ICSV20
20th International Congress on Sound & Vibration
ICSV20
Bangkok,Thailand
7-11 July 2013

ROBUST AUDIO WATERMARKING CAPABLE OF SURVIVING PLAY-RECORD PROCESS BASED ON CHROMA FEATURES

Wang Shuozhong; Feng Guorui; Zou Tingting; Qian Zhenxing
Communication & Information Eng., Shanghai University, Shanghai 200072, China

Wang Runtian
Shanghai Acoustics Laboratory, Chinese Academy of Sciences, Shanghai 200032, China

Chen Yunfei
Key Laboratory for Underwater Test and Control Technology, Dalian 116013, China

e-mail: shuowang@shu.edu.cn

This paper proposes an audio watermarking scheme robust to the play-record process therefore allows watermark extraction with a microphone while the music is being played. The key is the synchronization technique that uses the chroma features representing the semantic contents in the audio. Peaks of the processed chroma sequence are used to locate embedding points, and the audio segmented accordingly. Spectral lines in a chosen frequency range are properly altered to carry the data. Experimental results show effectiveness of the method.

1. Introduction

Audio watermarking is a technique to embed secret data in digital audio signals. The hidden information may be used, for example, to identify copyright of the music or track the source and distribution path of the program. Basic specifications of audio watermarks are embedding capacity, imperceptibility, and robustness against various attacks. These are mutually contradictive therefore require appropriate compromise for specific applications.

Commonly considered attacks include noise contamination and lossy compression [1]. Some authors study such attack as add/remove, filtering and modification [2]. In this paper, we propose a highly robust audio watermarking scheme that can resist play-record attacks. The watermark can be extracted from a played-and-rerecorded version, therefore is suitable to be used in applications that prohibit intrusion in watermark detection.

2. Sematic-based watermark synchronization

The key to the success of play-record tolerant watermarking lies in the reliable synchronization. Since the audio is a stream, one must choose suitable time instances for data embedding, and these embedding points must allow accurate identification upon extraction. If designed improperly, even slight attacks may destroy the synchronization, making the whole system useless. When a

play-record process is involved, things become much more critical as the process can be viewed as a combination of many strong attacks including linear and nonlinear filtering, additive and multiplicative noise contamination, and time-stretching or squeezing. Distortion in the re-recorded signal is so severe that none of the waveform or spectral features can be used to locate the precise embedding spots. Fig. 1 shows the waveforms of a piece of music before and after play-record, and Fig. 2 is their spectrograms. In our experiments that use an ordinary commercial music player and a small microphone, the objective difference grades (ODG) of 10 pieces of re-recorded music are close to minus 4, indicating a poor fidelity.

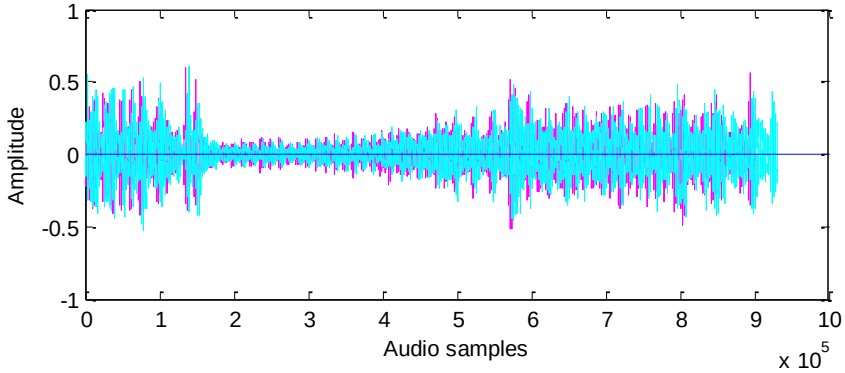
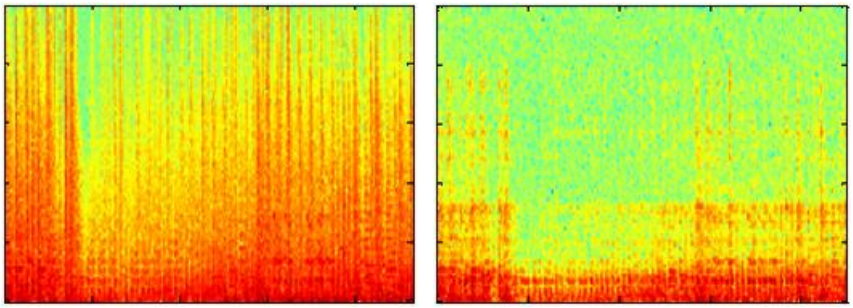


Figure 1. Waveforms before and after play-record.



(a) Original spectrogram (b) After play-record

Figure 2. Spectrograms before and after re-recording.

Although the play-record process causes severe distortion, the semantic contents remain unchanged. Therefore we seek features based on the chroma chracteristics (beat and tempo) to realize synchronization. According to Daniel *et al.*[3], we take the 12 halftones in an octave around 1 kHz and use the corresponding spectral energy as the chroma feature for synchronization in the following steps.

1) Segment the audio signal into L frames, each containing B samples.
2) Calculate FFT of each frame, and get energy of the 12 halftones in 707 Hz~1414 Hz, denoted $C_{i,j}$, ($1 \le i \le L$, $1 \le j \le 12$). Fig. 3 shows halftones in 3600 frames. Each color strip represents a $C_{i,j}$, *i.e.*, energy in that time-frequency strip. Colors represent the values of the energy. Codes provided in [4] are used to calculate the chroma features.
3) A sequence of the 6-th halftone (about 1 kHz), $C_{i,6}$ (i=1, 2, … , L), is taken, and smoothed with a Gaussian convolution mask to produce a curve as shown in Fig. 4. Experiments show that, despite severe distortion of the waveform and spectrum, the major peak locations on the curve are quite stable after play-record.
4) Map the curve to the original time axis, and use the large peaks to locate reference points in the audio waveform for watermark embedding.

It is observed from Fig. 4 that there are still slight shifts in the play-record versions, which inevitably cause errors in watermark extraction.

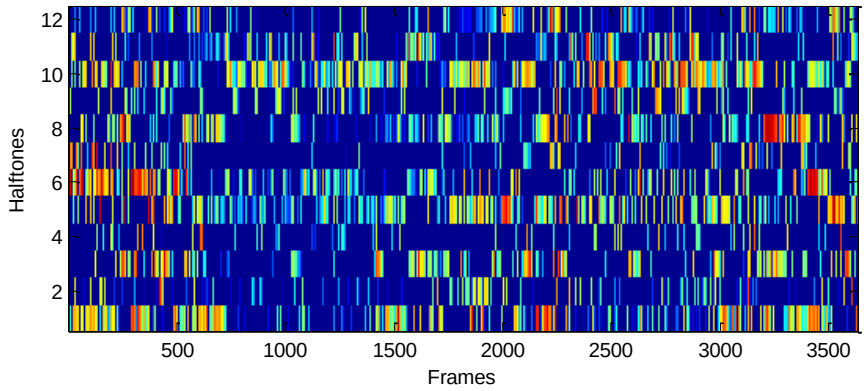


Figure 3. Energy in halftones within an octave. Colors represent $C_{i,j}$ values.

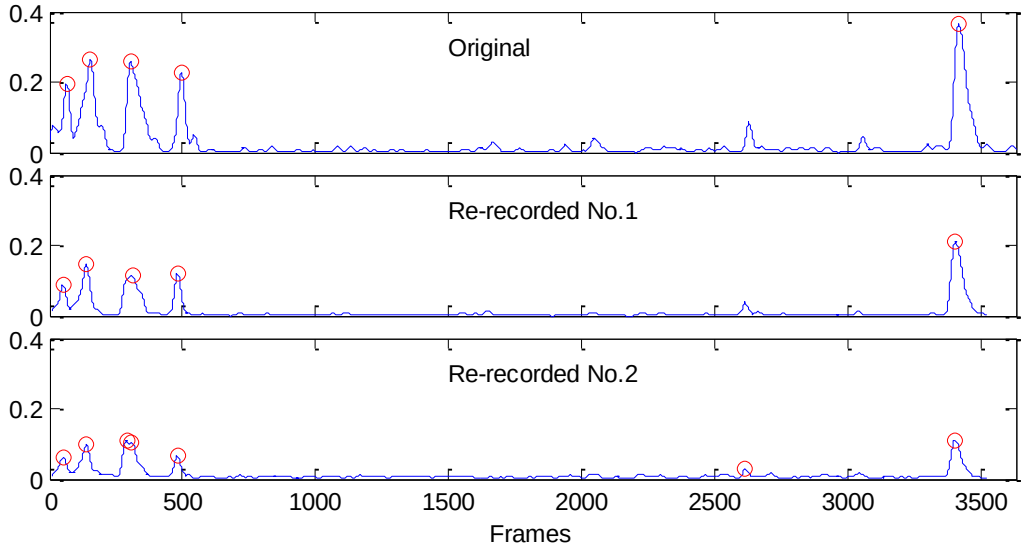


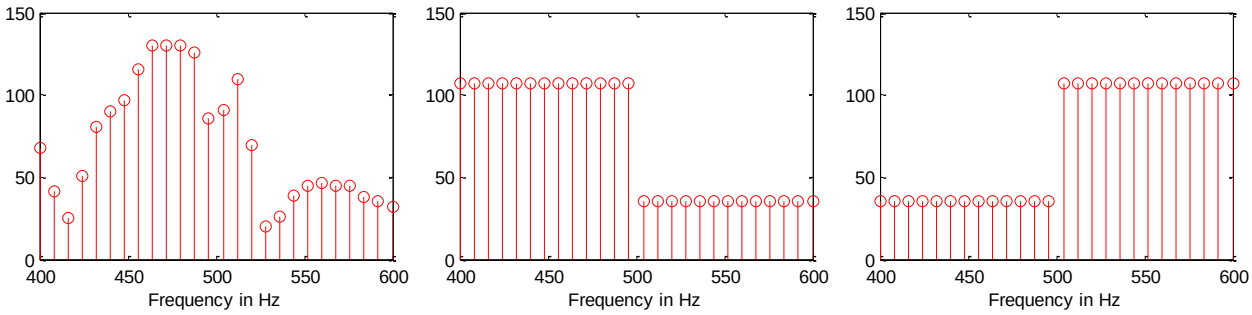
Figure 4. Chroma feature sequences of the original and two versions of play-recorded signals.

3. Watermark embedding and extraction

We use a frequency domain scheme to embed the watermark data that is usually in the form of a binary sequence. The embedding procedure is as follows.

1) Using the method described in Section 2, a series of M reference points in the time axis are located, which partition the audio signal into (M+1) parts, each again is divided into equal-length segments containing K samples.
2) Calculate FFT of each segment to obtain K complex spectral coefficients. Use the magnitude of the coefficients in the range of 400Hz~600Hz to carry one bit of the watermark data, see Fig. 5(a). Find the average of the spectral magnitudes, E.
3) To embed a "1", force the first half of the spectral lines to be 1.5E, and the second half to be 0.5E, see Fig. 5(b).
4) To embed a "0", force the first half of the spectral lines to be 0.5E, and the second half to be 1.5E, see Fig. 5(c).

This way, the embedding rate of the watermark is close to but slightly less than $n = fs/K$ where fs is the sampling rate of the audio, since the length of each part divided by K generally does not give an integer. Fig. 6 gives a comparison of the waveforms before and after watermark embedding, and Fig. 7 compares the spectrograms. It is seen that the distortion introduced by watermark embedding is quite small.



(a) Spectra in 400Hz~600Hz (b) Embed a "1" (c) Embed a "0"

Figure 5. Spectral lines before and after data embedding.

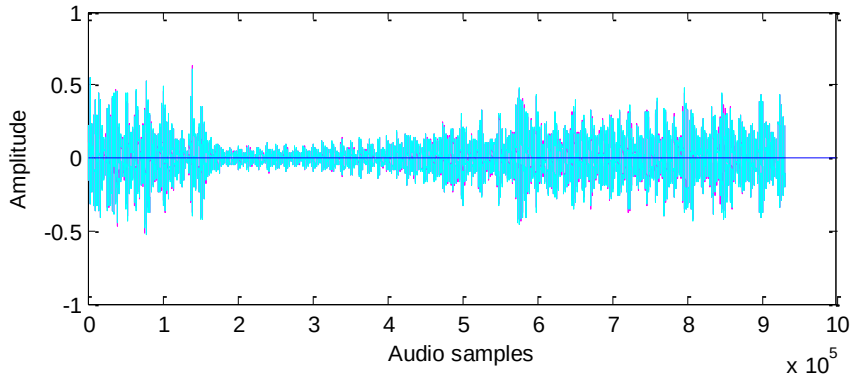
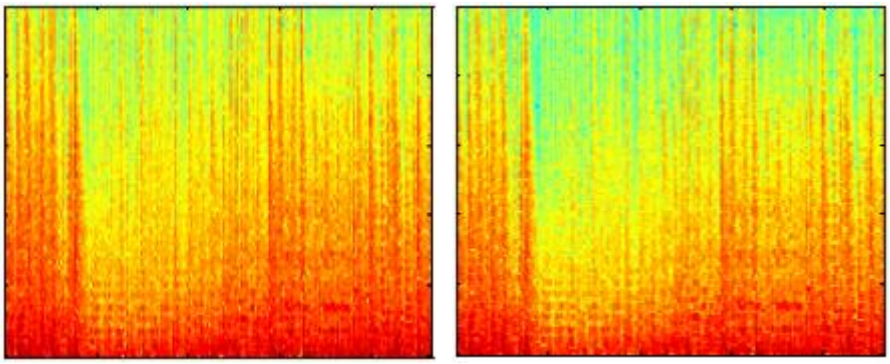


Figure 6. Watermarked waveform almost coincide with the original.



(a) Original (b) Watermarked

Figure 7. Spectrograms before and after embedding.

In the watermark extraction, the audio is segmented according to the synchronization reference instances using the same method as in the embedding, and spectral lines in 400Hz~600Hz are found. A decision as to whether the embedded bit is "1" or "0" can be made from the comparison between the two halves of the spectral lines. An example of extracted "1" is shown in Fig. 8.

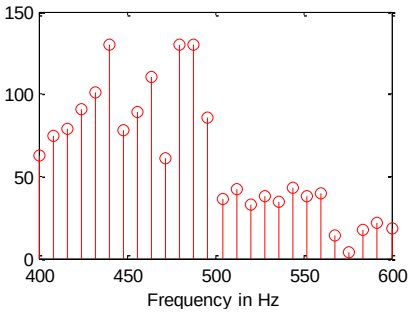


Figure 8. The extracted bit is "1".

4. Experiments and discussion

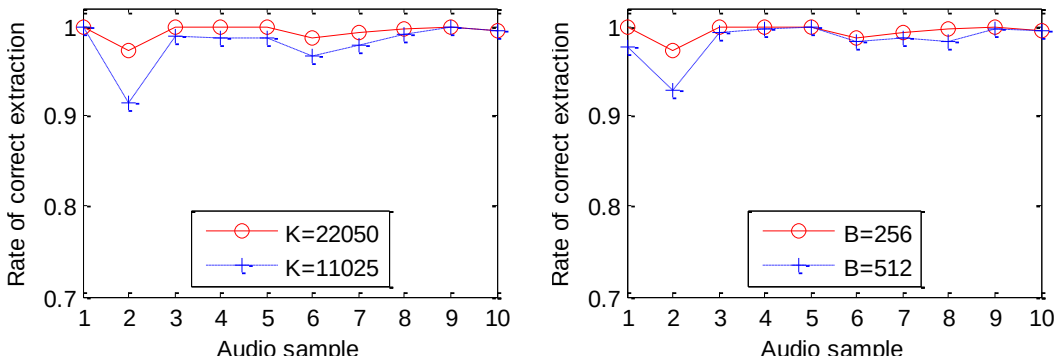
We take 10 audio pieces in the WAV format including pop songs and light music, all sampled at 44100 Hz. The lengths of the music range from 21.5 to 279.3 seconds. The parameters are chosen as B=256 to give satisfactory synchronization, and K=22050 resulting in an embedding rate of nearly 2 bps. Imperceptibility of the watermark is measured with ODG. Bit-error-rates (BER) are calculated for attacks including MP3 compression, contamination of additive white Gaussian noise (AWGN) with SNR being 30 dB and 40 dB respectively, and the play-record processes. Table 1 gives the results.

Table 1. Imperceptibility and robustness performance.

No.	Length (sec)	Payload (bits)	ODG	BER (%)			
				MP3 128kbps	AWGN		Play-record
					30dB	40dB	
1	21.5	43	–0.203	0	0	0	0
2	56.8	113	–0.370	0.75	0	0	2.68
3	73.6	146	–0.321	0	0	0	0
4	137.7	276	–0.390	0	0	0	0
5	109.0	218	–0.459	0	0	0	0
6	115.3	230	–0.093	0.70	0	0	1.33
7	279.3	558	–0.454	1.80	0.43	0	0.72
8	187.7	375	–0.332	0.50	0.54	0.36	0.30
9	165.3	330	–0.419	0.72	0.31	0	0
10	90.7	181	–0.271	0.22	0.55	0.35	0.54

It is observed from Table 1 that the watermark is almost transparent in all tested audio clips with ODG close to 0. The worst one is –0.459. As a comparison, the results given in [5] are mostly around –0.7. Robustness is also good, especially to the play-record attack, with the worst case of BER=2.68%.

By changing the parameters B and K, the behavior of resisting play-record attacks will change. A smaller K increases the embedding capacity but reduces robustness, as shown in Fig. 9(a). Reducing B makes location of the synchronization references more accurate, shown in Fig. 9(b), but too small a B value, say less than 128, will cause the algorithm unstable.



(a) Different K with B=256 (b) Different B with K=22050

Figure 9. Performance with different parameters.

5. Conclusions

We have proposed an audio watermarking scheme that can survive the play-record process. Synchronization is achieved by using the semantic of the audio, that is, the chroma features. This solves the problem of heavy distortion in the waveform and spectrum due to play-record, giving reliable reference points in the audio signal for watermark embedding and extraction. The audio signal is segmented according to the chroma-based references, and some selected spectral lines are modified to carry the watermark data. Imperceptibility and robustness are tested in experiments. The proposed technique is applicable to streaming media and allows watermark extraction without interrupting the playing of the program.

This work was supported by Natural Science Foundation of China (61071187, 61103181) and Key Laboratory Foundation for Underwater Test and Control Technology (9140c260201110c26).

REFERENCES

1 Czyżyk, P., et al., Analysis of impact of lossy audio compression on the robustness of watermark embedded in the DWT domain for non-blind copyright protection, *Communications in Computer and Information Science*, **287**, 36-46, (2012).

2 Lang, A., Dittmann, J., Spring, R., Vielhauer, C., Audio watermark attacks: from single to profile attacks, *Proceedings of ACM Multimedia and Security Workshop*, (2005).

3 Daniel, P. W. E., Graham, E. P., Identifying cover songs' with chroma features and dynamic programming beat tracking, *Acoustics, Speech and Signal Processing*, **4**, 1429-1432, (2007).

4 Ellis, D., (2007). *Chroma Feature Analysis and Synthesis*. [Online.] available: http://www.ee.columbia.edu/~dpwe/resources/matlab/chroma-ansyn/

5 Chen, B., Zhao, J., An adaptive and audio watermarking algorithm for MP3 compressed audio signal, *International Instrumentation and Measurement Technology Conference*, 1057-1060, (2008).