Int. J. Electron. Commun. (AEÜ) 69 (2015) 188–197

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

# Perception-based audio watermarking scheme in the compressed bitstream

Baiying Lei[a,*], Ing Yann Soon[b]

[a] *Department of Biomedical Engineering, School of Medicine, Shenzhen University, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen 518060, China*
[b] *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore*

ARTICLE INFO

*Article history:*
Received 25 January 2012
Accepted 23 August 2014

*Keywords:*
Audio watermarking
MPEG-4 SLS audio
Perception
Copyright protection

ABSTRACT

In this paper, a new perception-based watermarking scheme for MPEG-4 scalable to lossless (SLS) audio is proposed. Under the control of the psychoacoustic model, the significant part of integer modified discrete cosine transform (IntMDCT) coefficients are adaptively modified during MPEG-4 SLS audio compression taking robustness, imperceptibility and security into consideration. The chaotic watermark generated by chaos is simply embedded to ensure security. Moreover, the adaptive spread spectrum method is exploited to further tradeoff robustness and transparency of this scheme. Extensive experimental results confirm that the proposed scheme is robust against common signal processing attacks while the inaudibility of the scheme is preserved.

## 1. Introduction

In the last few years, with the development of Internet and availability of the state-of-the-art digital multimedia, it is possible to share a large amount of audio or video files, but it also runs the risks of widespread multimedia production forgeries and copyright violation. Consequently, the problem of digital property and copyright protection of multimedia products has attracted a lot of interest and is gaining more attention. As a solution to copyright protection issues, digital audio watermarking technology has become a focus in information security and achieved unprecedented development.

Meanwhile, with the advancement in the audio standards, audio techniques and features can be exploited to protect copyright with watermarking method. Therefore, multitudes of watermarking schemes are available for the protection of the audio formats like MP3, MPEG-2 and MPEG-4 advanced audio coding (AAC) to prevent copyright pirating and malicious modifications. However, as the newest standard released in 2006 by ISO/IEC, MPEG-4 scalable lossless audio [1] coding integrates the functions of lossless, perceptual and fine granular scalable audio coding in a single framework and allows the scaling up of a perceptually coded representation such as MPEG-4 AAC, there is no report on watermarking schemes for MPEG-4 SLS audio watermarking schemes in the literature to our best of knowledge. The widespread use of the compressed MPEG-4 SLS audio format on the Internet, the commercial importance of copyright protection, and the potential profit loss caused by illegally copying necessitate a proposal for the MPEG-4 SLS audio copyright protection. Actually, embedding watermarks for MPEG-4 SLS audio in the compressed domain based on IntMDCT is very feasible.

Recently, some work has shown that the chaotic maps can be adopted in digital watermarking to increase the security [2–7]. In [8], due to the unique features of chaos, chaotic sequences have shown superior robustness when compared to the widely used pseudo noise (PN) sequences in watermarking applications. It is evident that watermark embedding process can easily integrate the existing coding schemes with chaotic watermark generation. Thus chaotic sequences are adopted to enhance the security, anticounterfeit and noninvertibility of the proposed watermarking scheme.

The goal of this paper is to design and analyze a robust, feasible and applicable watermarking scheme based on chaos and human auditory system (HAS) model to protect MPEG-4 SLS audio clip in the IntMDCT domain. Finding a balance point between robustness and transparency will be an important problem adequately solved and addressed in this paper. The watermark should be embedded into the MPEG-4 SLS audio compressed bitstream and extracted directly. The chaotic watermark data as copyright information is embedded into the IntMDCT coefficients and detected from the coded bitstream successfully. The modifications should be as small as possible to ensure the watermark inaudibility. Moreover, the watermark should also be robust to various attacks like

* Corresponding author. Tel.: +86 755 26534314; fax: +86 755 26534940.
*E-mail addresses:* leiby@szu.edu.cn, baiying@email.unc.edu (B. Lei).

http://dx.doi.org/10.1016/j.aeue.2014.08.012
1434-8411/

signal processing, and decoding/recoding attacks. Lastly, the proposed method should introduce some possibility of commercial realization and be tailored for a wide range of applications with the exploitation of MEG-4 SLS audio features and encoding and decoding process. All in all, the main contributions of the proposed watermarking framework are threefold. Firstly, it is compatible with the SLS bitstream without much audio quality distortion and robust to MPEG-4 SLS audio compression and coding process. Secondly, it is based on psychoacoustic model and significant state to ensure the imperceptibility. There is no additional psychoacoustic model as we use the SLS perceptual model directly. Chaotic sequence rather than PN sequence is adopted to improve the security. Our method is simple and convenient to implement. Thirdly, it is applied in the IntMDCT domain and adaptive spread spectrum method is used. The watermark can be adaptively modified in a way that tradeoffs between robustness and transparency.

The rest of this paper is organized as follows. A brief review of the recent watermarking schemes is introduced in Section 2. Section 3 describes MPEG-4 SLS structure and related characteristics. Psychoacoustics and the significant state applied in the proposed watermarking scheme are presented in Section 4. Section 5 introduces chaotic sequences in the watermarking field and the proposed watermarking method. The watermark extraction is given in Section 6. Preliminary performance analysis is conducted in Section 7. Experimental results are discussed in Section 8 followed by the paper summary in Section 9.

## 2. Related work

Currently, digital watermarking techniques mainly focus on the design of watermark generation, embedding and extraction in the frequency, spatial, cepstrum or mixed domain with symmetric and public key. Indeed, irrelevant and perceptual audio characteristics can be exploited to hide the extra signal. The essential principles and techniques are of great significance to design a novel and hybrid scheme. A large number of audio watermarking techniques can be found in the literature. The most widely used watermarking techniques are echo hiding, spread spectrum, patchwork and content based methods. Echo hiding method [9,10] embeds a watermark by adding an echo signal to increase robustness. Echo defined as delay or offset can determine imperceptibility of the modifications. However, it is signal dependent or offset dependent and not suitable for speech signal with frequent silent intervals. Spread spectrum method is the mainstream audio watermarking technique weighted in the time (or frequency) domain by adopting PN or chaotic sequence to generate and spread watermark [11,12]. In this scheme, computational complexity and synchronization are the main challenging problems. Patchwork method is implemented by adding and subtracting a constant value from two corresponding sample sets [13]. It is a typical watermarking method that takes advantage of host signal to tradeoff between robustness and imperceptibility. Increasing the watermark strength within the constraints of masking threshold is a good way to find the balance point in this method. Content-based method explores dynamic and unique audio features for watermarking approaches [14,15]. This scheme is signal dependent or adaptive by modifying audio frames or scale factor. Besides, HAS model is usually adopted in this technique.

At the same time, there are some recent watermarking techniques based on MPEG-2 AAC audio [16–19]. For example, watermarking prototype for MPEG-2 AAC advocated in [16] was a solution to the Pulse Coding Modulation (PCM) watermarking problem. The weak point of the method is that the overall performance of the system is not optimal and the bit error rate is still very high. Besides, the watermark is not a blind detection system. Tachibana et al. [17] proposed a bitstream watermarking scheme for MPEG-2 AAC audio based on a two-dimensional pseudo random array and detect the watermark with the correlation method. It has lower robustness as there is no HAS model. Meanwhile, the algorithm still has much room for improvement because no listening test was performed to verify the imperceptibility. Another novel watermarking scheme for MPEG-compressed audio designed by Quan et al. [18] employed wet paper codes and inserted data directly by modifying the MPEG audio quantization process. The novel enhanced spread spectrum AAC watermarking scheme proposed by Cheng et al. was very fast for real-time application as it used the quantization indices directly [19]. The general scheme applied in the DCT domain is also suitable for other domain and other audio formats. This robust watermarking algorithm with low complexity adopts spectral filtering to reduce noise and improve the detection bitrate.

Apart from MP3 and MPEG-2 AAC data protection framework, the latest high data rate audio watermarking techniques are designed in the latest publications [20,21] as well. In [20], Li et al. introduced three techniques to embed watermark adaptively based on advanced audio zip (AAZ). The masking threshold is used in order to hide a watermark signal. Under the constraints and control of the masking threshold, the watermark is adaptively and transparently embedded. The system is highly dependent on the threshold that makes its computation intensive for real-time application. A similar data hiding scheme for audio property protection based on IntMDCT was suggested in [21] for high data rate audio such as MPEG-4 AAC. The data hiding technique using the IntMDCT time–frequency transforms is still possible and applicable for other compressed audio data and spatial sound information. The disadvantage of the hiding technique is that the perceptual model is too simple to achieve higher data rate with less audio distortion. In other words, higher data rate has higher audible degradation that is not suitable for real application.

Furthermore, there are some watermarking schemes based on chaos and HAS in the compressed domain in the literature too. For instance, a novel watermarking method introduced in [2] used the chaotic sequence and MPEG-1 psychoacoustic model to improve robustness in the frequency domain. Besides, this scheme has relatively high robustness, flexibility, and supports multiple watermarking. However, listening test was not conducted to test the imperceptibility. There is also no performance comparison with other methods. Bassia et al. [22] developed an algorithm that modified each audio sample by adding a signal-dependent, low-pass shaped watermark signal. This amplitude modification watermarking scheme generates watermark by thresholding a chaotic map to form a bipolar sequence. It is a non-blind scheme and can resist most of attacks. However, it is not robust to time scale modification and synchronization attacks.

Finally, a description of perceptual audio watermarking schemes can be found in many other publications too [15,23,24]. In 1996, Boney et al. [23] put forward an algorithm to utilize the MPEG psychoacoustic model in order to obtain necessary frequency-masking values to achieve audio transparency. The watermark is generated by filtering a PN sequence with a filter that approximates HAS frequency masking characteristics. Thus these watermarks are signal dependent and different for other audio signals. This technique was further explored by Swanson et al. [15] which took advantage of perceptual coding techniques in order to embed the watermark efficiently and appropriately.

## 3. MPEG-4 SLS structure

MPEG-4 SLS codec adopts the IntMDCT based lossless coding approach [25,26]. The input integer PCM format is losslessly

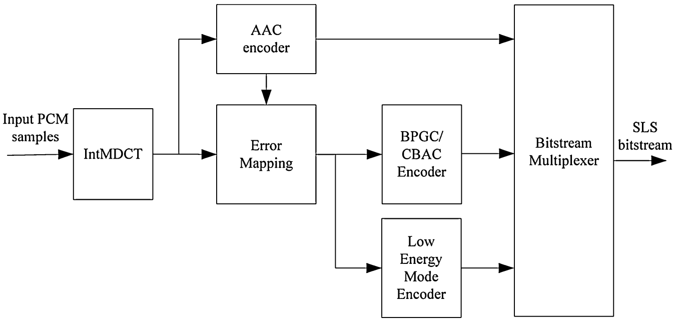

**Fig. 1.** Block diagram of SLS encoder.

transformed into the frequency domain using the IntMDCT. IntMDCT is a lossless integer to integer transform that approximates the normal MDCT transform. The IntMDCT spectral data are coded with a core MPEG-4 AAC layer and another complementary Lossless Enhanced (LLE) layer. Bit-plane coding method is utilized to produce the fine grain scalable to lossless portion of the lossless bitstream. The error mapping process manages to preserve the probability distribution skew of the original IntMDCT coefficients. The resulting coefficients are then coded by the entropy coder used in the LLE layer to generate the core layer AAC bitstream.

Fig. 1 illustrates the basic coding principle of the SLS encoder. Either bitplane Golomb coding (BPGC) or context-based arithmetic coding (CBAC) and low energy mode coding (LEMC) can be used to code the residual spectrum. BPGC is adopted in SLS as the major arithmetic coding scheme. It uses a probability assignment rule that is derived from the statistical properties (Laplacian distributed) of the residual spectrum in SLS. CBAC complements BPGC to further improve the coding efficiency. LEMC is adopted for coding signals from low energy regions to improve the coding efficiency of BPGC by further incorporating more sophisticated probability assignment rules. The low energy bit-planes will be coded at last using LEMC until it reaches the plane of the least significant bit (LSB) for all scale factor bands (SFB). Finally, the LLE bitstream is multiplexed with the core AAC bitstream to produce the final SLS bitstream.

To this end, AAC core encoder can be taken as a quantization and coding process where the IntMDCT spectral data are first grouped into different SFBs, which are then quantized with different quantization step size and coded to produce the AAC bitstream. Assuming that an IntMDCT coefficient $c(k)$ from a particular SFB $s$ is quantized at the AAC core encoder, for $k = \{0, 1 \ldots, N-1\}$ where $N$ is the dimension of IntMDCT, which generates an output $i(k)$, the following property holds:

$$|thr[i(k)]| \leq |c(k)| < |thr[i(k)]| + \Delta[i(k)], \quad k \in s, \tag{1}$$

where $thr[i(k)]$ is the next quantized value closer to zero with respect to $i(k)$ and calculated via table lookup and linear interpolation to ensure the deterministic behavior necessary for lossless coding, $i(k)$ is the quantized IntMDCT spectral data vector produced by the AAC quantizer, that is, $Q[c(k)] = i(k)$, where $Q(\bullet)$ is the non-uniform quantizer in the SLS core layer. $\Delta[i(k)]$ is the quantization step size for $i(k)$. If the SF of the band $s$ to which [0,1] belongs is denoted as $scale\_factor(s)$, then the $thr[i(k)]$ is computed by:

$$thr[i(k)] = \begin{cases} sgn[i(k)]\sqrt[4]{2^{scale\_factor(s)}}(|i(k)| - C)^{4/3}, & i(k) \neq 0, \\ 0, & i(k) = 0, \end{cases} \tag{2}$$

$$\Delta[i(k)] = thr[|i(k)| + 1] - thr[|i(k)|], \tag{3}$$

Here, $scale\_factor(s)$ is the SF that determines the quantization step size for SFB $s$, sgn[$\bullet$] is the sign operator, and the rounding offset $C$ is equal to 0.4054 in the standard.

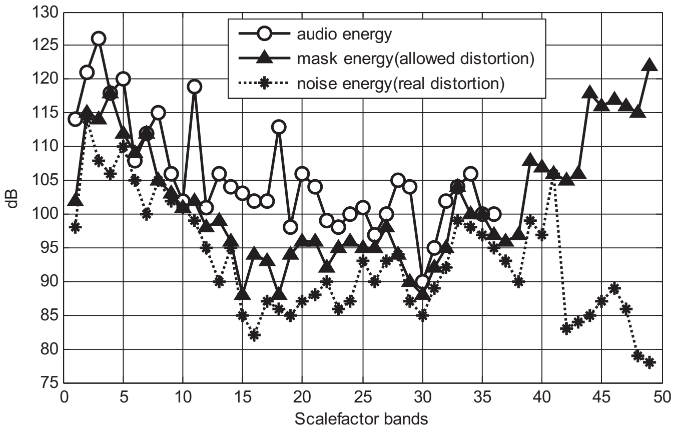

**Fig. 2.** Signal, mask energy and allowed distortion in one frame.

Intuitively, given the SLS quantized value $i(k)$, it is sufficient in most cases to use the minimum mean square error (MMSE) residual obtained by subtracting the IntMDCT coefficient $c(k)$ to its MMSE reconstruction, i.e., $\hat{c}(k) = E\{c(k)|i(k)\}$, where $E\{\bullet\}$ is the expectation operation. However, such an error mapping process will generally produce uniformly distributed errors, $e(k)$, which is virtually incompressible. In order to fully utilize the statistical properties of $c(k)$ implied in Eq. (1), MPEG-4 SLS adopts a somewhat anti-intuitive and different approach in order to produce a residual signal $e(k)$ with the smallest entropy, where the residual signal $e(k)$ is given by:

$$e(k) = \begin{matrix} c(k) - \lfloor thr[i(k)] \rfloor, & i(k) \neq 0 \\ c(k), & i(k) \neq 0 \end{matrix} \quad k = 1, \ldots 1024. \tag{4}$$

## 4. Psychoacoustic model and significant state

### 4.1. Psychoacoustics

Most recent audio coding standards are based on psychoacoustics which describes HAS properties and exploits the “undetectable” signal information. The undetectable information is identified during signal analysis by incorporating into the coder several psychoacoustic principles, including absolute hearing thresholds (ATH), simultaneous masking, the spread of masking along the basilar membrane, and temporal masking. Masking is a fundamental property of HAS and a basic element of MPEG-4 SLS audio. The masking phenomenon occurs when a sound is made inaudible by a “masker”, a noise or unwanted sound of the same duration as the original sound. HAS model delivers a masking threshold that quantifies the maximum amount of distortion at each point such that quantization of the time–frequency parameters does not introduce any audible artifacts. Integrating psychoacoustic notions with basic properties of signal quantization and coding can lead to the theory of transparent audio watermarking.

Actually, the frequency-masking model can be used to obtain the spectral characteristics of a watermark based on HAS inaudible information. Perceptual distortion control is achieved by a psychoacoustic signal analysis that estimates signal masking power based on psychoacoustic principles. Signal, mask and noise energy plot for one frame in an SLS test sequence excerpt is illustrated in Fig. 2, where the excerpt is coded by MPEG-4 SLS reference codec at 128 kbps. It can be observed that the watermark material can be transparently embedded when the quantization and distortion is controlled to be lower than the masking threshold.

### 4.2. Significant state

The probability distribution of bit-plane symbols is usually correlated with their frequency location and the significant state of the adjacent spectral lines [27]. The amplitudes of adjacent spectral lines are correlated due to the leakage of the IntMDCT filterbank. In addition, the amplitude of the IntMDCT spectrum is also highly correlated with the quantization interval of SLS core quantizer if it is present. The significant states of the adjacent spectral lines are designed to capture the correlations among the amplitude of the current IntMDCT spectral data, those of the adjacent IntMDCT spectral lines and the quantization interval of the AAC core quantizer. Besides, a vector, $sig\_cx(k, j)$ is defined to represent the significant states for adjacent IntMDCT spectral line of bit-plane symbol, where $k = 1, \ldots N$ and $j$ is bit-plane, that is:

$$\begin{aligned} sig\_cx(k,j) = \{ & sig\_cx(k-2,j), sig\_cx(k-1,j), \\ & sig\_cx(k+1,j), sig\_cx(k+2,j)\}, \end{aligned} \tag{5}$$

where the significant state $sig(k, j)$ is denoted as:

$$sig(k,j) = \begin{cases} 0 & \hat{c}_j(k) = 0, \\ 1 & \hat{c}_j(k) \neq 0, \end{cases} \tag{6}$$

and $sig(k, j)$ is taken as 0 if $k$ is outside of the IntMDCT spectrum. Here $\hat{c}_j(k)$ is the partial reconstruction for $c(k)$ up to bitplane $j$,

$$\hat{c}_j(k) = \sum_{f=j+1}^{M-1} b(k,f)2^f + thr[i(k)], \tag{7}$$

where $b(k, f) \in \{0, 1\}$ is a bit-plane symbol which is either 0 or 1. The $sig\_cx(k, j)$ context is only used in coding bit-plane symbols from insignificant IntMDCT spectral lines. IntMDCT coefficient $c(k)$ is insignificant when $i(k)$ is equal to zero in the core layer. For those from significant ones (i.e., $i(k) \neq 0$ or the near DC components), the $sig\_core$ context is further introduced. The value of $sig\_core$ context is given by:

$$sig\_core(k) = \begin{cases} 0 & c(k) \text{ is from an insignificant} sfb. \\ 1 & c(k) \text{ is from a significant} sfb. \end{cases} \tag{8}$$

Furthermore, for $sig\_core(k) = 1$, the following rule holds:

$$0 \leq e(i) \leq \Delta(i(k)). \tag{9}$$

Actually, the proposed watermarking algorithm explores both the perceptual model and significant state theories. In this method, the selected coefficients with the constraints of perceptual model and significant state can be modified slightly and adaptively to embed watermark in a way that does not produce any perceived effect.

## 5. Chaotic watermark generation and embedding

As the chaotic functions can generate similar pseudo random sequences with almost identical performance, the simplest chaotic map, logistic map, is chosen to produce the chaotic sequence that encrypts the binary watermark because it is simple to implement. As the sequence generated by the map is composed of real numbers, the output chaotic sequence is quantized into binary stream by the following rule:

$$\beta(n) = \begin{cases} 1, & x(n) \geq 0.5, \\ 0, & x(n) < 0.5, \end{cases} \tag{10}$$

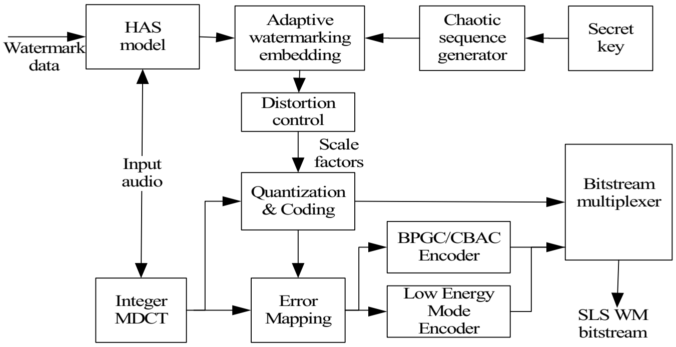

**Fig. 3.** Proposed watermark embedding algorithm.

Finally, the watermark to be embedded into the host audio signal in this scheme is generated by:

$$w(n) = \sum_{n=1}^{N_w} b(n) \oplus \beta(n), \tag{11}$$

where $N_w$ is the length of the watermark, $b(n)$ is binary data which can be any digital signature for copyright or other purpose. All types of digital data (text, images, etc.) that are represented using binary numbers in the computer can be used as $b(n)$. $\oplus$ denotes the Exclusive-OR operation.

Generally, in the proposed watermark embedder, a SS based approach is implemented in the compressed domain. The binary digital signature is spread and encrypted by a chaotic sequence with a secret key. The watermark is embedded in the quantized IntMDCT coefficients after the nonuniform quantization in the core layer, which is usually with different quantization steps in different SFBs to shape the quantization noise so that it can be best masked. In the proposed scheme, adaptive watermark embedding is employed to shape the local watermark strength under the control of psychoacoustic model. The allowed distortion control is also adopted to improve the system performance. Finally, the multiplexer produces the bitstream including data and side information to complete the watermarking process. The proposed watermark embedding system is summarized in Fig. 3.

In order to embed the watermark, the original input audio is segmented into different frames, and each frame is decomposed into different subbands. The IntMDCT transform in the SLS encoder is performed to obtain the transform coefficient. Besides, in this watermarking scheme, the transformed coefficients are grouped into SFBs that are generally fixed at 49 for a sampling rate of 48 kHz.

In order to enhance the audio fidelity, watermark embedded into the original audio is conditional on the just noticeable distortion constraints and perceptual shaping. The watermark is chosen to be added to the perceptually significant scalefactor bands of the spectrum in such a way that the watermark is not easily and illegally removed from the original audio without causing much audio distortion. The psychoacoustic model is applied on those subbands to determine the masking thresholds for each subband. The masking threshold is the allowed distortion for each IntMDCT coefficient. Here it is assumed that similar to perceptual audio coding, if the distortion created in the scalable coding is below the mask, the transparent quality can be achieved. Actually, the allowed distortion (masking threshold), $M(s)$, can be obtained by MPEG-4 SLS psychoacoustic model. Therefore, in the proposed watermarking scheme, $M(s)$ for each SFB $s$ in terms of dB is computed as:

$$M(s)=\begin{cases}\dfrac{E(s)}{SMR(s)}, & E(s)>70\,\text{dB and}SMR(s)>1.\\ E(s), & E(s)>70\,\text{dB and}SMR(s)\le 1.\\ E(s)\times 1.1, & E(s)\le 70\,\text{dB}.\end{cases} \tag{12}$$

Note that Eq. (12) is the default mask implementation in the MPEG-4 SLS, where $s$ $(0\le s\le S)$ and $S$ is the total number of SFB, $SMR$ is the signal to mask ratio. $SMR(s)$ for each SFB $s$ is calculated using psychoacoustic model in the MPEG-4 SLS reference model encoder. $E(s)$ is the signal energy (in terms of dB) for SFB $s$ and computed by:

$$E(s)=\sum_{k=O[s]}^{O[s+1]-1} c^2(k), \tag{13}$$

where $O(s)$ is the starting index of spectrum coefficient for SFB $s$ and $c(k)$ is IntMDCT coefficient. $k$ is the number of the coefficients. The watermark energy is compared with $M(s)$, if the watermark energy is far below $M(s)$, the introduced watermark will be imperceptible. As a result, there will be watermark embedded. Otherwise, if the watermark energy happens to be higher than $M(s)$ during some spectral range, or they are at similar level, there will be no watermark embedded. This adherence to the psychoacoustic model makes it very difficult to "hear" the hidden data, and maintains the transparency of the scheme at the same time.

In an SLS codec, the bit-plane coding is performed in a sequential order. The plane of the most significant bit (MSB) for spectral data from the lowest SFB to the highest SFB is coded first. To be more specific, the illustration of the watermarking scheme for MPEG-4 SLS audio exploring the perceptual model and significant state context information is shown in Fig. 4, where $n/w$ means watermark will be embedded adaptively. For each SFB $s$, if it is declared as significant and under the constraints of allowed distortion, the watermark can be embedded. Otherwise, there will be no watermark data embedded in this SFB. The significant state and allowed distortion determines the region to be embedded and how much can be embedded. As a result, this selection makes the watermark embedding adaptive in nature according to the contents of the audio and MPEG-4 SLS psychoacoustic model.

To further enhance the adaptiveness of the proposed algorithm, the embedding mechanism also employs an adaptive method to determine watermark embedding intensity. The power spectrum $S(k)$ of the audio segment is first computed. Each segment of the signal $s(n)$ is weighted by a Hanning window, $h(n)$. The maximum is normalized to a reference sound pressure level of 96 dB.

$$S(k)=10\log_{10}\left[\frac{1}{N}\left|\sum_{n=0}^{N-1}s(n)h(n)\exp\left(-j2\pi\frac{nk}{N}\right)\right|^2\right]. \tag{14}$$

Suppose that the watermark data to be embedded is represented by $w(k)$, the host signal to be embedded is $S(k)$. The method of embedding watermark into the host signal is expressed as:

$$S'(k)=S(k)+a(k)\times w(k), \tag{15}$$

where $S'(k)$ is the watermarked signal. Generally, if $S'(k)$ is the watermarked signal, and $S(k)$ is the original signal, then the $R_{SNR}$ is usually defined as:

$$R_{SNR}=10\log_{10}\frac{\sum_k S^2(k)}{\sum_k (S'(k)-S(k))^2}, \tag{16}$$

Substitute Eq. (16) to (15), then the scaling factor $a(k)$ can be obtained by:

$$a(k)=\sqrt{\left(\sum_k S^2(k)\right)10^{R_{SNR}/10}}. \tag{17}$$

Thus the adopted SS method can be expressed as:

$$S'(k)=S(k)+\sqrt{\left(\sum_k S^2(k)\right)10^{R_{SNR}/10}w(k)}. \tag{18}$$

For a predefined threshold $R_{SNR}$, the scaling parameter $a(k)$ can be adaptively adjusted by audio signals to control watermarking strength. Consequently, the watermarking system can be adjusted by changing $R_{SNR}$, which is an advantage of the proposed method. Different $R_{SNR}$ is chosen to tradeoff between robustness and imperceptibility of the proposed scheme.

## 6. Watermark extraction

Fig. 5 describes the watermark extraction process. After the SLS watermarked bitstream is parsed, it is entropy decoded to recover $S'(k)$. As the SS method is adopted in the proposed scheme, it is a non-blind method, thus the extraction and detection procedure needs the original signal. The watermark extraction is the reverse process of the watermark embedding process. Specifically, the watermark extraction rule is:

$$w'(k)=\frac{S'(k)-S(k)}{a(k)}, \tag{19}$$

The obtained $w'(k)$ are de-spread with the same quantized chaotic sequence in Eq. (11) in order to detect and recover the hidden binary data. The inserted binary data can be obtained by:

$$b'(k)=\sum_{k=1}^{N_w} w'(k)\oplus\beta(k), \tag{20}$$

In order to detect the existence of the inserted binary data, a similarity measure is used to decide whether there is inserted watermark data or not which is defined as:

$$Sim(b,b')=\frac{\sum_{k=1}^{N_w} b(k)\times b'(k)}{\sqrt{\sum_{k=1}^{N_w} b(k)^2}\sqrt{\sum_{k=1}^{N_w} b'(k)^2}}, \tag{21}$$

If $Sim(b,b')>\delta$, where $\delta$ is the decision threshold, then the watermark exists and output the watermark message. Otherwise there will be no output data.

## 7. Performance analysis

### 7.1. Error analysis

Two types of errors may occur while searching for the watermark sequence: the false positive error (i.e., false watermark detection) and the false negative error (i.e., failure to detect an existing watermark). It is rather difficult to give an exact probabilistic model of false positive and negative errors. Here, a simplified model is utilized based on binomial probability distribution to provide an analysis to estimate the probability of a false positive and negative error for the proposed technique. For comparing the similarities between

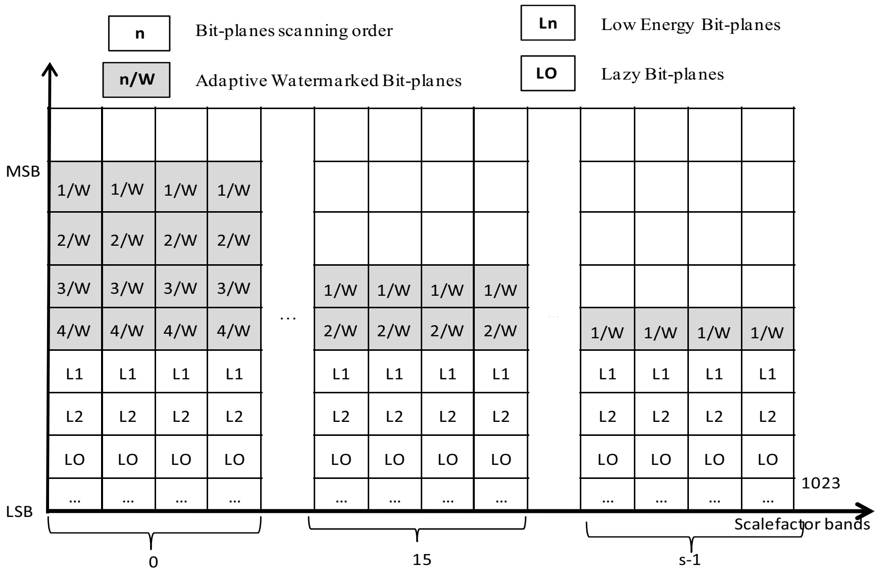

**Fig. 4.** Watermark embedding with perceptual model.

$$b'(k) = \sum_{k=1}^{N_w} w'(k) \oplus \beta(k),$$

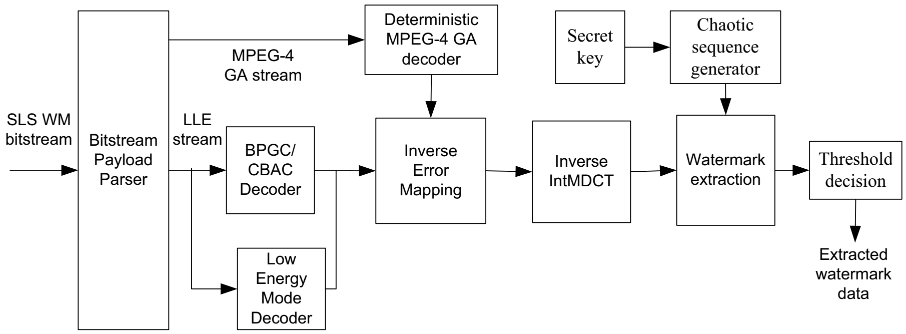

**Fig. 5.** Diagram of watermark message extraction.

the original and extracted watermark signals, the normalized cross-correlation coefficient (NC) is used, which is computed as:

$$NC(w, w') = \frac{\sum_{l} w(i)w'(i)}{\sqrt{\sum_i w^2(i)}\sqrt{\sum_i w'^2(i)}}. \tag{22}$$

(1) False positive error

The probability of false watermark detection is defined as:

$$P_{fp} = P\{NC(w, w') \geq T_P | \text{no watermark}\}, \tag{23}$$

where $P\{A/B\}$ is the probability of event $A$ given event $B$, $T_P$ is a threshold. Since $w(i)$ and $w'(i)$ are either 0 or 1, correspondingly, $w^2(i)$ and $w'^2(i)$ are either 0 or 1, NC can be rewritten as:

$$NC(w, w') = \frac{\sum_l w(i)w'(i)}{\sqrt{\sum_i w^2(i)}\sqrt{\sum_i w'^2(i)}} \geq \frac{\sum_i w(i)w'(i)}{M}, \tag{24}$$

where $M$ is the length of the binary watermark in 1-D format. To decide whether the watermark is present or not, NC is compared with a threshold. If NC is larger than a minimum threshold, then the watermark is present. Suppose $n$ bits of errors occur when extracting watermark data, then $m = M - n$ bits are identical to the original watermark data. It is easy to derive:

$$\sum_i w(i)w'(i) = m - n = M - 2n. \tag{25}$$

If Eqs. (25) and (24) are substituted into Eq. (23), the probability of false positive error can be further derived as:

$$\begin{aligned} P_{fp} &= P\{NC(w, w') \geq T_\rho | \text{no watermark}\} \\ &= \left\{ \frac{\sum_i w(i)w'(i)}{M} \geq T_\rho | \text{no watermark} \right\} \\ &= \left\{ \frac{M - 2n}{M} \geq T_\rho | \text{no watermark} \right\}. \end{aligned} \tag{26}$$

when $n \le M(1-T_\rho)/2$, the false positive error will occur, which is calculated as:

$$P_{fp} = \sum_{n=0}^{\lceil M(1-T_\rho)/2\rceil} P\left\{\sum_i w(i)w'(i) = M-2n|\text{no watermark}\right\} = \sum_{n=0}^{\lceil M(1-T_\rho)/2\rceil} \begin{pmatrix} M \\ n \end{pmatrix} p^n(1-p)^{M-n}, \qquad (27)$$

where $\begin{pmatrix} M \\ n \end{pmatrix} = (M!/n!(M-n)!)$, $p$ is the probability of error in BER when some watermark extraction is performed, which is defined as:

$$BER = \frac{1}{M}\sum_{i=1}^{M} w(i) \oplus w'(i), \qquad (28)$$

where $\oplus$ is the exclusive or (*XOR*) operator.

Since watermark values are either 0 or 1, thus $p = 0.5$, then

$$P_{fp} = \sum_{n=0}^{\lceil M(1-T_\rho)/2\rceil} \begin{pmatrix} M \\ n \end{pmatrix} 0.5^M. \qquad (29)$$

(2) False negative error

False negative error is the probability of declaring a watermarked audio as an unwatermarked one. Less false negative errors imply a better watermarking scheme. Similarly, the probability of false negative error can be determined as:

$$P_{fn} = P\{NC(w, w') < T_\rho|\text{no watermark}\} = \sum_{n=\lceil M(1-T_\rho)/2\rceil+1}^{M} \begin{pmatrix} M \\ n \end{pmatrix} p^n(1-p)^{M-n}, \qquad (30)$$

where $p$ is the BER of the extracted watermark.

In the proposed scheme, the false positive and negative rate is nonlinear with respect to the watermark length. From the above analysis, when the watermark length is 1024 in the proposed scheme, the probability of the false positive and negative error is approximately to zero, which indicates the good performance of this scheme.

### 7.2. Security analysis

The security of the embedded watermark should be considered in a watermarking system for copyright protection due to the fact that, if an attacker guesses the positions of the embedded watermark successfully, the embedded watermark can be easily detected and altered. This section describes the security analysis since security is an important component of a secure watermarking scheme. To improve the confidentiality, the key space should be large enough to make the attacks by brute force infeasible. In the digital world, 24-bit are often used to represent a floating-point number (i.e., initial condition) due to limitation of numerical precision of a computer. In fact, the security of a watermarking scheme usually depends on keys rather than the privacy of scheme. In this proposed audio watermarking scheme, key $K_1$ is used to generate logistic sequence for enhancing the security of the proposed scheme. Therefore, the size of key value space affects the confidentiality of the proposed scheme. The key value space $K_1$ is calculated as follows: suppose $K_1 = \{0 < K_1(i) < 1|i = 1, 2, \ldots, N_1\}$, $N_1$ is an integer which is used to generate the logistic sequence, hence it should be large enough to produce chaotic sequences $Y = \{y(i, j)|i = 1, 2, ..., N_1, j = 1, 2, ..., N_2\}$, where $N_1$ denotes the number of chaotic sequences and $N_2$ represents the length of each chaotic sequence. When $K_1' = \{0 < K_1(i) + d < 1|i = 1, 2, \ldots, N_1\}$, it will generate another group of chaotic sequences $Y' = \{y'(i, j)|i = 1, 2, \ldots, N_1, j = 1, 2, \ldots, N_2\}$. The function $f_k = SS(d)$ is used to test key space of $K_1$ as below:

$$f_k = SS(d) = \frac{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}|y(i, j) - y'(i, j)|}{N_1 \times N_2}. \qquad (31)$$

Fig. 6 plots the function of $f_k = SS(d)$. It can be seen that $f$ is equal to 0 when $d_0 = 10^{-17}$ in this method. Thus the key space of $K_1$ is $1/d_0 = 10^{-17}$, which means that there is enough key space to guarantee high confidentiality of the proposed watermarking system. Based on this security analysis, it can come to a conclusion that the embedded watermarks are secure to attackers who try to exhaustively or statistically detect and read them. All in all, the proposed scheme with such a long key is enough for reliability and practical usage.

**Table 1**
MPEG-4 audio test sequences (16 bits stereo).

| Index | Item (.wav) | Index | Item (.wav) | Index | Item (.wav) |
|---|---|---|---|---|---|
| 1 | avemaria | 6 | cymbal | 11 | haffner |
| 2 | blackandtan | 7 | dcymbals | 12 | mfv |
| 3 | broadway | 8 | etude | 13 | unfo |
| 4 | cherokee | 9 | flute | 14 | violin |
| 5 | clarinet | 10 | fouronsix | 15 | waltz |

## 8. Experimental tests and results

The main audio editing and attacking tools adopted in the experiments are GoldWave (v5.25), CoolEditPro (v2.1), and EAQUAL v0.1.3 alpha for measuring the objective difference grade (ODG) value. Some signal processing functions are implemented in MATLAB. A total of 15 MPEG-4 standard test sequences (16 bits/sample, 48 kHz and WAV format) listed in Table 1 are used to analyze the performance of the watermarking algorithm.

### 8.1. Time and spectrum waves

Tests were run to evaluate whether the watermarked data in MPEG-4 SLS audio can be detected through more qualitative methods, and the characteristics of the changes. This research analyzes the sound waveforms in the time and frequency domains. The original SLS audio waveforms are compared with the watermarked audio containing the hidden data. The resulting waveforms in the time domain and the power spectrum density (PSD) in the frequency domain are shown in Figs. 7 and 8 respectively. Both plots do not show distinguishing visual differences between the original audio and the watermarked audio. However, there seems to be a slight amount of signal loss and a slight increase in distortion due to the hidden data.

### 8.2. Watermark robustness

In the experiment, the watermark robustness is evaluated in terms of BER after the watermark extraction. The robustness performance after the following common signal processing manipulations are evaluated and tested: (1) MP3, AAC and SLS compression of watermarked PCM audio attacks; (2) low pass filtering (LPF) (with cut-off frequency 44.1 kHz); (3) requantization test (16–8–16 bits); (4) delay (500 ms, 10%); (5) echo addition (500 ms, 10%); (6) resampling (44.1–22.05–44.1 kHz); (7) equalization; (8) noise addition (add white noise to 20 dB SNR). The BER results after the common signal processing attacks are shown in Table 2. It can be seen from the table that the BER results of the proposed scheme are quite satisfactory.

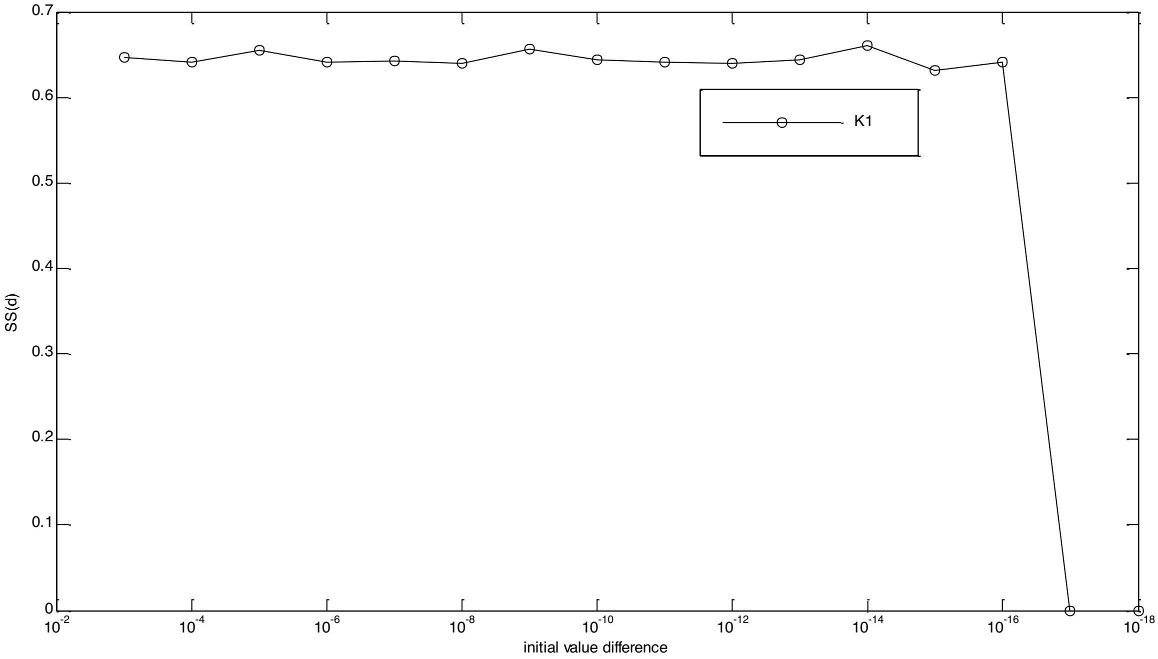

**Fig. 6.** Key space under different initial value difference.

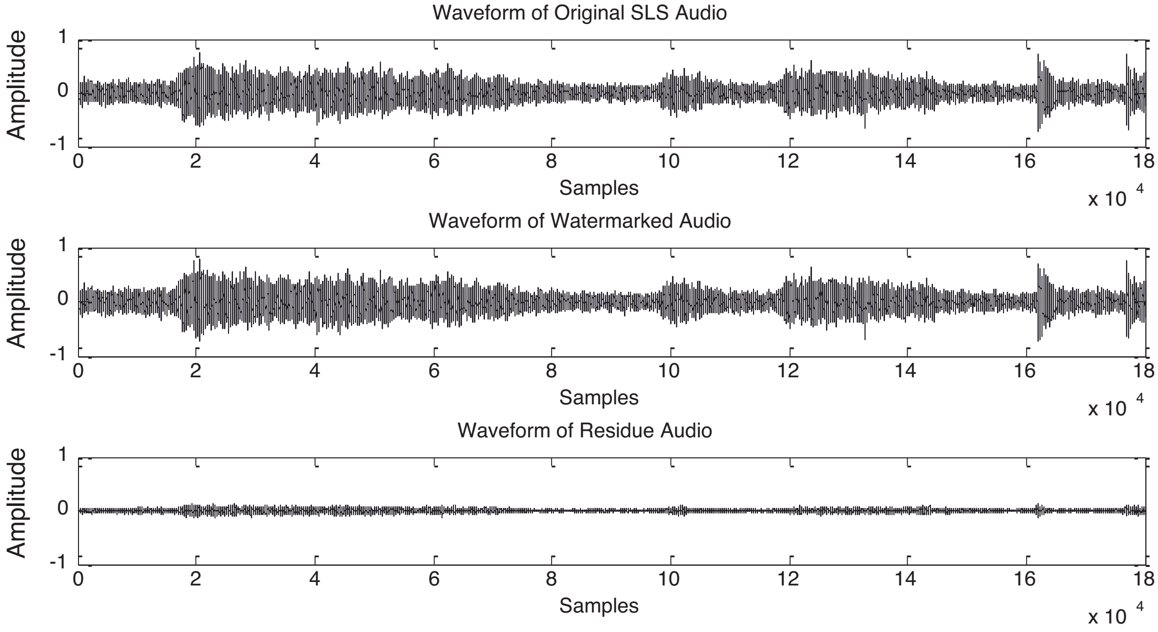

**Fig. 7.** Original, watermarked and residue audio waveform in time domain.

### 8.3. Objective evaluation test

ITU-R Recommendation standard BS.1387 is used to measure the audible distortion of the embedding process and provides the definition for perceptual evaluation of audio quality (PEAQ). PEAQ is used for rating of the quality of audio signal modifications performed by a watermark embedder. EAQUAL tool [23,28] implements PEAQ and computes the ODG and noise mask ratio (NMR) by means of a neural network. The computed ODG values are in a range of [− 4, 0], where −4 is the worst (very annoying) and 0 is the best (imperceptible). Specifically, the embedding distortion caused by the proposed watermarking algorithm expressed with the ODG value is very close to the threshold of becoming audible as shown in Table 3. The NMR is relatively high indicated the robustness of the watermarking scheme. The main reason of having satisfactory ODG is due to the fact that only a small part of samples are modified for watermarking and only small modifications of the magnitudes are being made.

### 8.4. Subjective listening test

In the subjective listening tests, the subjects are asked to discern the watermarked and host audio clips. The evaluation of the subjective quality of the watermarked material was done by a MUSHRA listening test, which stands for “Multi Stimulus test with Hidden Reference and Anchors” and is described in ITU-BS.1534 [29].

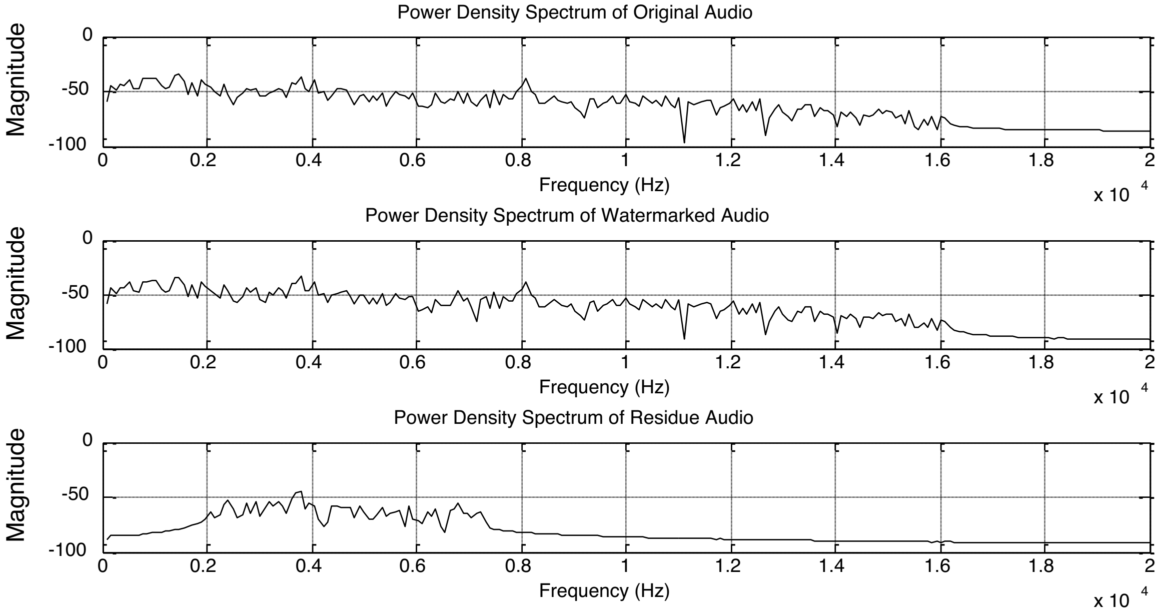

**Fig. 8.** PSD of original, watermarked and residue audio.

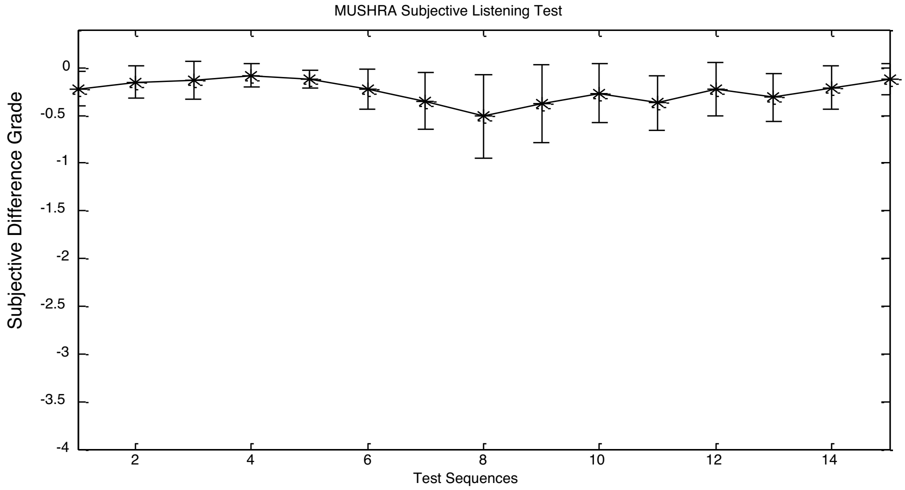

**Fig. 9.** Subjective listening test results.

**Table 2**
BER of watermark extraction under attacks.

| Item (.wav) | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| No manipulation | 0 | 0 | 0 | 0 | 0 | 0 |
| MP3@128 kbps | 0 | 0 | 0 | 0 | 0 | 0 |
| MP3@64 kbps | 0.002 | 0.002 | 0.001 | 0.04 | 0.05 | 0 |
| AAC@128 kbps | 0 | 0 | 0 | 0 | 0 | 0 |
| AAC@64 kbps | 0 | 0 | 0 | 0 | 0.003 | 0 |
| SLS@128 kbps | 0 | 0 | 0 | 0 | 0 | 0 |
| SLS@96 kbps | 0 | 0 | 0 | 0 | 0 | 0 |
| SLS@64 kbps | 0 | 0 | 0 | 0 | 0 | 0 |
| LPF | 0.001 | 0.001 | 0 | 0 | 0.001 | 0 |
| Requantization | 0 | 0 | 0 | 0 | 0 | 0 |
| Delay | 0.005 | 0.005 | 0 | 0.007 | 0.008 | 0 |
| Echo addition | 0 | 0 | 0 | 0 | 0 | 0 |
| Resampling | 0.021 | 0.022 | 0.02 | 0.036 | 0.048 | 0 |
| Equalization | 0 | 0 | 0 | 0 | 0 | 0 |
| Noise addition | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 3**
Objective evaluation test results.

| Item (.wav) | ODG | NMR | Item (.wav) | ODG | NMR | Item(.wav) |
|---|---|---|---|---|---|---|
| 1 | −0.10 | −30.49 | 9 | −0.11 | −30.15 | 1 |
| 2 | −0.11 | −30.18 | 10 | −0.10 | −30.87 | 2 |
| 3 | −0.09 | −31.67 | 11 | −0.14 | −28.94 | 3 |
| 4 | −0.14 | −28.22 | 12 | −0.11 | −32.11 | 4 |
| 5 | −0.15 | −26.78 | 13 | −0.09 | −33.02 | 5 |
| 6 | −0.08 | −34.23 | 14 | −0.13 | −29.67 | 6 |
| 7 | −0.11 | −30.28 | 15 | −0.16 | −32.15 | 7 |

The watermarked signal is graded with respect to the host signal according to a five-grade impairment scale defined in ITU-R BS.562, called subjective difference grade (SDG), which equals to the subtraction between the subjective ratings given separately to the host and watermarked signal. The test sequences have been used extensively for assessment of subjective sound quality in the MPEG-4 audio development process. To give an anchor for comparison, the

**Table 4**
Watermark payload for standard test sequences.

| Item (.wav) | Payload (bps) | Item (.wav) | Payload (bps) | Item (.wav) | Payload (bps) |
|---|---|---|---|---|---|
| 1 | 4 | 6 | 9 | 11 | 4 |
| 2 | 5 | 7 | 10 | 12 | 2 |
| 3 | 8 | 8 | 3 | 13 | 5 |
| 4 | 5 | 9 | 3 | 14 | 4 |
| 5 | 3 | 10 | 5 | 15 | 6 |

quality of the watermarked items is compared with the quality of original items. Besides, the ten listeners who are both experienced and familiar with the set of test items are invited to perform the test. The results of the listening test are shown in Fig. 9. As can be seen from the above results, the quality degradation of the bitstream watermarking system is very small for the vast majority of the test items. For all items, the small variations of the signals indicate that there is no significant distortion introduced by this scheme.

### 8.5. *Watermark payload*

Table 4 shows the data payload results of all the test sequences. As different test sequences have different features which determine how much watermark data can be embedded, the data payload differs accordingly.

## 9. Conclusions

In this paper, the perception and chaos based audio watermarking scheme in the compressed bitstream is proposed to provide a feasible and effective audio copyright protection scheme for MPEG-4 SLS audio signal. Performance analysis shows good property of the watermarking system. Several objective and subjective tests are carried out to analyze the performance of the proposed scheme for various signal manipulations and standard benchmark attacks. The experimental results show that the proposed scheme is inaudible and robust against common signal processing and Stirmark benchmark attacks.

## Acknowledgement

This work was supported partly by National Natural Science Foundation of China (Nos. 61402296, 61101026, 61372006, 60871060, 61031003, 61427806 and 81270707), Shenzhen Key Basic Research Project (Nos. 201101013, 201202076, JCYJ20130329105033277, and JSE201109150013A), China Postdoctoral Science Foundation Funded Project (Nos. 2013M540663 and 2014T70824) and National Natural Science Foundation of Guangdong Province (No. S2013040014448).

## References

[1] ISO./IEC. Information technology – MPEG audio technologies – Part 1: MPEG surround; 2007.
[2] Giovanardi A, Mazzini G, Tomassetti M. Chaos based audio watermarking with MPEG psychoacoustic model I. In: Proceeding of Joint Conference of the Fourth International Conference onInformation, Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia, vol.1603. 2003. p. 1609–13.
[3] Lei B, Soon IY, Zhou F, Li Z, Lei H. A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition. Signal Process 2012;92(9):1986–2001.
[4] Lei B, Soon IY, Tan EL. Robust SVD-based audio watermarking scheme with differential evolution optimization. IEEE Trans Audio Speech Lang Process 2013;21(11):2368–78.
[5] Lei B, Soon IY, Li Z. Blind and robust audio watermarking scheme based on SVD–DCT. Signal Process 2011;91(8):1973–84.
[6] Lei B, Song I, Rahman SA. Robust and secure watermarking scheme for breath sound. J Syst Softw 2013;86(6):1638–49.
[7] Lei B, Song I, Rahman SA. Optimal watermarking scheme for breath sound. In: Proceedings of International Joint Conference on Neural Networks. 2012. p. 1–6.
[8] Mooney A, Keating JG, Pitas I. A comparative study of chaotic and white noise signals in digital watermarking. Chaos Solitons Fractals 2008;35(5): 913–21.
[9] Gruhl D, Lu A, Bender W. Echo hiding. Lect Notes Comput Sci 1996;1174:295–315.
[10] Chen OTC, Wen-Chih W. Highly robust, secure, and perceptual-quality echo hiding scheme. IEEE Trans Audio Speech Lang Process 2008;16(3):629–38.
[11] Ko BS, Nishimura R, Suzuki Y. Time-spread echo method for digital audio watermarking. IEEE Trans Multimed 2005;7(2):212–9.
[12] Yeo IK, Kim HJ. Modified patchwork algorithm: a novel audio watermarking scheme. IEEE Trans Speech Audio Process 2003;11(4):381–6.
[13] Kirovski D, Malvar HS. Spread-spectrum watermarking of audio signals. IEEE Trans Signal Process 2003;51(4):1020–33.
[14] Seok J, Hong J, Kim J. A novel audio watermarking algorithm for copyright protection of digital audio. ETRI J 2002;24:3.
[15] Swanson MD, Zhu B, Tewfik AH, Boney L. Robust audio watermarking using perceptual masking. Signal Process 1998;66(3):337–55.
[16] Neubauer C, Herre J. Audio watermarking of MPEG-2 AAC bitstream. In: Proceedings of 108th AES Convention. 2000.
[17] Tachibana R, Shimizu S, Kobayashi S, Nakamura T. An audio watermarking method using a two-dimensional pseudo-random array. Signal Process 2002;82(10):1455–69.
[18] Quan X, Zhang H. Data hiding in MPEG compressed audio using wet paper codes. In: Proceedings of International Conference on Pattern Recognition. 2006. p. 727–30.
[19] Cheng S, Yu H, Xiong Z. Enhanced spread spectrum watermarking of MPEG-2 AAC audio. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing Institute of Electrical and Electronics Engineers Inc. 2002. p. IV/3728–31.
[20] Li Z, Sun Q, Lian Y. Design and analysis of a scalable watermarking scheme for the scalable audio coder. IEEE Trans Signal Process 2006;54(8): 3064–77.
[21] Geiger R, Yokotani Y, Schuller G. Audio data hiding with high data rates based on IntMDCT. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. 2006. V-V.
[22] Bassia P, Pitas I, Nikolaidis N. Robust audio watermarking in the time domain. IEEE Trans Multimed 2001;3(2):232–41.
[23] Boney L, Tewfik AH, Hamdy KN. Digital watermarks for audio signals. In: Proceedings of the International Conference on Multimedia Computing and Systems. 1996. p. 473–80.
[24] Malik H, Ansari R, Khokhar A. Robust audio watermarking using frequency-selective spread spectrum. IET Inf Secur 2008;2(4):129–50.
[25] Yu R, Rahardja S, Lin C-C, Ko CC. A fine granular scalable to lossless audio coder. IEEE Trans Audio Speech Lang Process 2006;14(4):1352–63.
[26] Geiger R, Yu R, Herre J, Rahardja S, Kim SW, Lin X, et al. ISO/IEC MPEG-4 high-definition scalable advanced audio coding. J Audio Eng Soc 2007;55(1–2):27–43.
[27] Yu R, Geiger R, Rahardja S, Herre J, Lin X, Huang H. MPEG-4 scalable to lossless audio coding. In: Proceedings of Audio Engineering Society Convention, vol. 117. 2004.
[28] http://wiki.hydrogenaudio.org/index.php?title=EAQUAL
[29] Kozamernik F, Stoll G. EBU listening tests on Internet audio codecs. EBU Tech Rev 2000;283:20–33.