Blind Audio Watermark Decoding Using Independent Component Analysis

Sait Sener\textsuperscript{1,2} \hspace{1cm} Bilge Gunesl\textsuperscript{1}

\textsuperscript{1}Multimedia Signal Processing and Pattern Recognition Lab. \\
Electrical-Electronics Faculty. Dept. of Electronics and Communications Eng. \\
\textsuperscript{2}Institute of Informatics, Satellite Communication and Remote Sensing \\
Istanbul Technical University, 34469 Maslak, Istanbul, Turkey \\
http://www.ehb.itu.edu.tr/~bgunesl/mspr

Abstract

This paper proposes an audio watermark extraction technique which adopts Independent Component Analysis (ICA) for blind watermark decoding. Unlike the existing work, our method allows to combine data synchronization and watermark decoding into one optimization procedure, thus robust to transmission over different channels. Watermark encoder is designed as a nonlinear data embedding machine which is compatible to MPEG Layer1 Model 1. It is shown that the proposed ICA based watermark decoding scheme allows decoding watermark info accurately even though the watermark to signal ratio is less than -20 dB. The method is robust to stereo-to-mono conversions and performs very well when the channel noise level is high.

1. Introduction

Recent growth in the distribution of digital multimedia data over networks and internet caused authentication and copyright problems. Digital watermarking is proposed as a solution to these problems. Digital watermarking is a technique to embed narrow-band data called watermark robustly and imperceptibly into a host signal such as digital image, video, audio, fingerprint and etc, mainly to protect IP rights [1,2]. This paper deals with the audio watermarking. Many techniques have been proposed for audio watermarking in time domain as well as frequency domain [1, 3]. Using the frequency domain watermarking potentially improves the robustness against attacks. To build a robust audio watermarking system is also requires the design of a robust decoding mechanism. Most of the watermarks decoding methods use a kind of correlation-based decision rule to extract the watermark bits. This is mainly because of the simplicity of correlator receivers [3, 4]. The decoding performance of correlation based watermark extractors rely on a number of parameters, such as Watermark-to-Signal Ratio (WSR), signal to noise ratio (SNR), and the length of watermark bit streams. On the other hand, synchronization delays coming from transmission through the communication channels may reduce the watermark extraction capability of the decoder significantly.

This paper proposes a new audio watermark extraction framework which adopts ICA for blind watermark decoding. ICA is a commonly used signal processing method developed for Blind Source Separation (BSS) [5]. The application of ICA to digital watermarking for the purpose of multimedia authenticity protection was first introduced in [6]. Then a variety of researches have been used ICA for detection and for image watermarking [7]. A novel, flexible audio watermarking approach by ICA presented in [8]. Unlike the reported works, our method employ ICA not only an extraction tool but also a synchronization tool between the transmitter and the receiver. While preserving inaudibility by psychoacoustic masking model, we also keep high decoding rate by ICA decomposition. It is shown that the proposed watermark decoding method is robust to stereo-to-mono conversions and independent channel noise. Also it does not require the original data for decoding.

2. Watermark Embedding

An adaptive audio watermarking technique introduced in [4] is used for watermark embedding. It allows maximizing watermark to signal ratio in an iterative way and controls the decoding accuracy at watermark encoding stage, while preserving inaudibility. The encoding scheme is compatible to MPEG Layer 1 Model 1 and does not require the original audio data for decoding, and robust to stereo-to-mono conversions as well as to additive channel noise, even though the watermark strength is low. The audio data is processed frame by frame. At each instant, the encoder takes an original audio frame, \( s^0_i \), as its input and transmits the

0-7695-2128-2/04 $20.00 (C) 2004 IEEE
corresponding watermarked frame, $s_{hw}^t$, over the communication channel. Watermark embedding is performed by

$$s_{hw}^t = s_i^0 + \sum_{c=1}^{t-1} f(s_i^c, w_j)k,$$

$$i=1,..,(Lx RP), j=1,...,L \quad (1)$$

where Refresh Period (RP) refers to the repeat number of block embedding. In Eq.(1), $w_j$ is the watermark bit which is transmitted in frame $i$. Watermark bits can be either 1 or -1, $j$ can be any integer from 1 to $L$ where $L$ is the watermark length. $k$ is a secret key sequence with zero mean generated by a Pseudo Noise generator (PN). $f(.)$ is a nonlinear function of $s_i^c$, the audio frame at step $c$, and the watermark bit $w_j$. $f(.)$ models the watermark generation thus the second term in Eq.(1) represents nonlinear distortions coming from watermarking.

### 3. Watermark Detection and Extraction

Our watermarking scheme can be described as the process of embedding a key sequence, specified by the original audio and the watermark bit stream, into an original audio stream. Therefore, we consider the secret key and the original audio data as two independent data sources. ICA is used to decompose the embedded key and the original signal by using received observations. Then, by looking at the correlation between the decomposed key and the secret key, watermark data is decoded as 1 or -1. In order to guarantee the extraction of watermark signal by ICA, it is required that the number of observed inputs is at least equal to or larger than the number of independent audio sources. Thus we need to acquire at least two watermarked noisy data sequence at the receiver site.

The proposed watermark decoder receives a number of watermarked noisy data as its input and first applies a whitening operation on them. Then ICA block performs decomposition on the whitened audio resulting in estimated source signals. This is followed by a correlation based traditional decoding scheme.

Let $S_{hw}^t$ refers to the received watermarked signal at $t^{th}$ observation, and $t = 1,2,...,m$ where $m \geq 2$. Note that here usage of the index “$t$” is different from the previous section. In order to guarantee the independence of observations, our decoder first applies a PCA based whitening process on $S_{hw}^t$ and provides uncorrelated, unit variance observations, $\tilde{S}_{hw}^t, t=1,2$. $\tilde{S}_{hw}^t$ can be computed by using Eq.(2)

$$\tilde{S}_{hw}^t = ED^{-1/2}E^TS_{hw}^t \quad (2)$$

where $E$ is a $m \times m$ (here $m=2$) orthogonal matrix of eigenvectors of covariance matrix $E[S_{hw}^1S_{hw}^T]$ and $D$ is a $m \times m$ diagonal matrix of its eigenvalues.

While the PCA whitening process estimates the uncorrelated components, it is also quite useful to reduce the dimension of the data. In our application, reducing the dimension allows detecting whether the received signal is watermarked or not. If both of or one of received audio streams are unwatermarked, the PCA whitening automatically reduces the dimension to one.

Let observed watermarked audio signals are arranged into a vector $S_R = [S_{hw}^1, S_{hw}^2, ..., S_{hw}^m]$ and the independent source signals into $S_s = [S_1, S_2, ..., S_m^m]$, respectively; then the ICA relationship is given by

$$S_R = AS_S \quad (3)$$

where $A$ is an unknown mixing matrix of full rank. The basic problem of ICA based watermark decoding is then to estimate the original signal vector $S_s$ from observed mixtures $S_R$ or, equivalently, to estimate the inverse of mixing matrix such that $W = A^{-1}$. In order to estimate $W$, we used FASTICA [10], one of the commonly used algorithmic solutions to the ICA estimation. After finding $W$, it is used to separate the mixtures to obtain $\hat{k}_i$, estimated secret key embedded in audio frame $i$. Note that $\hat{k}_i$ is not a noise free signal and the additive channel noise affects the performance of ICA. Eq.(4) formulates ICA decomposition.

$$S_S = WS_R \quad (4)$$

The ICA decomposition provides us the source signal vector, $S_s = [S_1, S_2]$, and according to our model, one of the source signals corresponds to the transmitted original audio signal while the second one is the estimated embedded key. Since our decoding scheme does not require the original signal, the only signal that we have at the receiver site is the original secret key, $k$. Thus, detection of the transmitted watermark bit is accomplished by looking at the correlation between $k$ and decomposed source signals. Eq.(5) defines the correlation

$$r(\tau) = \sum_{l=0}^{M-1} k(l)S_s(l+\tau) \quad (5)$$
where \( t = 1, 2, \ldots, 512 \), and \( \tau = 0, 1, \ldots, (L x R \ P) \).

Ideally, the correlation for original audio should be equal to zero, since the key and original audio are uncorrelated. On the other hand, it becomes maximum for the estimated embedded key and the inserted watermark bits can be extracted according to the sign of the correlation as follows:

\[
\hat{w}_j = \begin{cases} 
1 & \text{sign}(r) > 0 \\
-1 & \text{sign}(r) < 0
\end{cases}
\]

(6)

The reason that we have variable \( \tau \) in Eq.(5) is synchronization problem between the transmitted and received audio bits. In practice, the watermarked data is transmitted through a variety of communication channels resulting delays in the received bit streams. Thus, the received audio bit stream is delayed at time \( \tau \). Therefore, our decoding scheme first specifies the exact \( \tau \) value through the audio stream then switches to the decoding, thus integrates the bit synchronization to the watermark extraction. Figure 1(a) and (b) show two audio frames asynchronous watermarked with 306 bits delay. Figure 1(c) shows correlation values obtained by our system versus \( \tau \). As it is shown in Fig. 1(c), decoder finds the maximum correlation values at 306th, 818th (306+512) bits and so on. Thus the proposed ICA-based synchronization tool enables to install data synchronization.

\[\text{Table 1. Correlation values between the original secret key and the extracted key.}\]

<table>
<thead>
<tr>
<th></th>
<th>Fr.1</th>
<th>Fr.2</th>
<th>Fr.3</th>
<th>Fr.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>With ICA</td>
<td>-0.8679</td>
<td>0.8608</td>
<td>0.8511</td>
<td>-0.8744</td>
</tr>
<tr>
<td>Without ICA</td>
<td>-0.0066</td>
<td>-0.0503</td>
<td>0.0542</td>
<td>0.0302</td>
</tr>
<tr>
<td>Extracted WM Bit</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The watermarked audio files were attacked, respectively; by adding Gaussian noise. Table 2 presents the number of correctly decoded (C), false (F) and missed (M) watermark blocks with or without ICA, at different SNRs. The watermark embedding has been performed at two different frequency bands, 2-22050 Hz and 25-1300 Hz. As it is seen from the Table, ICA improves the WM extraction performance significantly. Results obtained for two different frequency bands are slightly different, thus the proposed scheme is robust to band pass filtering.

Finally, we conducted experiments to see how a “stereo-to-mono conversion attack” would affect the proposed scheme. First, we extracted Left (L) and Right(R) channels of the same audio file. Both channels are watermarked at two different watermark strength, simultaneously. Then the watermarked stereo audio is generated by using a professional editing software. Robustness to stereo-to-mono conversions is analyzed on domain. The secret key and the original audio data extracted by ICA decomposition are shown in Figure 1(e) and (f), respectively. Observe that the embedded key and the estimated key obtained by ICA are almost the same. It is also true for the original audio frame and the estimated one. Table 1 presents the correlation values between the original secret key and the extracted key. Note that ICA increases the correlation significantly therefore reduces the WM decoding errors.
the mono recorded stereo file. Table 3 presents the WM decoding results obtained for Left (L), Right (R) and mono recorded stereo (S-to-M) files. Alternatively, mono audio is watermarked with the same parameters and decoding results are reported as Mono in the Table 3. It is observed that the proposed decoding scheme is robust to conversions and ICA decomposition improves the decoding performance significantly. Note that the false alarm rate is equal to zero for noise free audio. Robustness to band pass filtering is also preserved under the stereo-to-mono conversions.

5. Conclusions

In this paper, we present a new approach based on ICA to audio watermark extraction. Our embedding scheme analyses the stability of the audio samples by using a Model 1 Layer 1 compatible Psychoacoustic Auditory Model, at encoding stage. ICA is used to eliminate the de-synchronization problem encountered in recording watermarked audio transmitted from different sources. The experiments show that our approach is robust to channel noise, stereo-to-mono conversions as well as de-synchronization attacks and preserves inaudibility.

6. References


Table 2. Performance comparison of the traditional decoding and the proposed ICA-based decoding at different SNRs and robustness to filtering attack.

Table 3. Robustness to stereo-to-mono conversions.