AN IMPROVED PATCHWORK-BASED DIGITAL AUDIO WATERMARKING IN CQT DOMAIN

Peng Hu, *Qin Yan*, *Luan Dong*, *Meng Liu*

College of Computer and Information, Hohai University, Nanjing, China
hupeng4412@gmail.com, yan_qin@hhu.edu.cn, phd.luan.dong@gmail.com, liumeng.hhu@gmail.com

ABSTRACT

In nowadays digital audio watermarking still remains one of the hot research topics in the view of multimedia copyright protection. In this paper an improved patchwork-based audio watermarking algorithm in Constant-Q Transform (CQT) domain has been proposed. The advantage of CQT in music analysis lies in its nonlinear frequency spacing. However the absence of exact invertible transform still prevents CQT from the wide application. In this paper it is overcome by frame pair selection, which is carefully performed according to frame pair energy ratios of the middle frequency range to avoid the disturbance of watermarking embedding and degradation in signal quality afterwards. Watermarks are then embedded by modifying the energy of selected frame pairs. The experimental results indicate that the proposed method outperforms the latest patchwork-based audio watermarking algorithm in Discrete Cosine Transform (DCT) domain with better signal quality of embedded signals and yet more robust to the conventional attacks.

Index Terms— Audio watermarking, CQT, ICQT, frame pairs, patchwork

1. INTRODUCTION

With the rapid development of multimedia and social network, digital data can be easily distributed without any authorization. Hence watermarking remains one of important techniques for copyright and integrity authentication [1-3]. Main requirements of digital audio watermarking are imperceptibility, robustness and data capacity [4]. Current audio watermark schemes tend to play a delicate balance between imperceptibility and robustness. Different techniques have been employed in audio watermarking schemes, such as echo hiding [5], vector regression [1], spread spectrum [6] and patchwork [7,8].

Among watermarking methods, the patchwork-based schemes present great performance to resisting conventional attacks and show good imperceptibility [8]. The technique of patchwork watermarking is originated from image processing [9]. Later it is extended to audio signals by Arnold [7], where the patchwork method embeds watermarks by modifying Discrete Fourier Transform (DFT) coefficients in patches formed by audio signal segments. However, the method is only feasible assuming that those patches have same statistical property which is not always be satisfied. In [8], patchwork-based audio watermarking scheme in DCT domain is proposed where two patches and many frame pairs are produced for each audio segment. If a frame pair meets the same limitation, watermarks will be embedded by modifying their DCT coefficients. Nevertheless, the robustness of this method against most of traditional attacks is ensured only when embedding watermarks multiple times which requires large data capacity inevitably.

CQT is essentially a transform of a signal from time domain to frequency domain such that the center frequencies of the frequency bins are geometrically spaced and their Q-factors are all equal [10]. Accordingly the advantage of CQT over conventional DFT in analysis of musical signals lies in that it gives better frequency resolution for low frequencies and high time resolution for high frequencies. This is in contrast with linearly spaced frequency bins in the conventional DFT which fails to satisfy the varying time and frequency resolution requirements over the wide range of audible frequencies. Although there is no exact invertible CQT (ICQT) available, a computationally efficient algorithm for computing CQT with flexible bins per octave and assured quality of reconstructed audio contents is proposed by Schörkhuber and Klapuri [10]. In this paper we apply an improved patchwork-based audio watermarking algorithm in CQT domain. In the proposed scheme, watermarks are only embedded into the selected frame pairs according to their energy in the octave of middle frequency range to ensure its inaudibility.

The rest of the paper is organized as follows. Section 2 describes CQT briefly followed by an overview of the whole system in section 3. The details of watermark embedding and extraction are presented in Section 4 and 5. Experiments and analysis are provided in Section 6. Finally section 7 concludes the paper.

2. CONSTANT-Q TRANSFORM

CQT was initially proposed by Brown to calculate a log-frequency spectrogram [11]. Later an approximate and effi-

Thanks to National Natural Science Foundation of China for funding (No.61170297) and Qing Lan Project of Jiangsu Province of China.

cient CQT and ICQT computation was proposed in [10]. We brief it here and more details can be found in [10]. CQT can be written as

$$\mathbf{X}_{\boldsymbol{d}}^{\mathbf{CQ}} = \mathbf{A}^{*}\,\mathbf{X}_{\boldsymbol{d}}. \tag{1}$$

where d=1,2,…,D indexes the number of octaves calculated, matrix $\mathbf{A}^{*}$ is the conjugate of $\mathbf{A}$ which is used to produces the CQT for the highest octave only, matrix $\mathbf{X}_{\boldsymbol{d}}^{\mathbf{CQ}}$ and $\mathbf{X}_{\boldsymbol{d}}$ contains CQT and DFT coefficients of audio signal in octave d respectively.

The approximate ICQT is then given by

$$\boldsymbol{Y}_{\boldsymbol{d}} = \mathbf{V}^{*}\,\boldsymbol{X}_{\boldsymbol{d}}^{\boldsymbol{CQ}}. \tag{2}$$

where $\mathbf{V}^{*}$ is the conjugate of $\mathbf{V}$**,** $\mathbf{V}$=$\mathbf{A}^{*}$, $\boldsymbol{Y}_{\boldsymbol{d}}$ contains the DFT values of reconstructed signal.

3. OVERVIEW OF THE PROPOSED WEATERMRK SCHEME

The procedures of watermark embedding and extraction are shown in Fig.1(a) and Fig.1(b). Since the high frequency coefficients are vulnerable to attacks and low frequency coefficients are easily audible, CQT coefficients in a middle frequency range is a balanced choice to embed watermarks by modifying the energy of selected frame pairs. At extraction stage the watermark is recovered according to the energy ratios of selected frame pairs of reconstructed signal.

4. WATERMARK EMBEDDING

4.1. CQT frame pairs selection

Assume that octave d corresponding to a middle frequency for inserting watermarks, the key in watermark embedding in CQT domain is to select the proper frame pairs in octave d. Fig.2 shows the process of selecting CQT frame pairs. j^{th} frame of $\mathbf{X}_{\boldsymbol{d}}^{\mathbf{CQ}}$ in octave d is denoted by $X_d^{CQ}(j)$, j=1,2,…,R. R is the total number of frames in octave d. Note different octaves contain different numbers of frames. Due to lack of ICQT allowing perfect reconstruction of the original signal from its transform coefficients, the modification of CQT coefficients of a frame will trigger the modifications of those of its adjacent frames. For instance, the increase of $X_d^{CQ}(j)$ of frame j will raise the energy of adjacent frames. Furthermore, the experiments show the closer to the frame j, the stronger the intensity change of CQT of adjacent frames. Hence the intra-pair distance within a frame pair s is set by (3):

$$s = \lceil N/L \rceil + 1. \tag{3}$$

where N denotes the DFT frame length of $\mathbf{X}_{\boldsymbol{d}}$ and L denotes the frame shift of $\mathbf{X}_{\boldsymbol{d}}$, $\lceil * \rceil$ denotes rounding towards positive infinity. Hence $X_d^{CQ}(j)$ and $X_d^{CQ}(j+s)$ are taken as a frame pair, j=1,2,…,R-s.

Unfortunately not all frame pairs $\{X_d^{CQ}(j), X_d^{CQ}(j+s)\}$ are suitable for watermark embedding due to following three reasons.

(a) The small CQT coefficients (e.g. close to zero) tend to have a weakly stable energy values which are very easily influenced by external changes such as reconstruction of signals. Therefore they are not suitable for watermark embedding.

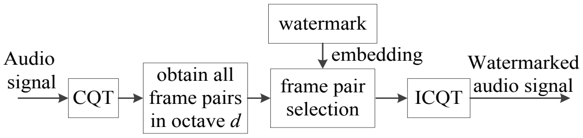

Fig.1(a). The procedure of audio watermark embedding.

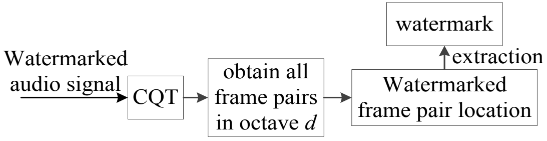
watermark
extraction
Watermarked audio signal
CQT
obtain all frame pairs in octave d
Watermarked frame pair location

Fig.1(b). The procedure of audio watermark extraction.

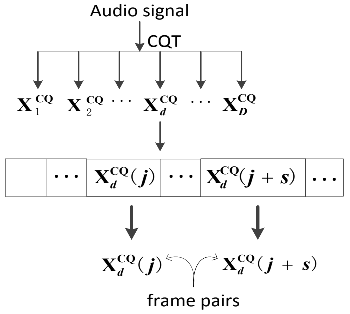
Audio signal
CQT
$\mathbf{X}_1^{CQ}$ $\mathbf{X}_2^{CQ}$ ··· $\mathbf{X}_d^{CQ}$ ··· $\mathbf{X}_D^{CQ}$
··· $\mathbf{X}_d^{CQ}(j)$ ··· $\mathbf{X}_d^{CQ}(j+s)$ ···
$\mathbf{X}_d^{CQ}(j)$
$\mathbf{X}_d^{CQ}(j+s)$
frame pairs

Fig.2. The procedure of forming CQT frame pairs. $X_d^{CQ}(j)$ and $X_d^{CQ}(j+s)$ are the coefficients of frame j and $j+s$ in octave d.

(b) The watermarked frame pairs cannot be located precisely as the modification of CQT coefficients of a frame leads to the energy fluctuation of its adjacent frames and alters the energy ratios of the frame pairs consequently.

(c) If the energy ratio of a frame pair is too high, more efforts are required in modification of CQT coefficients of both frames, which will inevitably degrade the perceptual quality of audio signals considerably.

In order to overcome problem (a), the energy $P_{d,j}$ of frame j in octave d is calculated.

$$P_{d,j} = \sum_{m=1}^{B}\left(X_d^{CQ}(j,m)\right)^2 \tag{4}$$

where j=1,2,…,R-s, B is the number of bins per tave, $X_d^{CQ}(j,m)$ denotes m^{th} coefficient of frame j in octave d, $P_{d,j}$ denotes the energy of frame j in octave d. The energy of frame pairs is then controlled by θ to avoid that they are too small to embed watermark as in (5):

$$\min(P_{d,j}, P_{d,j+s}) \geq \theta. \tag{5}$$

where $P_{d,j+s}$ denotes the energy of frame $j+s$ in octave d. In order to select a suitable value of θ, frames with too large or small energy are eliminated as in (6). In our case, U is set with 10^{-12} according to the energies of CQT frames in octave d. θ is set by (6) , (7) and (8) to ensure the energy of watermarked frame pairs from the non–exact reconstructed signal still satisfies (5) such that the watermark location could be correctly sought at the extraction stage.

$$U \leq P_{d,j}. \tag{6}$$

$$\left| \lfloor \log_{10} P_{d,j} \rfloor - \lfloor \log_{10} P_{d,j+1} \rfloor \right| > 2. \tag{7}$$

$$\theta = 10^{(\lfloor \log_{10} P_{d,j} \rfloor - 1)}. \tag{8}$$

where $\lfloor * \rfloor$ denotes rounding towards negative infinity.

In order to overcome problem (b) and (c), we introduce the inter-pair distance *T* and energy ratio $r_d(j)$ of a frame pair in octave *d* as follows:

$$T \geq 3 \times s. \tag{9}$$

$$1 \leq r_d(j) = \max\left(\frac{P_{d,j}}{P_{d,j+s}}, \frac{P_{d,j+s}}{P_{d,j}}\right) \leq c. \tag{10}$$

where $r_d(j)$ is the maximum energy ratio of frame pair $\{X_d^{CQ}(j), X_d^{CQ}(j+s)\}$. According to the generation of CQT frame pairs and watermark embedding rules in section 4.2, the inter-pair distance *T* is set as no less than 3 times of intra-frame distance to control the disturbance between preceding and following frame pairs which may lead to the false alarm detection. In our case, *c* is experimentally set as 3 to avoid the high energy of selected frame pairs to degrade the quality of watermarked audio signals. Only frame pairs satisfying criteria (5), (9) and (10) would be chose to insert watermark.

4.2. Watermark embedding

Before the watermark embedding, the eliminated frame pairs are slightly modified to ensure their energy ratios $r_d(j)$ is not within the range of selected frame pairs after non-exact signal reconstruction.

Assume that i^{th} frame pairs $\{X_d^{CQ}(i), X_d^{CQ}(i+s)\}$ is chosen, the watermark embedding is accomplished by altering the energy ratio of $X_d^{CQ}(i)$ and $X_d^{CQ}(i+s)$. Let $\hat{X}_d^{CQ}(i,m)$ and $\hat{X}_d^{CQ}(i+s,m)$ be the modified coefficients of $X_d^{CQ}(i,m)$ and $X_d^{CQ}(i+s,m)$ respectively. $\hat{P}_{d,i}$ and $\hat{P}_{d,i+s}$ denote the modified energy of $P_{d,i}$ and $P_{d,i+s}$ respectively. The embedding rules are:

Embed watermark bit "0":

- If $\frac{P_{d,i+s}}{P_{d,i}} \in [\frac{1}{3}, \frac{1}{2}]$, $\hat{P}_{d,i} = \frac{P_{d,i+s}}{r1}$, $\hat{P}_{d,i+s} = r1 \times P_{d,i}$.
- If $\frac{P_{d,i+s}}{P_{d,i}} \in (\frac{1}{2}, 1]$, $\hat{P}_{d,i} = \frac{P_{d,i+s}}{r2}$, $\hat{P}_{d,i+s} = r2 \times P_{d,i}$.
- If $\frac{P_{d,i+s}}{P_{d,i}} \in (1, \frac{3}{2}]$, $\hat{P}_{d,i} = P_{d,i}$, $\hat{P}_{d,i+s} = r3 \times P_{d,i}$.
- If $\frac{P_{d,i+s}}{P_{d,i}} \in (\frac{3}{2}, 2]$, $\hat{P}_{d,i} = P_{d,i}$, $\hat{P}_{d,i+s} = P_{d,i+s}$.
- If $\frac{P_{d,i+s}}{P_{d,i}} \in (2,3]$, $\hat{P}_{d,i} = r4 \times P_{d,i+s}$, $\hat{P}_{d,i+s} = P_{d,i+s}$.

Embedding of watermark bit "1":

- If $\frac{P_{d,i}}{P_{d,i+s}} \in [\frac{1}{3}, \frac{1}{2}]$, $\hat{P}_{d,i} = r1 \times P_{d,i+s}$, $\hat{P}_{d,i+s} = \frac{P_{d,i}}{r1}$.
- If $\frac{P_{d,i}}{P_{d,i+s}} \in (\frac{1}{2}, 1]$, $\hat{P}_{d,i} = r2 \times P_{d,i+s}$, $\hat{P}_{d,i+s} = \frac{P_{d,i}}{r2}$.
- If $\frac{P_{d,i}}{P_{d,i+s}} \in (1, \frac{3}{2}]$, $\hat{P}_{d,i} = r3 \times P_{d,i+s}$, $\hat{P}_{d,i+s} = P_{d,i+s}$.
- If $\frac{P_{d,i}}{P_{d,i+s}} \in (\frac{3}{2}, 2]$, $\hat{P}_{d,i} = P_{d,i}$, $\hat{P}_{d,i+s} = P_{d,i+s}$.
- If $\frac{P_{d,i}}{P_{d,i+s}} \in (2,3]$, $\hat{P}_{d,i} = P_{d,i}$, $\hat{P}_{d,i+s} = r4 \times P_{d,i}$.

where the watermarking parameters *r1*, *r2*, *r3* and *r4* are positive constants satisfying

$$r1 > 0.5,\ r2 > 1,\ r3 > 1.5,\ r4 > 0.5 \tag{11}$$

The embedding rules are designed above by altering the energies of each frame in selected frame pair towards an opposite orientation such that the watermark can be obtained by energy ratios. The corresponding CQT coefficients are modified as (12) and (13):

$$\hat{X}_d^{CQ}(i,m) = X_d^{CQ}(i,m) \times \sqrt{\frac{\hat{P}_{d,i}}{P_{d,i}}} \tag{12}$$

$$\hat{X}_d^{CQ}(i+s,m) = X_d^{CQ}(i+s,m) \times \sqrt{\frac{\hat{P}_{d,i+s}}{P_{d,i+s}}} \tag{13}$$

5. WATERMARK EXTRACTION

The watermark extraction procedure is basically a reverse order of watermark embedding including location of watermarked frame pairs and recovering of the embedded bits without the original audio signals. The energy of the watermarked audio signal in octave *d* is obtained by (14):

$$P'_{d,j} = \sum_{m=1}^{B} \left(Y_d^{CQ}(j,m)\right)^2 \tag{14}$$

where $Y_d^{CQ}(j,m)$ denotes m^{th} CQT coefficient of frame *j* in octave *d* of reconstructed signal. The watermark extraction is listed as follows:

- Locate the watermarked frame pairs: If $\{Y_d^{CQ}(j), Y_d^{CQ}(j+s)\}$ satisfy (5), (9) and (10), the frame pair has been watermarked, *j*=1,2,…,*R*.
- Recover the watermark bits: If the energy of the watermarked frame pair $\{Y_d^{CQ}(i), Y_d^{CQ}(i+s)\}$ satisfies $P'_{d,i} > P'_{d,i+s}$, the embedded watermark bit is "1". Otherwise, the watermark bit is "0".

6. RESULTS AND ANALYSIS

In this section, several experiments are conducted to compare the imperceptibility and robustness of the proposed algorithm and [8]. In total four types of mono-channel audio signals (western pop, eastern classic, south Asian folk, subcontinent country) have been used for tests. Each includes ten 10-second clips, sampled at the rate of 44.1kHz and quantized with 16bits. The watermark is a random sequence distributed by N(0,1) with a length of 50. In proposed scheme, CQT is set with *D*=8 octaves and *B*=12 bins per octave. The DFT frame length *N* and frame shift *L* is set with 128 and 39. The intra-pair distance *s* is 5. As different octaves occupy different frequency ranges leading to fluctuating capacities, the middle octave *d* (*d*=5) is selected to ensure the reasonable watermark payload. The embedding constants are experimentally set as *r1*=2.9, *r2*=2.5, *r3*=3 and *r4*=0.6. In addition the parameters for the method [8] are as below: the half length of frame *M*=15, the number of fragment in each audio segment *R*=40, threshold Θ_{th}=0.0005, the limit frequency f_{limit}= 6KHz and the watermarking parameters *a1*=0.6, *a2*=0.5. Note that all the watermarks are only embedded once.

The imperceptibility of watermarked audio signals and the robustness of the proposed algorithm are evaluated by Perceptual Evaluation of Audio quality (PEAQ) [12] and bit error rate (BER) [8] respectively.

6.1. Imperceptibility Tests

In this section PEAQ compares the audio signal with the watermarked signal with a return of objective difference grade (ODG) value. The closer ODG is to 0, the better the quality of the watermarked signals. The ODG quality of audio signals after two types of watermark embedding is shown in Fig.3. It can be seen that ODG of the proposed scheme is lower than those of [8] in all the occasions by 10.99% on average.

6.2. Robustness Tests

The robustness of the proposed watermark scheme and the baseline system [8] is measured by BER.

$$\text{BER}=\frac{\text{false bits extracted}}{\text{bits embedded}}\times 100\%. \tag{15}$$

They are tested under 8 types of attacks including re-quantization (from 16bits to 8bits), noise (20dB), amplitude (increased by 1.8 times), MP3 (128kbps), AAC (128kbps), re-sampling (from 44.1kHz to 16kHz, then back to 44.1kHz), high-pass filtering (100Hz) and low-pass filtering (8KHz). As shown in Table 1, it is evident that the proposed algorithm gives better performance than the method in [8] in all attacks. In particular it has lower BER than those of [8] under high-pass and low-pass attacks. This might be due to the impact of Q-factor which gives different resolution for different frequency ranges.

7. CONCLUSION

This paper proposes an improved patchwork-based audio watermarking scheme in CQT domain. To improve robust ness and ensuring imperceptibility, watermarks are only embedded into audio signals by modifying energy rate selected CQT frame pairs in certain octave. The results suggest that the proposed system achieves better ODG than [8] in terms of watermarked signals. In addition the proposed system gives better BER than that of [8] under all conventional attacks, especially in high-pass and low-pass attacks. This might be due to the nonlinear characteristics of CQT. Currently the test database is limited to 4 types of audio signals with 40 clips in total. In future the algorithm will be tested on a wider range of audio clips to further confirm the stability of watermarking parameters while keeping the quality of audio signals**.**

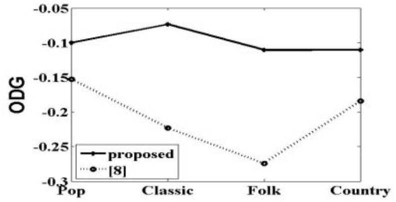

Fig.3. The comparison of ODG of watermarked signals by the proposed scheme and [8]. The ODG is the average value of ten clips in each type.

Attacks	BER (%)							
	Western Pop		Eastern Classic		South Asian Folk		Subcontinent Country	
	[8]	PS	[8]	PS	[8]	PS	[8]	PS
No attack	0	0	0	0	0	0	0	0
Re-quantization	0	0	0	0	0	0	0	0
Noise (20dB)	0.6	0.2	0	0	0.2	0.2	0	0
Amplitude	0	0	0	0	0	0	0	0
MP3 (128kbps)	0	0	0	0	0	0	0	0
AAC (128kbps)	0	0	0	0	0	0	0	0
Re-sample	0	0	0	0	0	0	0	0
High-pass (100Hz)	0.6	0	0.8	0	0	0	0.4	0
Low-pass (8KHz)	1	0	1.2	0	1	0	0.6	0

Table 1. Comparison of BER of the proposed scheme (PS) and the method in [8]. The BER is the average value of ten clips in each type.

REFERENCES

[1] X. Y. Wang, W. Qi, and P. P. Niu, “A new adaptive digital audio watermarking based on support vector regression,” *IEEE Trans. on Audio, Speech and Language Processing*, vol.15, no.8, pp. 2270-2277, Nov. 2007.

[2] I. K. Yeo, H. J. Kim, “Modified patchwork algorithm: A novel audio watermarking scheme,” *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 4, pp. 381-386, July 2003.

[3] N. K, Kalantari, M. A. Akhaee, and S. M. Ahadi, “Robust multiplicative patchwork method for audio watermarking,” *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1133-1141, Aug. 2009.

[4] K. Khaldi, A. Boudraa, “Audio watermarking via EMD,” *IEEE Trans. on Audio, Speech and Language Processing,* vol. 21, no. 3, pp. 675-680, Mar. 2013.

[5] O. T.-C. Chen, W.-C. Wu, “Highly robust, secure, and perceptual-quality echo hiding scheme,” *IEEE Trans. On Audio, Speech and Language Processing*, vol.16, no. 3, pp. 629-638, Mar. 2008.

[6] H. Malik, R. Ansari, and A. Khokha, “Robust audio watermarking using frequency-selective spread spectrum,” *IET Information Security* , vol. 2, no. 4, pp. 129-150, Dec. 2008.

[7] M. Arnold, “Audio watermarking: features, applications and algorithm,” in *Proc. Multimedia and Expo*, New York, 2004, pp. 1013-1016.

[8] I. Natgunanathan, Y. Xiang, and Y. Rong, “Robust patchwork-based embedding and decoding scheme for digital audio watermarking,” *IEEE Trans. on Audio, Speech and Language Processing*, vol. 20, no. 8, pp. 2232-2239, Oct. 2012.

[9] W. Bender, D. Gruhl, and N. Morimoto, “Techniques for data hiding,” *IBM Systems. J*, vol. 35, no. 3&4, pp. 313-335, 1996.

[10] C. Schorkhuber, A. Klapuri, “Constant-Q transform toolbox for music processing,” in *Proc. 7th Sound and Music Computing Conference*, Barcelona, Spain, 2010, pp. 3-64.

[11] J. C. Brown, “Calculation of a constant Q spectral transform”, JASA, vol. 89, no. 1, pp. 425-434, 1991.

[12] P. Kabal, *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*, McGill University, 2003.