, while in panel (d), most $distr_f$ are above 0.5. Thus, embedding the watermark information by modifying the first column of $U$ introduces much more spectrum distortion than modifying the second column of $U$. However, from (a) and (c) of Fig. 2, we can see there are some frames whose $dist_f$ are above 20 and also some frames whose $distr_f$ are above 0.6. Since $dist_f$ and $distr_f$ cannot be exactly mapped to subjective perception, instead of further lowering of the threshold value set in the above iterative algorithm which may not achieve sufficiently good

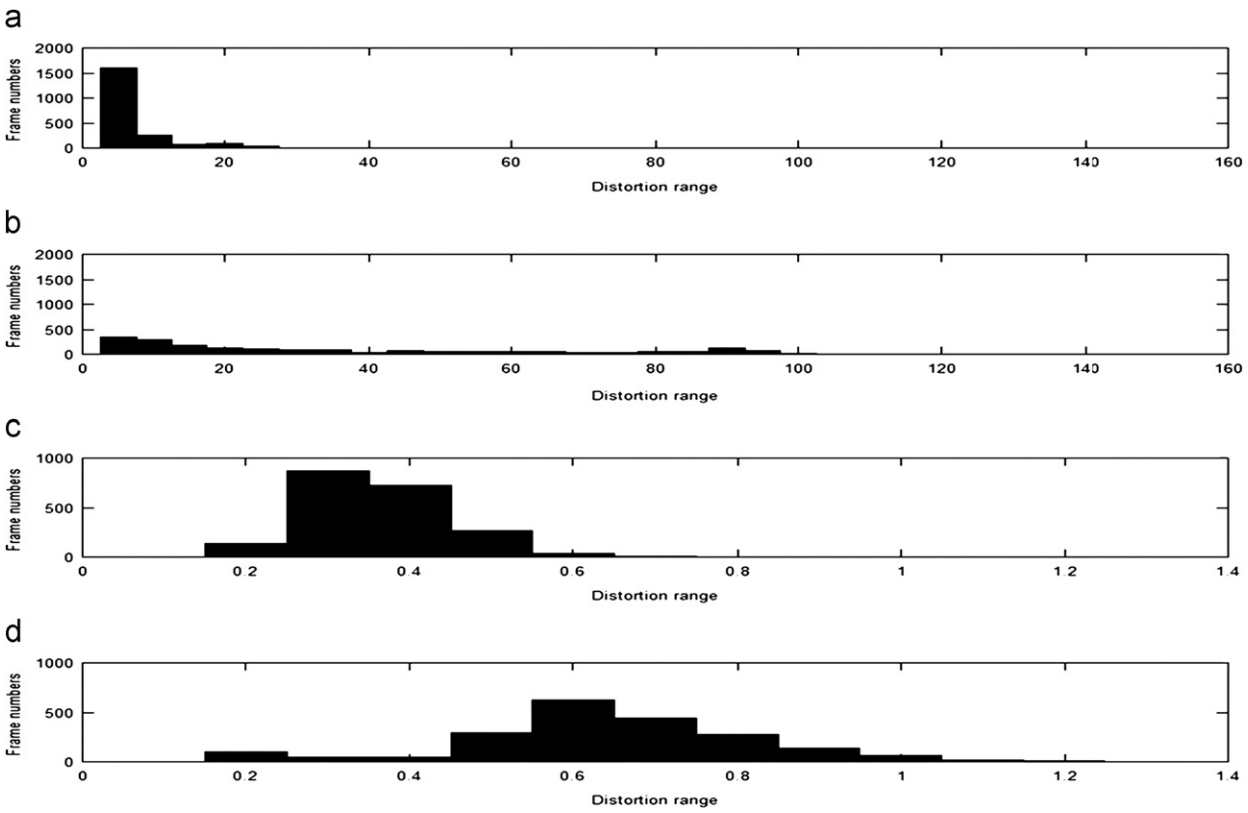


**Fig. 2.** (a) The $dist_f$ distribution of the example music by modifying the second column of $U$, (b) $dist_f$ distribution of the example music by modifying the first column of $U$, (c) $distr_f$ distribution of the example music by modifying the second column of $U$ and (d) $distr_f$ distribution of the example music by modifying the first column of $U$

imperceptibility, we should use a psychoacoustic approach to check if these distortions are audible or not. If audible, then further distortion suppression will be performed.

*3.1.5. Audible distortion detection and further suppression*

Before taking a psychoacoustic approach, subjective listening tests were carried out on watermarked signals using the distortion control algorithm outlined in Section 3.1.4. The listening result showed that imperceptibility was almost acceptable except for a very small number of audible artefacts appearing in few music signals. Thus, we decided to remove these artefacts to achieve perceptual transparency independent of music style or genre. The psychoacoustic model [13] was used as a means to evaluate the perceptibility of the introduced distortion and, by extension, to detect any audible artefacts remaining.

In psychoacoustics, the noise-to-mask ratio (*NMR*) has been used as a measurement of audio quality in many different applications [36,37]. It also has been included into recommendation BS.1387 for perceptual evaluation of audio quality (PEAQ) [38]. This section will firstly give a concise introduction to the steps involved in deriving the *NMR*, which will then be adapted to our presented watermarking algorithm. The steps involved are [13]:

(a) Spectral analysis to derive power spectral density (PSD);
(b) Identify tonal and noise maskers;
(c) Decimation and reorganization of maskers to discard those inaudible maskers;
(d) Obtaining individual masking thresholds;
(e) Obtaining global masking threshold;
(f) Derive the minimum masking threshold $LT_{min}$;
(g) Finally, computing the *NMR* as follows:

$$NMR(i) = SMR(i)\text{-}SNR(i) \tag{10}$$

where $i$ denotes sub-band number, signal-to-mask ratio (*SMR*) is defined for each sub-band as the ratio of signal energy to $LT_{min}$, *SNR* is defined for each sub-band as the ratio of signal energy to noise energy.

A negative *NMR* indicates that the noise energy is below $LT_{min}$. Conversely, a positive *NMR* indicates that the noise energy is above $LT_{min}$. In order to minimize audible distortions, the corresponding *NMR* should be at least less than or equal to 0 dB [65]. As far as our watermarking scheme is concerned, the *NMR* is calculated as

$$E_{ns}(i) = \max_{j=1:n} |E_o(i,j) - E_w(i,j)| \tag{11}$$

$$NMR(i) = E_{ns}(i) - LT_{min}(i) \tag{12}$$

where $i$ denotes sub-band number, $j$ denotes frequency bin number in each sub-band, $LT_{min}$ is the minimum masking threshold energy in dB, QUOTE is the noise energy in dB, QUOTE is the original signal energy in dB

and QUOTE is the watermarked signal energy in dB. The $i$th sub-band noise energy is determined by the largest noise introduced into each frequency bin of this sub-band.

All these values can be obtained using an implementation of the psychoacoustic model [39]. Based on the experiment performed in [37], it is known that by ensuring the *NMR* is less than $-5$ dB, distortion introduced will be inaudible. More specifically, if the *NMR* exceeds $-5$ dB for a given frame, it indicates that the distortion may be audible and should be further reduced. Otherwise, no further suppression is required. The algorithm to identify and suppress these distortions is described as follows:

(1) Calculate the *NMR* for each frame.
(2) Obtain position $(i,k)$, representing the sub-band number and the bin number, respectively, from the frame where the corresponding *NMR* is great than $-5$ dB. If no position $(i,k)$ is found in the frame that satisfies this, repeat with the next frame from step (1); else, go to step (3).
(3) In this case, the magnitude of the corresponding component in position $(i,k)$ is tuned according to the following equations:

$$E'_w(i,k) = E_o(i,k) - LT_{min}(i) + a \tag{13}$$

$$A'_w(i,k) = 10^{1/20*(E_w(i,k)+20*\log_{10}(FFT\_SIZE))} \tag{14}$$

where $E'_w$ is the tuned signal energy, $A'_w$ is the tuned signal spectrum, *FFT_SIZE* is the frame size, $a$ represents the NMR threshold. In our implementation, we set $a$ equal to 5 dB meaning that only an NMR value that is equal to or less than $-5$ dB is acceptable so that distortion is inaudible.

We used 10 signals by way of an example to show the effect of including the psychoacoustic distortion suppression algorithm. The results are plotted in Fig. 3.

From Fig. 3(a), we can see that the ODG score of each music file has been improved, and the average improvement is 0.38. However, the detection precision of each music file has been decreased, and the average decrease is 0.035, as shown in Fig. 3(b). It is worth noting that this added procedure results in an increased computational cost at the embedding stage. The decision to include this can be made dependent on the application domain to which the watermarking technique is being applied. If perceptual transparency is a primary concern (for example, in copyright protection), then the perceptual distortion suppression would outweigh any reduction in the watermark detection precision and the additional computational cost of embedding.

### 3.1.6. Reconstruction using overlap

After embedding the watermark by modification, an inverse FFT will be applied to produce the time-domain signal. There is an issue of how to reconstruct the time-domain signal if we apply a window function in the analysis stage. There are two reasons for using overlapping at embedding stage when derive FFT spectrum. The first reason is that it can prevent spectrum discontinuity and restore the lost signal data at each frame boundary [61]. The second reason is that by using the overlapping window effects can

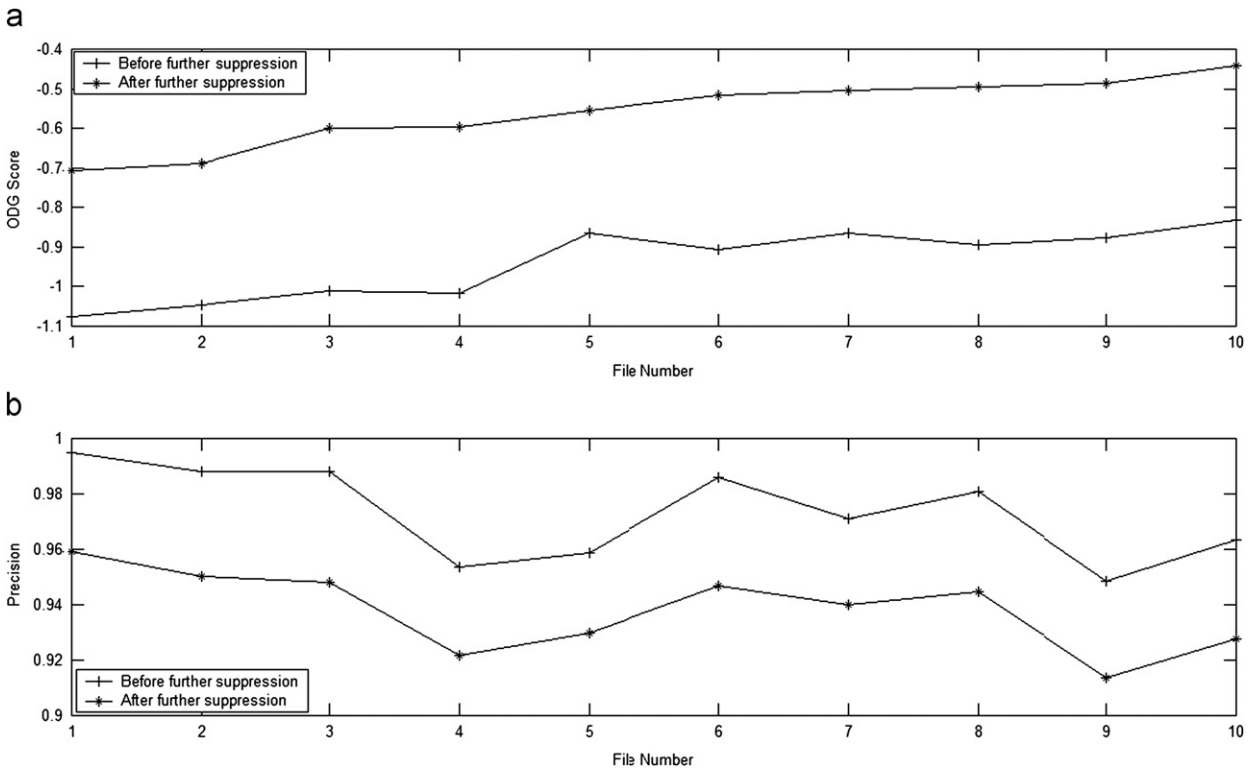


**Fig. 3.** (a) ODG score before and after applying further suppression and (b) precision before and after applying further suppression.

be removed when reconstructing the time domain signal. Specifically, when a Constant Overlap Add (COLA) constraint is satisfied [59], the window effects can be removed completely after reconstruction. The COLA is expressed as

$$\sum_{m\in z} W(mR-n) = 1 \ \forall n \tag{15}$$

where $W$ is the window function, $m$ is the frame number, $R$ is the hop size and $n$ is the sample index in a frame.

From Eq. (15), it can be seen that if the sum of the time-domain window samples applied in each frame is equal to 1, then the window effect can be canceled. For example, considering the well-known Hanning window, when the hop size is equal to half of the frame size, the COLA constraint will be fulfilled. However, we cannot use an overlap during the watermark embedding as it will have a negative impact on the watermarking recovery accuracy. For example, when the spectrum of the frame $m$ is modified to satisfy the embedding criteria, the next frame $m+1$ will also possibly be modified to meet the embedding criteria. As an overlap between spectrum $m$ and $m+1$ exists, it is very likely that the spectrum modification on frame $m$ will be affected by that on frame $m+1$, which may result in damage to the embedded watermark information in frame $m$. This phenomenon could appear in each frame and thus seriously degrade the precision of watermarking. Some experiments have been carried out which verify that doing this can reduce the precision to around 70%. However, overlap can be used in a different way to reconstruct the time domain signal without affecting watermark precision as following:

(a) Before embedding, Hanning-windowed FFT of each frame, without using overlap and use overlap, is computed, respectively, and thus obtains the entire signal spectrum $X_1(\omega)$, $X_2(\omega)$.
(b) During embedding, only manipulating $X_1(\omega)$ and then the modified spectrum $X'_1(\omega)$ is obtained.
(c) During reconstruction, $X'_1(\omega)$ and $X_2(\omega)$ will be combined to reconstruct time-domain signal.

Take one time-domain sinusoidal signal $x_1$ as an example to detail the procedure of reconstruction, where the frame length $l=1024$. $X'_1(\omega)$ and $X_2(\omega)$ will be firstly converted to time-domain signal $x'_1$ and $x_2$, respectively, using the inverse

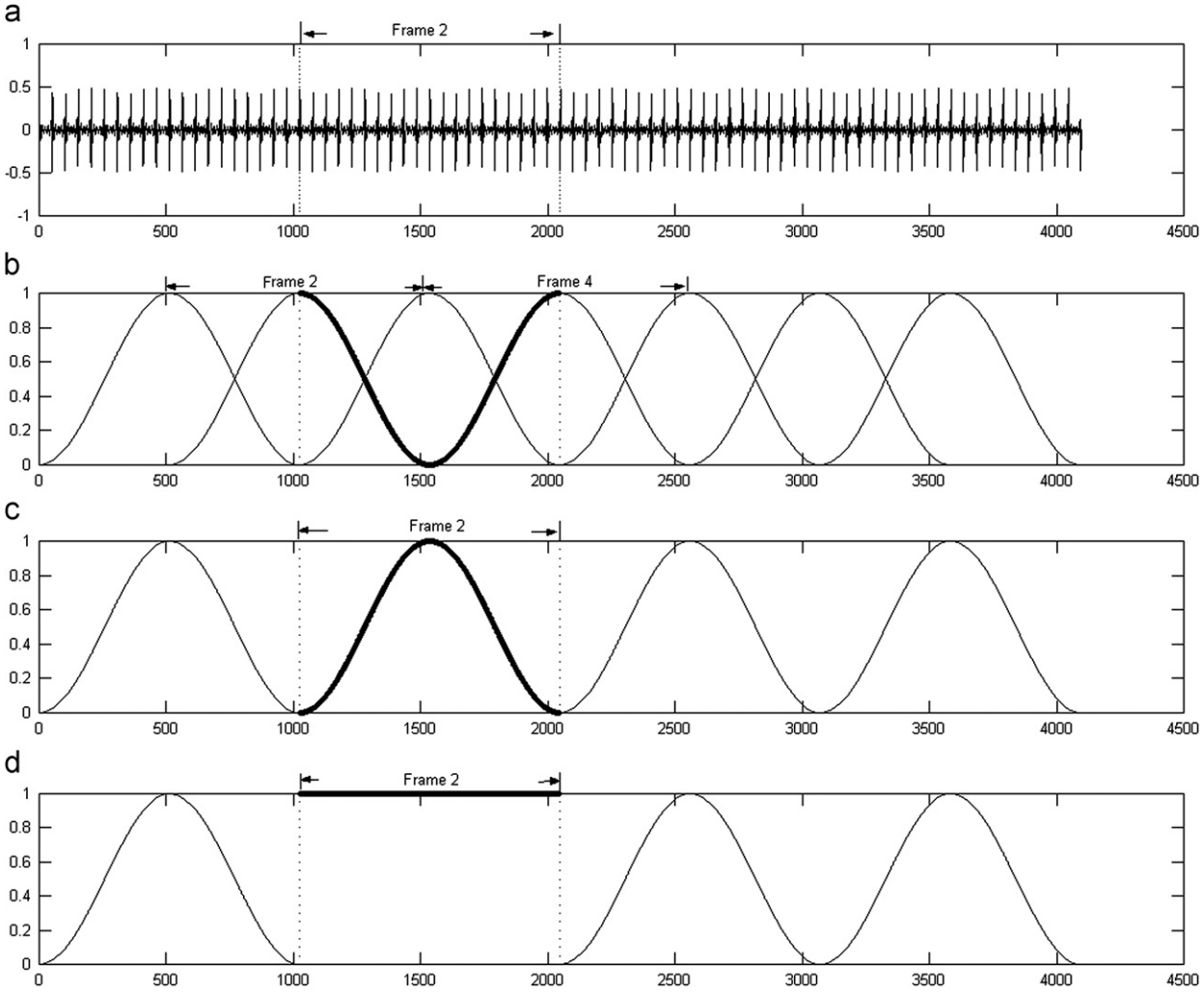


**Fig. 4.** (a) $x_1$, (b) $w_2$: window frame series used in overlap FFT, (c) $w_1$: window frame series used in non-overlap FFT and (d) COLA principal.

FFT. Denoting $w_1$, $w_2$ as the time-domain window frame series applied in non-overlap and overlap FFT, respectively. $x_1$, $w_2$ and $w_1$ are shown in Fig. 4(a)–(c), respectively, where $x$-axis denotes sample index, $y$-axis denotes amplitude.

Let us illustrate how to reconstruct the second frame of time-domain watermarked signal, as an example.

$x_1[n]$, $x_1'[n]$, $x_2[m]$ and $x_2[k]$ denote the second frame of $x_1$ and $x_1'$, the second and fourth frame of $x_2$, respectively. Firstly, $x_1'[n]$ can be rewritten as

$$x_1'[n]=x_1[n]w_1[n]+\Delta=x_1[n_1]w_1[n_1]+x_1[n_2]w_1[n_2]+\Delta \tag{16}$$

where $l+1\le n\le l^*2$, $l+1\le n_1\le l^*3/2$, $(l^*3/2)+1\le n_2\le l^*2$, $\Delta$ denotes the modification on the second time-domain signal as a result of embedding the watermark.

Similarly, $x_2[m]$ can be rewritten as

$$x_2[m]=x_1[m]w_2[m]=x_1[m_1]w_2[m_1]+x_1[n_1]w_2[n_1] \tag{17}$$

where $(l/2)+1\le m\le l^*3/2$, $(l/2)+1\le m_1\le l$.

Likewise, $x_2[k]$can be rewritten as

$$x_2[k]=x_1[k]w_2[k]=x_1[n_2]w_2[n_2]+x_1[k_2]w_2[k_2] \tag{18}$$

where $(l^*3/2)+1\le k\le(l^*5/2)$, $l^*2+1\le k_2\le l^*5/2$.

Based on COLA, we can derive

$$w_1[n_1]+w_2[n_1]=I \tag{19}$$

$$w_1[n_2]+w_2[n_2]=I \tag{20}$$

where

$$I=\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}_{(l/2)\times 1}$$

as shown in panel (d) of Fig. 4.

Based on the above Eqs. (16)–(20), adding$x_1'[n]$, $x_1[n_1]$ $w_2[n_1$ QUOTE and $x_1[n_2]w_2[n_2]$ element-wise, the second frame time-domain data of watermarked signal can be reconstructed as shown below:

$$\begin{aligned}&x_1'[n]+x_1[n_1]w_2[n_1]+x_1[n_2]w_2[n_2]\\ &\quad=x_1[n_1]w_1[n_1]+x_1[n_2]w_1[n_2]+\Delta+x_1[n_1]w_2[n_1]+x_1[n_2]w_2[n_2]\\ &\quad=x_1[n_1](w_1[n_1]+w_2[n_1])+x_1[n_2](w_1[n_2]+w_2[n_2])+\Delta\\ &\quad=x_1[n_1]+x_1[n_2]+\Delta=x_1[n]+\Delta\end{aligned} \tag{21}$$

where $x_1[n]+\Delta$ represents the second frame time-domain data of the watermarked signal.

By this way, the advantage of overlap has been taken and the window effect has been canceled completely.

### 3.2. Procedure

#### 3.2.1. Embedding procedure

A block diagram of our complete watermark embedding scheme is given in Fig. 5. The FFT is found of each frame of the original signal and then the RSVD is applied to each spectrum after its reorganization into matrix form. The elements in the vicinity of row $t$ in $u_{1\ldots m,2}$ are modified to satisfy the embedding conditions detailed above. The psychoacoustic model is incorporated to further remove audible distortion. Note that the procedure below only shows the case of embedding the bit '1', and $t$ is the embedding position.

#### 3.2.2. Extraction procedure

While synchronization of the candidate audio is not a consideration of the work at the current stage of development, we appreciate that it is an issue that we must consider for the future. For the purpose of this research, we assume synchronization has been achieved. Then, the watermark extraction procedure can be briefly described in the following steps:

(a) Split candidate watermarked signal into frames (frame size should match that in the embedding stage).

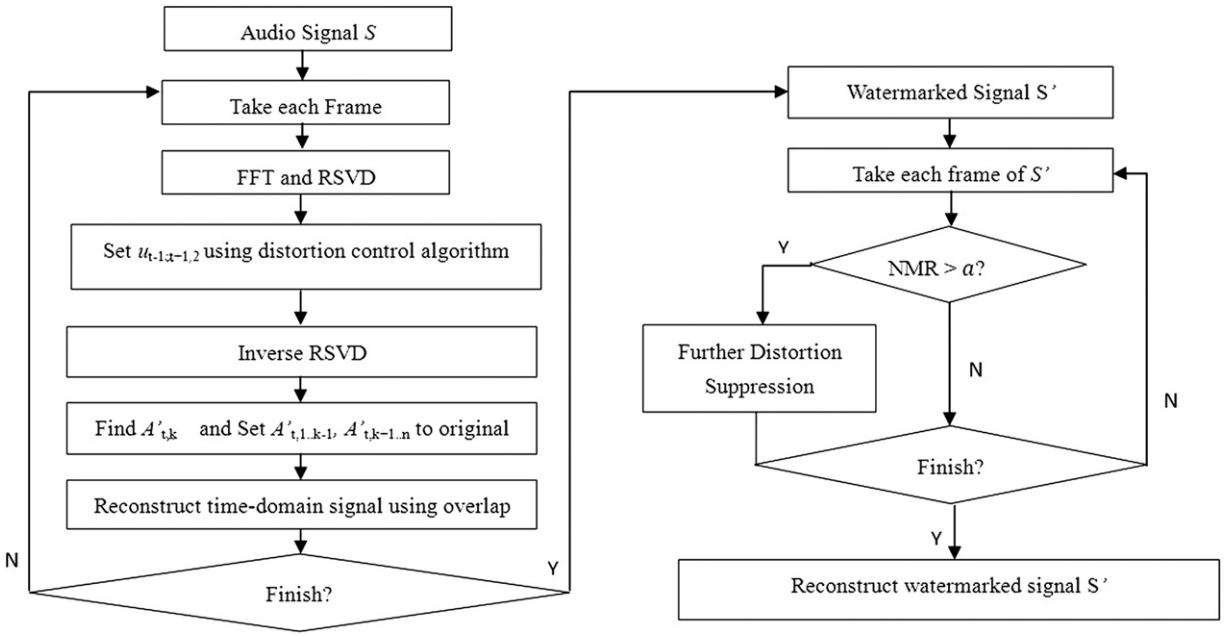


**Fig. 5.** The flowchart of watermark embedding.

(b) Apply the FFT to each frame and then organize its spectrum magnitudes to reconstitute the matrix $A'$.
(c) Apply the RSVD to $A'$, check if $u'_{t,2}-u'_{t+1,2}>0$, where $t$ is the pre-defined key from the embedding stage.
   I. If this condition is true, the bit extracted is a '1',
   II. Otherwise, the bit extracted is a '0'.

## 4. Evaluation

In this section, we evaluate the perceptual transparency of the proposed watermarking scheme, along with its robustness and capacity. For evaluation purposes, we randomly selected 25 music files of different styles and performed the full 'embed–extract–decode' cycle on each file. Each file was a 32-bit audio signal in WAV format sampled at 48,000 Hz. The watermark is a pre-defined random sequence of bits. The input signal was segmented into frames of length 1024 samples. By experiment, we found that embedding four bits into each frame can achieve a good compromise between capacity and imperceptibility. In this experiment, the length of the watermark bit sequence is 8192 bits.

### 4.1. Imperceptibility

#### 4.1.1. ODG

Subjective listening tests are not the ideal means of evaluating the perceptual transparency of an audio watermarking scheme for many reasons. The hearing ability of different listeners can vary enormously depending on their age, lifetime exposure to loud noises and even personal preference in musical tastes, not to mention that some listeners may be trained expert listeners, some may be audio professionals and some may be so-called 'average' music consumers [35]. Therefore, an objective audio quality measurement is more desirable to provide a fair evaluation of the perceptual transparency of a watermarking scheme, independent of the application or platform, in a consistent manner. The 'Perceptual Evaluation of Audio Quality' (PEAQ) is an International Standard used for such purposes for more than 10 years (Standard ITU-R in 1999) [40]. Estimates of audio quality by the PEAQ were compared to scores obtained from subjective listening tests, from which it was established that the correlation coefficient between objective and subjective scores was 0.837 for the Basic Version and 0.851 for the Advanced Version [40]. Though the PEAQ cannot be a complete replacement for subjective listening tests [38], it provides an objective quality measurement that is used by the audio industry [50]. The ODG is the measure used to quantify the perceived difference between a reference signal and a candidate. The ODG score ranges from 0 to $-4$, with 0 indicating that the two signals are perceptually identical and $-4$ indicating that there are perceptual differences between the two that would be so noticeable as to be described by listeners as 'annoying'. Therefore, the closer the ODG score is to zero, the more likely the signals are to be perceived as identical. By extension, if PEAQ analysis of an unwatermarked reference signal and a watermarked candidate signal produces an ODG score close to zero, this would suggest that any difference between them was imperceptible. A PEAQ test was carried out on 25 music files producing a mean ODG score of $-0.6352$, which is in the imperceptible range [4], and standard deviation of ODG score is 0.2258. Fig. 6 shows the distribution of ODG scores.

From Fig. 6 we can see that the ODG scores for all 25 files are between 0 and $-1$, indicating that any difference between the original signal and corresponding watermarked counterpart can be claimed to be either 'imperceptible' or 'perceptible but not annoying'. This is an acceptable result at this stage. We note that some of the ODG scores are very

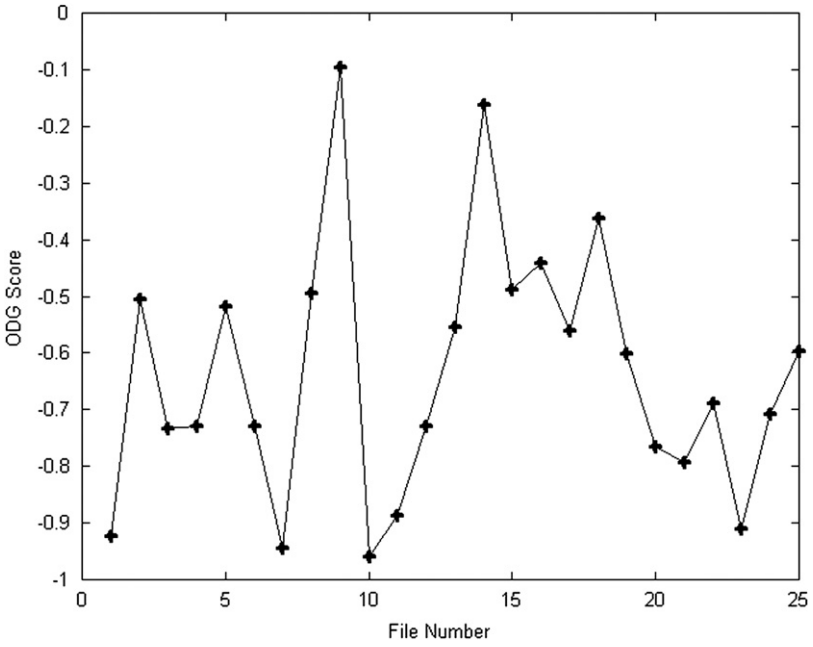


**Fig. 6.** ODG score distribution for all the 25 music files.

close to 0 while some other scores are close to $-1$. The reason for this variation is that the distortion control algorithm applied is not exactly mapped to the perceptual features of music. We believe that this result can be improved by refining the distortion measurement model to make it dynamically adaptable to the inherent content of each signal.

#### 4.1.2. SNR

According to International Federation of the Phonographic Industry (IFPI) [15,51], audio watermarking should be imperceptible when *SNR* is over 20 dB. *SNR* has been used widely to measure the quality of watermarked signal [52] which is formulated as

$$SNR = 10\log_{10}\frac{\sum_n S^2(n)}{\sum_n [s(n)-s'(n)]^2} \tag{22}$$

where $s(n)$ is the time-domain original signal and $s'(n)$ is the time-domain watermarked signal. Since Eq. (22) weights all time-domain errors equally, not taking time-varying energy and time-varying distortion into account, a much-improved quality measure could be obtained if *SNR* is measured over short frames and the results averaged. Thus, a frame-based measurement is defined, called the 'segmental signal to noise ratio' ($SNR_{seg}$), as follows [52]:

$$SNR_{seg} = \frac{1}{M}\sum_{j=1}^{M} 10\log_{10}\left[\sum_{n=N*(j-1)+1}^{N*j}\frac{s^2(n)}{[s(n)-s'(n)]^2}\right] \tag{23}$$

where $M$ is the number of frames and $N$ is the frame size, in our experiment, $M$ equal to 2048 and $N$ equal to 1024. Problems arise with the $SNR_{seg}$ if frames of silence are included as they produce large negative $SNR_{seg}$ values. This problem can be solved by setting a lower threshold and replacing all frames with $SNR_{seg}$ below this threshold equal to the threshold instead (a 0 dB threshold is reasonable) [52]. At the other extreme, frames with $SNR_{seg}$ greater than 35 dB are not perceived by listeners as being significantly different but affect the resulting $SNR_{seg}$. An upper threshold (normally 35 dB) can be used to reset any unusually high $SNR_{seg}$ to this upper threshold [52]. We calculated the $SNR_{seg}$ and *SNR* resulting from our watermarking scheme. These values can be used to compare our scheme with many other schemes. The distribution of *SNR* and $SNR_{seg}$ are shown in Fig. 7. The mean of *SNR* and $SNR_{seg}$ are 27.2304 and 27.5826 dB, respectively, and the standard deviations are 2.6556 and 3.1264 dB, respectively. Note from Fig. 7 that all *SNR* values are above 20 dB, which conforms to the IFPI standard.

### 4.2. Robustness

The ultimate watermarking method should resist any kind of manipulation introduced by accidental or deliberate manipulation or modification of the watermarked signal, whether direct or indirect. No such perfect method has been proposed so far and it is not clear whether an absolutely robust watermarking method can exist [11]. Any manipulation of an audio file can result in an attack on the embedded watermarks. Depending on the way the audio will be used, some attacks are more likely than others. For example, if the watermark is used as protection against illegal copying of

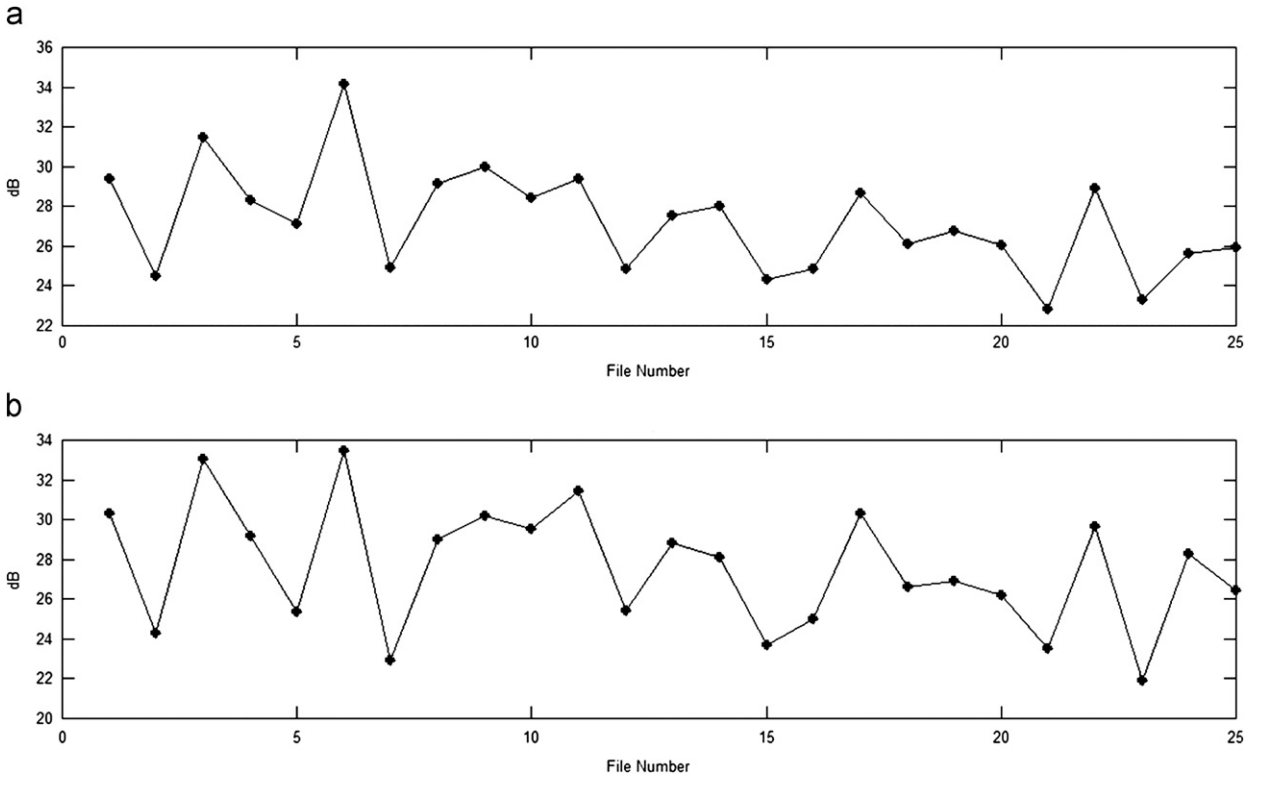


**Fig. 7.** (a) *SNR* distribution for all the 25 music files and (b) $SNR_{seg}$ distribution for all the 25 music files.

copyrighted music, the main attack will be lossy compression such as MP3, sometimes at very high compression rates. A detailed categorization of different attacks is described in [11]. Experiments were performed [53] to test if each attack can itself degrade perception of audio quality. A conclusion drawn was that the perceptual impact of the attack on audio quality is highly dependent on the attacked material. The same attack could have a noticeable impact in one case and be completely inaudible in another. If the attack itself impairs audio quality to such an extent that it reduces the commercial value of the watermarked audio, then this attack will not be considered as an appropriate candidate for this watermarking scheme. In this work, the attacks studied are restricted to MP3 compression, Equalization, Lowpass filtering, Highpass filtering and Noise removal, as they introduce less audible impairment.

Robustness of the watermarking scheme is dependent on correctly decoding the watermark from the signal after it has undergone one or more attacks. It is calculated as the percentage of watermark bits correctly decoded in the post-attack extraction, and is defined as

$$Precision_i = \frac{L - \sum_{i=1}^{L} |W(i) - W'(i)|}{L} \quad (24)$$

where $W$ is the binary bit sequence embedded, $W'$ is the binary bit sequence extracted, $L$ is the length of watermark bits (in our experiments, this is 8192 but can be any user-defined length) and $i$ denotes the file number.

#### 4.2.1. Robustness against MP3 compression

Any watermarking scheme intended for use in music should survive MPEG-1 Layer III compression algorithm [54], at a variety of compression bit rates. (Note, compression bit rates are calculated for a stereo signal, so for example, 128 kbps means 64 kbps in each of L and R channels.) Comparisons of the effect of MP3 compression were explored in [55]. Compression to 96 kbps makes the compressed audio clearly lack definition; compression to 128 kbps introduces noises slightly less defined than the original; compression to 192 kbps results in sound that does not feel the same as the original but it is impossible to tell exactly what the difference is. In order to ensure our watermark scheme is robust against strong MP3 compression, we chose to compress using bit rates from 64 to 160 kbps, as bit rates below 64 kbps are not recommended to compress music. Another experiment [56] was performed to evaluate the extent of degradation of audio quality by MP3 compression. The resultant ODG scores were $-3.506$ after 64 kbps compression, $-1.179$ after 128 kbps compression, and $-0.146$ after 192 kbps compression. This indicates that audio which has been compressed to 64 kbps using MP3 compression is perceptually noticeable, annoyingly so at lower bit rates.

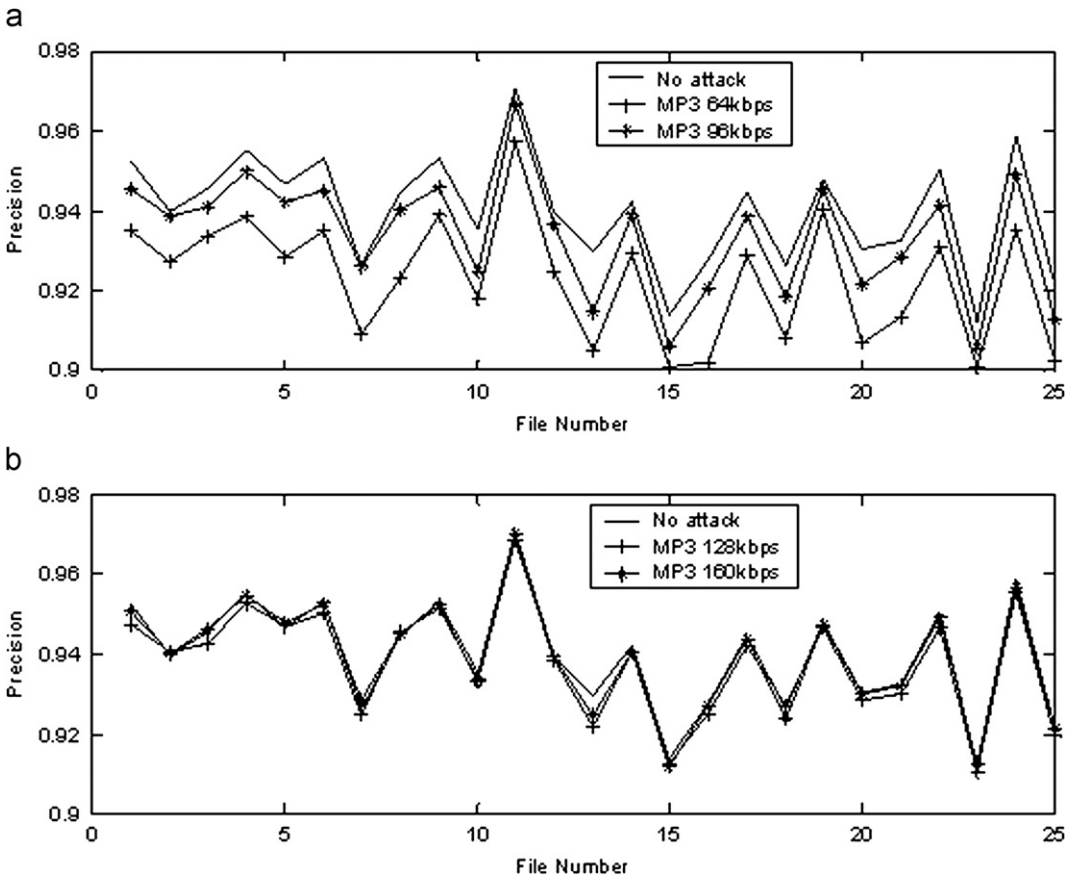


**Fig. 8.** (a) Robustness against MP3 attacks at 64, 96 kbps and (b) robustness against MP3 attacks at 128, 160 kbps

Recall from Section 4.1.1 that the mean ODG score achieved by our proposed watermarking scheme over 25 watermarked signals was −0.6352. This is actually approximately halfway between the ODG score achieved by MP3 compression at 128 and 192 kbps MP3, suggesting that the perceptual transparency of our proposed scheme is at least as acceptable as the quality of 128 kbps MP3 compression, which is widely used in many application domains, and approaching the almost imperceptible ODG score by 192 kbps MP3 compression.

The robustness of our proposed watermarking scheme against MP3 compression at different bit rates is shown in Fig. 8. We can see that the proposed scheme produces watermarked signals that can survive MP3 compression from 64 to 160 kbps to a high degree of precision. When the bit rate is 128 kbps or higher, the recovery precision is almost unchanged when compared to precision of recovery in the absence of any attack, while a 2–3% loss in precision is encountered when the compression bit rate is 64 kbps.

*4.2.2. Robustness against other selected attacks*

We also evaluated the robustness of the scheme against other encountered attacks as listed below:

(a) *Noise removal*: Noise reduction is 24 dB.
(b) *Equalization*: Use 'Amradio' curve to equalize, provided by Audacity tool [57].
(c) *Lowpass filtering*: The roll-off is 6 dB per octave; the cut-off frequency is 8000 Hz.
(d) *Highpass filtering*: The roll-off is 6 dB, the cut-off frequency is 2000 Hz.

As an aside, Equalization with 'Amradio' curve and Highpass filtering with cut-off frequency 2000 Hz introduced more audible distortion than the other two selected attacks: Lowpass Filtering and Noise removal.

Fig. 9 shows the recovery precision achieved against four different attacks as mentioned above.

The mean and standard deviation of the recovery precision for all attacks tested is shown in Table 1.

**Table 1**
Summary of the mean and standard deviation of precision against different attacks.

| Attacks | Mean | Standard deviation |
|---|---|---|
| No attack | 0.9400 | 0.0143 |
| Mp3 at 64 kpbs | 0.9229 | 0.0154 |
| Mp3 at 96 kpbs | 0.9338 | 0.0152 |
| Mp3 at 128 kpbs | 0.9381 | 0.0143 |
| Mp3 at 160 kpbs | 0.9393 | 0.0145 |
| Lowpass filtering | 0.9400 | 0.0145 |
| Highpass filtering | 0.9030 | 0.0288 |
| Noise removal | 0.9152 | 0.0584 |
| Equalization | 0.9083 | 0.0227 |

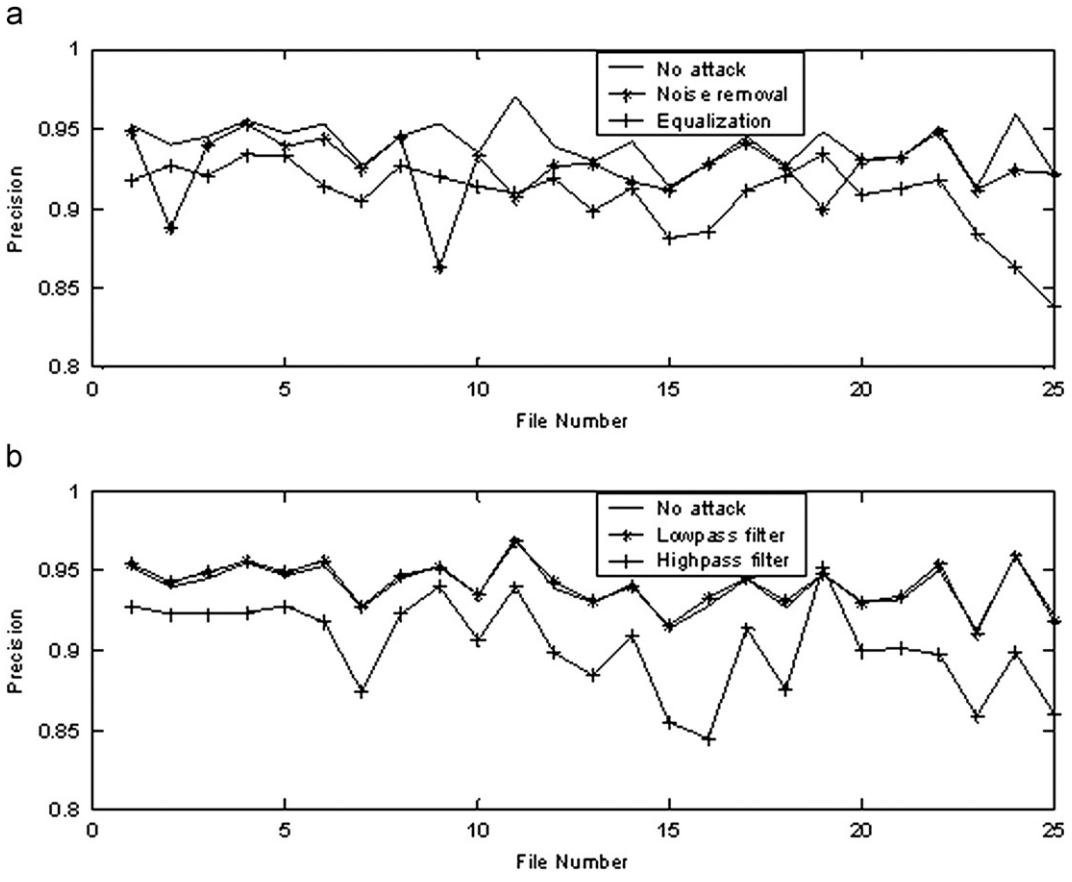


**Fig. 9.** (a) Robustness against 'Noise removal' and 'Equalization' and (b) robustness against 'Lowpass filtering' and 'Highpass filtering'.

### 4.3. Capacity

The data embedding capacity of a watermarking scheme refers to the number of bits that can be embedded into the audio, measured in bits per second (bps) [15]. As for our scheme, the payload, with 4 bits per frame, is ~187 bps. This is a comparatively high capacity when compared to audio watermarking schemes described in [15,41–49,51]. The payload capacity of the scheme can be improved by embedding more than 4 bits in each frame by modifying more elements in the second column of *U*.

### 4.4. Using the mode operation to increase robustness

As mentioned in Section 3.1.3, there is a trade-off between data capacity, robustness and imperceptibility. Robustness, in particular, is important in applications such as tracking illicit distribution or ensuring the watermark data survives the harshest transmission environment so it can still be decoded. A further development of the scheme was introduced to improve its robustness. This development was expected to have a negative impact on capacity but it was felt that the benefits would outweigh the loss in certain application scenarios.

In statistics, the 'mode' is used to find the data which occurs most frequently in a defined data set [58]. We can incorporate the mode idea into our watermarking scheme. The process is described as follows:

(a) Given an input watermarking sequence $W$ with the length of $nb$ bits;
(b) At the embedding stage, $W$ will be embedded repeatedly into $ns$ consecutive music sections.

At the extraction stage, we extract all the bits and then segment into $ng$ equal-size groups $W'_{1\ldots ng}$

The $i$th bit of the extracted watermark $W''$ is determined as the mode of data set $\{W'_{1,i}, W'_{2,i}, W'_{3,i}, \ldots, W'_{ng,i}\}$.

For our experiment, $nb$=8192, $ng$=5, that is, the input watermark $W$ is repeatedly embedded five times into each music file. The precision of the watermark scheme incorporating the mode operation is distributed as Fig. 10.

The mean precision of the scheme after incorporating the mode operation was 99.23% for the original and 98.78% after MP3 compression to 64 kbps. The respective standard deviations were 0.0118 and 0.0138. Therefore, the precision in the presence of a 64 kbps MP3 compression attack has been improved to almost 100% after applying the mode operation. As the algorithm is the same as that described in Section 3.2.1 except for the addition of the mode operation, the perceptual transparency will not be affected. Furthermore, robustness against other attacks such as those introduced in Section 4.2.2,

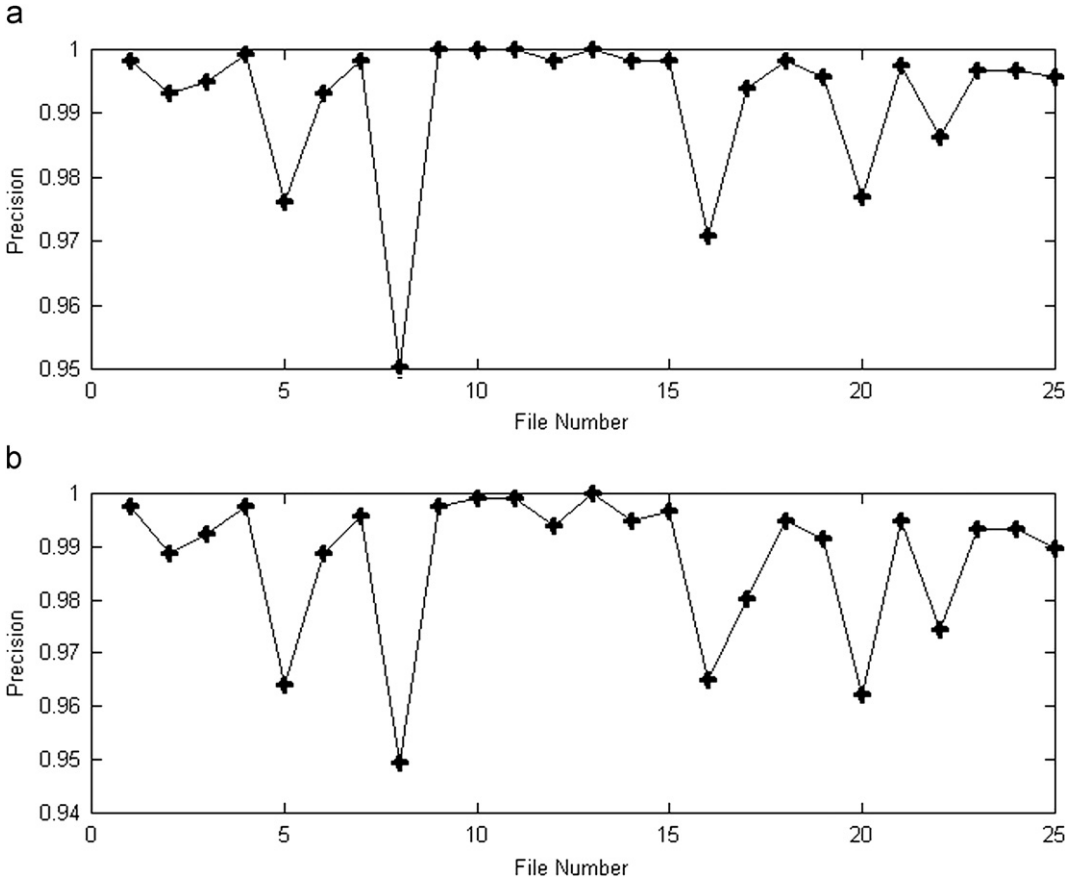


**Fig. 10.** (a) Precision distribution without any attack on 25 music files and (b) precision distribution after 64 kbps MP3 on 25 music files.

**Table 2**
Performance of audio watermarking schemes, sorted by data payload.

| Reference | Method | Payload | Blind | Robust to 64 kbps MP3 | SNR | Imperceptibility |
|---|---|---|---|---|---|---|
| Proposed | SVD | 187 | Yes | Yes | 27.2304 | Yes |
| V. Bhat K [15] | SVD | 45.9 | Yes | Yes | 24.37 | Yes |
| A.H. Ali [14] | SVD | N/A | No | N/A | 28.5525 | Yes |
| E. Ercelebi [57] | LBWT | N/A | No | No | 25.9263 | Yes |
| N. Cvejic [52] | SS | 27.1 | No | Yes | N/A | Yes |
| J. Wang [60] | CSPE | 46.75 | Yes | No | N/A | Yes |

'LBWT' denotes lifting-based wavelet transform; 'SS' denotes spread spectrum; 'CSPE' denotes complex spectral phase evolution.

would be expected to increase with the implementation of the mode operation. The capacity is decreased to $\sim$37 bps ($\sim$187/5).

In Table 2, we compare the performance of several recent audio watermarking schemes. From Table 2, we find that our proposed algorithm has the highest payload of those listed. Additionally, the proposed algorithm achieves a satisfactory compromise between imperceptibility and robustness.

## 5. Conclusion

The audio watermarking scheme presented in this work is based on transforming the spectrum of each frame of the signal using the RSVD. Modifications are then performed on the $U$ matrix resulting from the RSVD to embed watermark bits. Experiments have demonstrated that the proposed scheme has a high data payload and is imperceptible and robust against MP3 and other attacks. In future, the proposed distortion control method will be further studied to develop a means of adapting the scheme more closely to each specific music file. This would result in an even better guarantee for imperceptibility of the watermark. Finally, synchronisation [62] and error correction [63], along with investigation of multiple-column embedding, will be incorporated into the proposed scheme to improve its performance. The security of the scheme will also be explored by using cryptographic techniques [64] so that it can be deployed in environments where audio security is a paramount concern.

## References


[1] N. Cvejic, T. Seppanen, Digital Audio Watermarking Techniques and Technologies, Information Science Reference, USA, 2007.
[2] I. Cox, M. Miller, J. Bloom, Digital Watermarking, Morgan-Kaufmann Publishers, 2003.
[3] C.I. Podilchuk, E.J. Delp, Digital watermarking: algorithms and applications, IEEE Signal Process. Mag. 18 (4) (2001) 33–46.
[4] J. Blackledge, F. Omar, Audio data verification and authentication using frequency modulation based watermarking, international society for advanced science and technology, J. Electron. Signal Process. 3 (2) (2008) 51–63.
[5] 〈http://www.riaa.com/physicalpiracy.php〉 (last accessed 28 June 2010).
[6] A.N. Lemma, J. Aprea, W. Oomen, L. Van de Kerkhof, A temporal domain audio watermarking technique, IEEE Trans. Signal Process. 51 (4) (2003) 1088–1097.
[7] W.-N. Lie, L.-C. Chang, Robust high-quality time-domain audio watermarking based on low-frequency amplitude modification, IEEE Trans. Multimedia 8 (1) (2006) 46–59.
[8] X.-Y. Wang, H. Zhao, A novel synchronization invariant audio watermarking scheme based on DWT and DCT, IEEE Trans. Signal Process. 54 (12) (2006) 4835–4840.
[9] S. Wu, J. Huang, D. Huang, Y.Q. Shi, Efficiently self-synchronized audio watermarking for assured audio data transmission, IEEE Trans. Broadcast. 51 (1) (2005) 69–76.
[10] D. Kirovski, H.S. Malvar, Spread-spectrum watermarking of audio signals, IEEE Trans. Signal Process. 51 (4) (2003) 1020–1033.
[11] Stefan Katzenbeisser, Fabien A Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, Inc., Norwood, MA, 2000.
[12] Peter H.W. Wong, Oscar C. Au, Y.M. Yeung, A novel blind multiple watermarking technique for images, IEEE Trans. Circuits Syst. Video Technol. 13 (8) (2003) 813–830.
[13] T. Painter, et al., Perceptual Coding of Digital Audio, ETATS-UNIS, Institute of Electrical and Electronics, New York, NY, USA, 2000.
[14] A.H. Ali, M. Ahmad, Digital audio watermarking based on the discrete wavelets transform and singular value decomposition, Eur. J. Sci. Res. 39 (1) (2010) 6–21.
[15] V. Bhat K, et al., An adaptive audio watermarking based on the singular value decomposition in the wavelet domain, Digital Signal Process. doi:10.1016/j.dsp.2010.02.006.
[16] K. Chung, W. Yang, Y. Huang, S. Wu, Y. Hsu, On SVD-based watermarking algorithm, Appl. Math. Comput. 188 (1) (May 2007) 54–57.
[17] M. Fan, S. Li, Restudy on SVD-based watermarking scheme, Appl.-Math. Comput. 203 (2) (2008) 926–930.
[18] C.C. Chang, P. Tsai, C.C. Lin, SVD-based digital image watermarking scheme, Pattern Recogn. Lett. 26 (2005) 1577–1586.
[19] H.C. Andrews, C.L. Patterson, Singular value decomposition (SVD) image coding, IEEE Trans. Commun. 24 (1976) 425–432.
[20] A. Ranade, S.S. Mahabalaro, S. Kale, A variation on SVD based image compression, Image Vis. Comput. 25 (2007) 771–777.
[21] Y. Kim, H. Kang, K. Kim, S. Han, A Digital Audio Watermarking Using two Masking Effects, Lecture Notes in Computer Science, Springer, Berlin, 2002, pp. 105–115pp. 105–115.
[22] E. Ganic, N. Zubair, A. Eskicioglu, An optimal watermarking scheme based on singular value decomposition, in: Proceedings of the IASTED International Conference on Communication, Network, and Information Security, 2003, pp. 85–90.
[23] D. Chandra, Digital image watermarking using singular value decomposition, in: Proceedings of the IEEE 45th Midwest Symposium on Circuits and Systems, vol. 3, August 2002, pp. 264–267.
[24] R. Sun, H. Sun, T. Yao, A SVD and quantization based semi-fragile watermarking technique for image authentication, in: Proceedings of the 6th International Conference on Signal Processing (ICSP'02), August 2002, pp. 1592–1595.
[25] H. Ozer, B. Sankur, N. Memon, An SVD based audio watermarking technique, in: Proceedings of the 7th ACM Workshop on Multimedia and Security, August 2005, pp. 51–56.
[26] X. Zhang, K. Li, Comments on an SVD-based watermarking scheme for protecting rightful ownership, IEEE Trans. Multimedia 7 (3) (April 2005) 593–594.
[27] J. Liu, X. Niu, W. Kong, Image watermarking based on singular value decomposition, in: Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Pasadena, CA, USA, December 2006, pp. 457–460.
[28] W. Kong, B. Bian, D. Wu, X. Niu, SVD based blind video watermarking algorithm, in: Proceedings of the International Conference on Innovative Computing, Information and Control, Beijing, China, August 2006, pp. 265–268.
[29] W. Kong, B. Bian, D. Wu, X. Niu, Additive vs. image dependent DWT-DCT based watermarking, in: Proceedings of the International Workshop, MRCS, Istanbul, Turkey, September 2006, pp. 98–105.

[30] Y. Hu, Z. Chen, An SVD-based watermarking method for image authentication, in: Proceedings of 2007 International Conference on Machine Learning and Cybernetics, vol. 3, Hong Kong, China, August 2007, pp. 1723–1728.
[31] L. Trefethen, D. Bau, Numerical Linear Algebra, Society for Industrial and Applied Mathematics (SIAM), PA, USA, 1997.
[32] A.A. Mohammad, A. Alhaj, S. Shaltaf, An improved SVD-based watermarking scheme for protecting rightful ownership, Signal Process. 88 (2008) 2158–2180.
[33] R. Liu, T. Tan, An SVD-based watermarking scheme for protecting rightful ownership, IEEE Trans. Multimedia 4 (1) (2002) 121–128.
[34] X. Zhang, K. Li, Comments on "An SVD-based watermarking scheme for protecting rightful ownership", IEEE Trans. Multimedia 7 (2) (2005) 593–594.
[35] E. Zwicker, H. Fastl, Psychoacoustics Facts and Models, Springer-Verlag, 1990.
[36] J. Nikunen, T. Virtanen, Noise-to-mask ratio minimization by weighted non-negative matrix factorization, in: Proceedings of the 35th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, USA, 2010.
[37] Arnold, Michael, et al., 2002. Quality Evaluation of Watermarked Audio Tracks, vol. 4675 ( XI), International Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, 712 pp.
[38] C. Colomes, et al., Perceptual quality assessment for digital audio: PEAQ—the new ITU Standard for Objective Measurement of Perceived Audio Quality, in: Proceedings of the AES 17th International Conference, Florence, Italy, 1999, pp. 337–351.
[39] Psychoacoustic model implementation, Available 〈http://www.petitcolas.net/fabien/software/mpeg/index.html〉 (accessed: 15 June 2010).
[40] ITU-R Recommendation BS.1387, Method for objective measurements of perceived audio quality, International Telecommunications Union, Geneva, 1998.
[41] W.N. Lie, L.C. Chang, Robust high-quality time-domain audio watermarking based on low-frequency amplitude modification, IEEE Trans. Multimedia 8 (1) (2006) 46–59.
[42] I.-K. Yeo, H.J. Kim, Modified patchwork algorithm: a novel audio watermarking scheme, IEEE Trans. Speech Audio Process. 11 (4) (2003) 381–386.
[43] D. Kirovski, H.S. Malvar, Spread-spectrum watermarking of audio signals, IEEE Trans. Signal Process. 51 (4) (2003) 1020–1033.
[44] R. Tachibana, S. Shimizu, S. Kobayashi, T. Nakamura, An audio watermarking method using a two-dimensional pseudo-random array, Signal Process. 82 (10) (2002) 1455–1469.
[45] M.F. Mansour, A.H. Tewfik, Time-scale invariant audio data embedding, EURASIP J. Appl. Signal Process. (1) (2003) 993–1000.
[46] N. Cvejic, T. Seppanen, Spread spectrum audio watermarking using frequency hopping and attack characterization, Signal Process. 84 (1) (2004) 207–213.
[47] W. Li, X. Xue, P. Lu, Localized audio watermarking technique robust against time-scale modification, IEEE Trans. Multimedia 8 (1) (2006) 60–69.
[48] M.F. Mansour, A.H. Tewfik, Data embedding in audio using time-scale modification, IEEE Trans. Speech Audio Process. 13 (3) (2005) 432–440.
[49] S. Xiang, H.J. Kim, J. Huang, Audio watermarking robust against time-scale modification and MP3 compression, Signal Process. 88 (10) (2008) 2372–2387.
[50] Available 〈http://www.peaq.org/〉 (last accessed: 16 June 2010).
[51] E. Ercelebi, L. Batakcil, Audio watermarking scheme based on embedding strategy in low frequency components with a binary image, Digit. Signal Process. 19 (2) (2009) 265–277.
[52] J. Deller, J. Proakis, J. Hansen, Discrete-time Processing of Speech Signals, Macmillan, 1993.
[53] M. Steinebach, et al., Stir-Mark benchmark: audio watermarking attacks, in: Proceedings of the International Conference on Information Technology: Coding and Computing, Las Vegas, NV, 2001, pp. 49–54.
[54] M. Steinebach, A. Lang, J. Dittmann, StirMark Benchmark: audio watermarking attacks based on lossy compression, in: Edward J. Delp III, Ping Wah Wong (Eds.), Proceedings of the SPIE, vol. 4675, Photonics West, Bellingham, Washington, pp. 79–90.
[55] Available 〈http://www.mp3-tech.org/tests/gb/index.html〉 (last accessed: 16 June 2010).
[56] D. Campeanu, A. Câmpeanu, PEAQ—an objective method to assess the perceptual quality of audio compressed files, in: Proceedings of the International Symposium on System Theory, SINTES 12, 2005, Craiova, România, pp. 487–492.
[57] Available 〈http://audacity.sourceforge.net/download/〉 (last accessed: 16 June 2010).
[58] Available 〈http://en.wikipedia.org/wiki/Mode_(statistics)〉 (last accessed: 16 June 2010).
[59] T. Christos, F. Andreas, Real time spatial representation of moving sound sources, The 123rd AES Convention, New York, NY, USA, October 2007, pp 72–79.
[60] J. Wang, R. Healy, J. Timoney, Perceptually transparent audio watermarking of real audio signals based on the CSPE algorithm, in: ISSC2010, UCC, Cork, Ireland, 2010.
[61] W.Robert Ramirez, The FFT, Fundamentals and Concepts, Prentice-Hall, New Jersey, 1985.
[62] D. Megias, J.S. Ruiz, M. Fallahpour, Efficient self-synchronised blind audio watermarking system based on time domain and FFT amplitude modification, Signal Process. 90 (2010) 3078–3092, doi:10.1016/j.sigpro.2010.05.012.
[63] S. Baudry, J.-F. Delaigle, B. Sankur, B. Macq, H. Maître, Analyzes of error correction strategies for typical communication channels in watermarking, in: Proceedings of the Special Section on Signal Processing Information Theoretic Aspects Digital Watermarking, Signal Processing (2001) 1239–1250.
[64] Francois Cayre, Caroline Fontaine, Teddy Furon, Watermarking security, part one: theory, in: IS&T/SPIE International Symposium on Electronic Imaging, 2005, pp. 746–757.
[65] K. Hermus, Perceptual audio modeling based on total least squares algorithms, the 112nd AES convention, Munich, Germany, 2002.