

多媒體訊號處理 (II) 子計畫一： 多媒體數位浮水印技術 (II)

Multimedia Digital Watermarking Techniques (II)

計畫編號：NSC-89-2213-E-002-160

執行期限：89 年 8 月 1 日至 90 年 7 月 31 日

主持人：貝蘇章

台灣大學電機系教授

中文摘要

本研究針對數位聲音訊號提出一展頻浮水印方法。浮水印在加入到主音樂 (host audio) 之前, 先與一數列作相關運算 (correlation), 以打亂浮水印。把打亂的浮水印直接疊加到主音樂, 就形成了藏有浮水印的音樂。

要解出浮水印, 需要原始的主音樂。把藏有浮水印的音樂減去原始主音樂, 應得到被打亂的浮水印, 再把已打亂的浮水印與上述數列作一次相關運算, 則可以還原浮水印。

用來作相關運算的數列, 具有自相關函數為 delta function 的特性, 因此可以提高浮水印的效率與安全性。

此方法可以抵抗 MPEG-1 Layer III (MP3) 聲音壓縮攻擊。另外, 本研究中也採用了一種定量的方法來量測音樂之間的相似度, 以改進現行 MOS 方法的客觀性, 進而客觀評判浮水印技術的好壞。

關鍵詞：浮水印、相關運算、MP3。

ABSTRACT

A robust audio watermarking technique using spread spectrum approach is proposed in this paper. The watermark is scrambled by a spread spectrum sequence before it is inserted into the host audio. In the extraction process, the watermark must be restored by the same sequence. The spread spectrum approach can increase both the efficiency and security of the watermarking method. Two kinds of sequences, perfect sequences and uniformly redundant arrays (URA's), are tested. This technique proves to be robust against MP3 attack in most cases except when speech clips are used together with perfect sequences.

Keywords : Watermarking, Spread spectrum, MP3.

1. INTRODUCTION

It has been very common to distribute exact copies of data electronically nowadays. This makes data keeping far more convenient, but other problems may occur. It is possible that illegal copies may be made. To deal with pirating and to protect the intellectual property, the author can place some information (usually called watermark or digital signature) into his/her audio productions without being perceived. Only those with the right key can successfully extract the watermark.

Various techniques have been proposed to cope with pirating problems in audio data. As Tilki et al. mentioned in [1], watermark can be inserted by replacing the Fourier transform coefficients over the middle frequency bands by spectral components of the watermark. In [2], watermark is embedded by modifying phase values of Fourier transform coefficients. Another technique is echo hiding, which employs multiple decaying echoes to place a peak in the cepstrum at a known location [2]. Also, watermark embedding making use of perceptual masking has been proposed by Swanson et al. [3]

In the proposed technique, the watermark is scrambled before it is added into the host audio. This is done by correlating the watermark with a perfect sequence or a uniformly redundant array (URA), whose autocorrelation is a delta function, so that the scrambled watermark can be totally restored by correlating with the sequence again [4,5].

2. SEQUENCES WITH AUTOCORRELATION OF ZERO SIDELOBE

A sequence with autocorrelation of zero sidelobe is equivalent to the direct sequence in spread spectrum communications. After the signal is correlated with the sequence, it becomes a random signal, just like a white noise. If this noise is correlated with the sequence again, the original signal is restored.

Reconstruction using sequences with autocorrelation functions of low sidelobe is mentioned in [4], in which URA (Uniformly Redundant Arrays) was introduced. Perfect sequences were further developed in [5].

2.1 Properties of Perfect Sequences

Assume a perfect sequence $s(n)$ with length N , and its periodic sequence $s_p(n)$ with period N . Some properties of perfect sequences are shown as follows: [5]

2.1.1 Correlation properties

The autocorrelation function, or the PACF (Periodic Repeated Autocorrelation Function) of $s_p(n)$ is

$$\zeta(m) = \sum_{n=0}^{N-1} s_p(n)s_p(n+m) \quad (1)$$

Then

$$\zeta(m) = \begin{cases} E, & m = 0 \\ 0, & m \neq 0 \end{cases} \quad (2)$$

where the energy E of the sequence is given by

$$E = \sum_{n=0}^{N-1} s^2(n) \quad (3)$$

The magnitude of the spectrum of a perfect sequence is always the constant \sqrt{E} .

2.1.2 Product theorem of perfect sequences

Consider two periodic perfect sequences $s_1(n)$ and $s_2(n)$ whose periods are N_1 and N_2 , with N_1 and N_2 relatively prime, and energy efficiencies η_1 and η_2 . Then their product is also a perfect sequence with period $N_1 \cdot N_2$. Also, the energy efficiency of the product sequence is the product of the energy efficiencies of the two original sequences, i.e.

$$\mathcal{Y} = \mathcal{Y}_1 \bullet \mathcal{Y}_2 \quad (4)$$

This can be proven by the definitions of perfect sequences [5].

2.1.3 Synthesis of perfect sequences

From 2.1.2, each perfect sequence $s_p(n)$ possesses a DFT $S_p(k)$ of constant discrete magnitude. This property is used in perfect sequence synthesis. Combining a constant amplitude spectrum any odd-symmetrical phase spectrum

$$\mathcal{E}(N-k) = -\mathcal{E}(k), \quad \text{in the region } 0 \leq k < N \quad (5)$$

can always give a real, perfect sequence by inverse DFT.

2.2 Properties of Uniformly Redundant Arrays

Uniformly redundant arrays (URA's), as introduced in [4], are binary matrices with autocorrelation of zero sidelobe. They are firstly developed to enhance the performance of coded aperture array image processing. A complete URA set consists of a pair of matrices, A and G, where A is the key used in the embedding process, and G in the extraction process.

2.2.1 Synthesis of URA's

Given a URA of dimension r by s , it must be satisfied that r and s are both prime numbers and $r \cdot s = 2$. The elements in the matrix are denoted as $A(i,j)$, where $i = 0 \sim r-1$, and $j = 0 \sim s-1$. URA's are generated as follows: [4]

$$\begin{aligned} A(i,j) &= 0 \quad \text{if } i=0 \\ &= 1 \quad \text{if } j=0, i \neq 0 \\ &= 1 \quad \text{if } C_r(i)C_s(j)=1 \\ &= 0 \quad \text{otherwise} \end{aligned} \quad (6)$$

where $C_r(i)=1$ if there exists an integer $x, 1 \leq x < r$ such that $i = x^2 \pmod{r}$
 $= -1$ otherwise

The extraction key G is generated by assigning

$$\begin{aligned} G(i,j) &= 1 \quad \text{if } A(i,j)=1 \\ &= -1 \quad \text{if } A(i,j)=0 \end{aligned} \quad (10)$$

2.2.2 Correlation Property

The circular correlation function of A and G is a 2-D delta function with the element in the intersection of the first column and the first row proportional to the number of 1's in A, which is the value $(rs+1)/2$, and the rest all zeros.

3. PROPOSED TECHNIQUE

3.1 Watermark Embedding

The proposed technique makes use of spread spectrum approach and repeated insertion. The watermark W is first correlated with a sequence P , resulting in a

noise-like signal I . This signal is scaled by a factor k and added into the host audio A , producing the stego audio S .

$$S = A + kI = A + k(W \otimes P) \quad (7)$$

where \otimes stands for correlation.

3.2 Watermark Extraction

In the extraction process, it is necessary to refer to the original host audio A . The received stego audio S is subtracted by A , obtaining the noise-like signal I , which is then correlated with the sequence P to restore the watermark W .

$$\begin{aligned} W' &= (S - A) \otimes P \\ &= kI \otimes P \\ &= k(W \otimes P) \otimes P \\ &= kW \otimes (P \otimes P) \\ &= kW \otimes u \\ &= kW \end{aligned} \quad (8)$$

The block diagrams of watermark embedding and extraction are shown in Figs. 1 and 2.

3.3 Repeated Insertion

When I is added into the host audio A , repeated insertion is adopted. That is, each sample in I is added into several samples in A . This concept is illustrated in Fig. 3 [6].

4. AUDIO SIMILARITY MEASURE

This technique proves to be robust against MPEG I - Layer III (MP3) compression attack. To grade the watermark quality objectively and quantitatively, the approach in [7,8] is incorporated.

In the quality evaluation process, Measuring Normalizing Block (MNB) technique is developed. There are several steps in the measurement process:

- (1) Two signals are normalized by removing the mean values and normalized to a common RMS level.
- (2) Each signal is broken into overlapping frames. Each frame is multiplied by Hamming window and transformed into frequency domain. Only the samples of DC to Nyquist are retained.
- (3) Select frames with energy above a given threshold. Transform the frequency domain samples into dB scale by taking logarithm.
- (4) Apply the Frequency MNB (FMNB).
- (5) Apply either Time MNB (TMNB) structure 1 or structure 2.
- (6) Apply linear combination and logistic function to obtain Acoustic Distance (AD) and Logistic Function of AD (L(AD)).

The range of AD is from 0 to infinity, and the range of L(AD) is from 1 to 0. For AD closer to 0 and L(AD) closer to 1, the two audio signals are of higher similarity perceptually. For two identical signals, AD is 0 and L(AD) is 0.9909 [7,8].

5. EXPERIMENTAL RESULTS

5.1 Data Profile

Two types of audio clips are tested in this research, including both music and speech. The data profiles are listed in Tables 1 and 2. Two clips are used as host audio

while four are used as watermark audio clips. All the clips are 16-bit PCM stereo WAV format except that the host ones are of sample rate 44100 Hz and the watermark ones are of sample rate $44100/6 = 7350$ Hz. Both perfect sequences and URA's are tested, and the experimental results are shown.

5.2 Using Perfect Sequences

The perfect sequence is used to scramble the watermark before watermark embedding. The extracted watermark after the stego audio has gone through MP3 compression and decompression is compared with the original watermark and the similarity results corresponding to different repeating block sizes are measured, as shown in Figs. 4 and 5. This value stands for the robustness of the watermarking technique against MP3 attack. The MP3 encoding specification tested here is the standard bitrate of 128kb/s.

5.3 Using URA's

The similarities of extracted and original watermarks using URA's under MP3 attack with different repeating block sizes are also plotted in Figs. 4 and 5.

6. DISCUSSION

If the extracted watermark qualities under two combinations are carefully investigated, it is clear that URA outperforms the perfect sequences no matter what types of audio clips are under consideration.

In both cases, the host and watermark combinations can be classified into two categories: "music in music" and "speech in speech", with the former combination more robust against MP3 attack. Also, the repeated insertion does not work much in audio watermarking. Given the same host audio, watermark audio, and scrambling sequence, larger repeating block sizes do not necessarily yield better results in similarity measurement. Therefore, the major factor that counts in robustness improvement is not the repeating block size, but the efficiency of the scrambling sequence, in the URA case, the number of 1's in the matrix.

7. CONCLUSION

An audio watermarking technique based on the spread spectrum approach is proposed in this paper. Sequences with autocorrelation function of zero sidelobe are introduced, investigated, and tested in the experiments. Also, their results under MP3 compression attack are presented in a new objective and quantitative audio

similarity measure.

The experimental results show that URA provides better robustness for watermark embedding than the perfect sequence.

Repeated insertion is adopted but proved not very promising in robustness improvement. The major part that improves the robustness is the sequence correlation.

The main contribution of this research is that audio clips are used as watermark other than the commonly used binary sequences. The employment of audio clips as watermark introduces much more challenges than binary signals, but has also pointed out another way on digital watermarking. Watermarks can be larger and more meaningful signals other than binary sequences, carrying more information about the author, owner, or the creation. It will be of wide applications in the future multimedia and internet oriented environments.

8. REFERENCE

- [1] J.F. Tilki, A. A. Beex, "Encoding a Hidden Digital Signature Onto an Audio Signal Using Psychoacoustic Masking", Proc. 1996 7th Int. Conf. On Sig. Proc. Apps. And Tech., Boston, MA 1996.
- [2] W. Bender, D. Gruhl, N. Morimoto, "Techniques for Data Hiding", Tech. Rep., MIT Media Lab, 1994
- [3] M. D. Swanson, B. Zhu, A. H. Tewfik, L. Boney, "Robust Audio Watermarking Using Perceptual Masking", Signal Processing 66, 1998
- [4] E. E. Fenimore, T. M. Cannon, "Coded Aperture Imaging With Uniform Redundant Arrays", Applied Optics Vol. 17 No.3, 1 Feb 1978
- [5] H. D. Luke, "Sequences and Arrays With Perfect Periodic Correlation", IEEE Trans. Aerospace and Electronic Systems, May 1998
- [6] C. H. Lee, Y. K. Lee, "An Adaptive Digital Image Watermarking Technique For Copyright Protection", IEEE Trans. Consumer Electronics, Vol. 45, No. 4, Nov 1999
- [7] S. Voran, "Objective Estimation of Perceived Speech Quality – Part I: Development of the Measuring Normalizing Block Technique", IEEE Trans. Speech and Audio Processing, Vol. 7, No. 4, Jul 1999
- [8] S. Voran, "Objective Estimation of Perceived Speech Quality – Part II: Evaluation of the Measuring Normalizing Block Technique", IEEE Trans. Speech and Audio Processing, Vol. 7, No. 4, Jul 1999

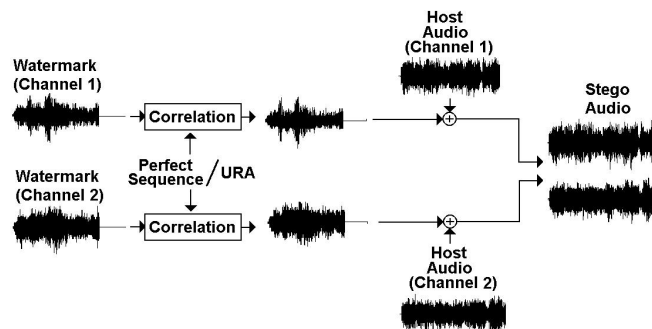


Fig. 1 Illustration of watermark embedding

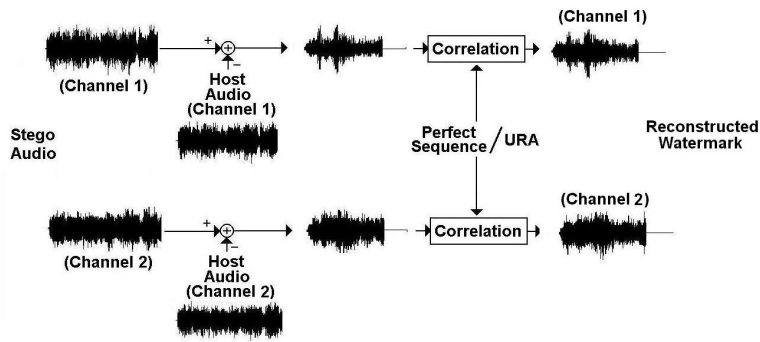


Fig. 2 Illustration of watermark extraction

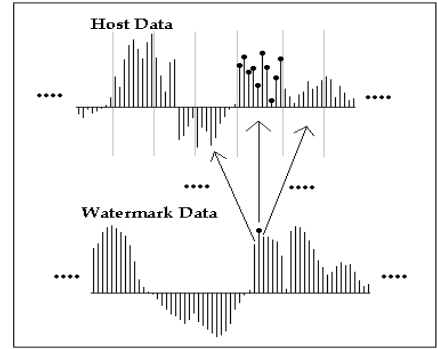


Fig. 3 Illustration of repeated insertion

Table 1: Audio clips used as host audio (sample rate = 44100Hz)

1	PIANO	Piano Solo: "Etude Op.25 No.12, Chopin"	13.531 seconds
2	SPFLE	English Speech: "Time Magazine", Female	17.580 seconds

Table 2: Audio clips used as watermark (sample rate = 7350Hz)

1	VOILN	Violin Solo: "Hungarian Dances No.1, Brahms"	4.232 seconds
2	GUITR	Guitar Solo: "Petenera Para Guitarra"	2.900 seconds
3	SPFSE	English Speech: "from Time Magazine", Female	5.538 seconds
4	SPMSE	English Speech: "from Time Magazine", Male	6.531 seconds

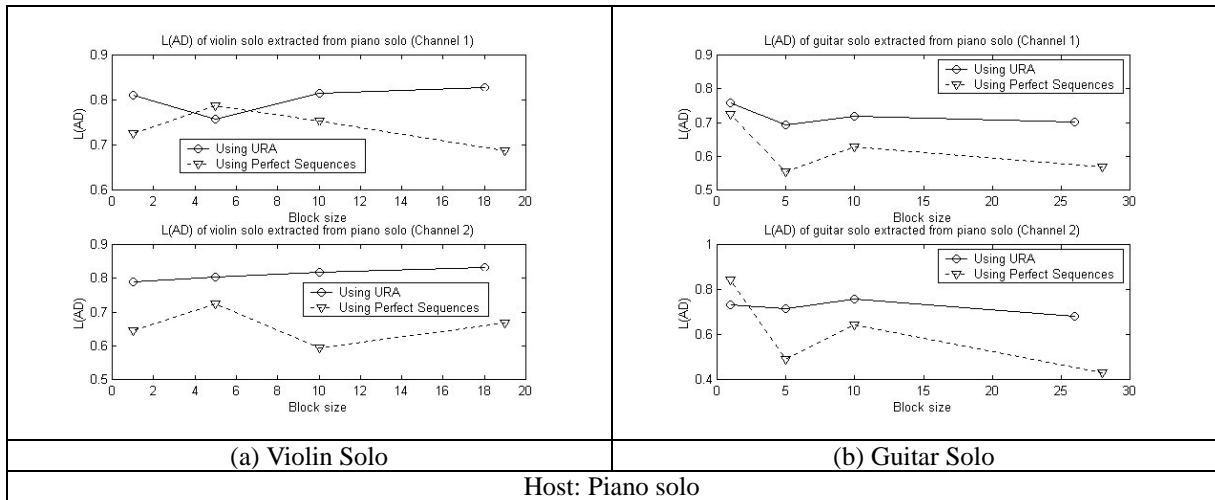


Fig. 4 Similarity values of music clips extracted from piano solo

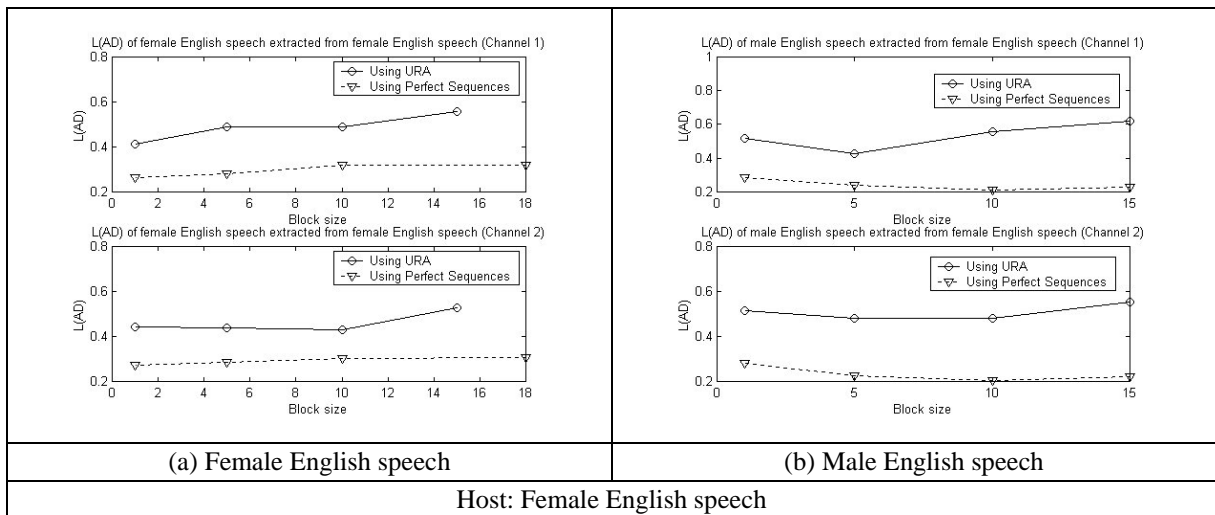


Fig. 5 Similarity values of English speech clips extracted from English speech clips