An Analysis-by-Synthesis Echo Watermarking Method

Wen-Chih Wu\(^{a,b}\), Oscal T.-C. Chen\(^c\)

Dept. of Electrical Engineering, National Chung Cheng University, Chia-Yi, 621 Taiwan\(^{a}\)
Dept. of Electrical Engineering, Wu-Feng Institute of Technology, Chia-Yi, 621 Taiwan\(^{d}\)

oscals@mail.bcu.edu.tw, wcwu@samlab.ee.ccu.edu.tw

Abstract

In the conventional echo watermarking methods, the robustness of watermarks could be improved by increasing the amplitudes of the echo signals, or as known as the decay rates, but at the expense of sacrificing audio quality. The objective of this work is to achieve a better audio quality and preserve the robustness of watermarks. Along with the proposed analysis-by-synthesis approach, there are two other considerations used to adjust the appropriate decay rates of each audio signal segment. One is the characteristic of audio signals after watermarking, and the other being the accurate watermarks that are decoded and recovered after MP3 attacks on the watermarked audio signals. Looking at the average decay rates and the signal-to-noise ratios, the proposed method exhibits better audio quality than those of the conventional method. Furthermore, the experimental results show that the recovery accuracy rates of watermarks from the watermarked audio signals after attacks are much higher than those of the conventional method.

1. Introduction

With the ever-increasing exploitation on intelligence property rights [1]-[6], and the growing concern for protecting such rights, digital watermarks provide an effective solution. However, as humans are more sensitive to hearing rather than seeing, designing a high-performance audio watermarking proves to be a challenge task. Audio watermarking embeds data in the original audio signals [1]. To achieve this, the system has a few prerequisites. First, the embedded data must not degrade the original audio quality. Secondly, audio quality must preserve its robustness after the run-through of some standard signal processing, so that the original watermarking information could be accurately recovered. Presently, major audio watermarking methods are functioned by using low bit coding, phase coding, spread spectrum coding, echo hiding, etc. [2].

The low-bit coding method embeds data in the least significant bit of the original signals. While this method could embed a large amount of data, it is unable to tolerate attacks from signal processing. The phase coding method embeds data in the phase values of Fourier Transform coefficients in audio segments. Phases of neighboring segments must adjust accordingly to maintain correlation. To prevent phase distortion from interfering audio quality, changes in phases must be smooth and the volume of embedded data need be restricted. The spread spectrum watermarking method attempts to disperse the data in different frequency spectrums [3], so to prevent signal interference on some frequency bands to affect watermark decoding. However this method would result in poor audio quality as noises are added on the original signals.

The echo hiding method embeds data by inserting echoes [4,5]. Echo positions are marked as data identification tags. The embedded data would be difficult to detect because these data have the same statistical and perceptual characteristics as the original signals. The robustness of the echo hiding method can be improved by increasing the amplitudes of the echo signals (or known as the decay rates), although such would degrade the audio quality. To decide what is the appropriate decay rate for the original audio signals is a difficult measure. To improve upon it, the psycho-acoustic model has been proposed to modify the decay rate [6].

In the previous paper, the authors of this work had already proposed using the analysis-by-synthesis approach to decide on the decay rate value [7]. With the amplitudes of echo signals adjusted only under the condition of no attacks, the robustness of watermarks did not improved much at all. Of all the attacks, it is especially important to watch out for MP3 encoding and decoding attacks. MP3 attack is a common and popular procedure used in audio signal processing. A watermark method that cannot effectively withstand MP3 attacks is considered impractical. In this work, the decay rate value would be adjusted by in response to the recovery of watermarks after signals have undergone MP3 attacks. With the analysis-by-synthesis approach, the impacts of MP3 attacks are first considered before adjusting the appropriate decay rate value. Such step would allow the embedded watermarks to have high robustness against those MP3 encoding and decoding attacks encountered at regular basis. As our simulation data reveals, this approach not only improves audio quality, it also demonstrates a much better performance in robustness against attacks, when compared to conventional echo watermarking methods [2]. Hence the proposed echo watermarking method is applicable in any industrial applications [8].
2. Proposed echo watermarking method

The proposed echo hiding method is realized with the analysis-by-synthesis approach. Before adding the watermarks, considerations are given to see if a watermark can accurately recover signals, especially after MP3 signal processing. Echo amplitudes are adjusted appropriately according to the characteristics of each audio signal segment. This was done so that the embedded echoes have minimal impacts on the original audio signals and it also ensures that data be accurately recovered after various common signal attacks. An encoder and a decoder are designed based on the proposed echo watermarking method.

2.1. Encoder design

During the encoding process, watermark data would be embedded into the original audio signals. The encoder shown in Fig. 1 is divided into three stages, including segmentation, echo watermark by using the analysis-by-synthesis approach and segment recombination.

**Segmentation:** First, the original signals are partitioned into many segments. Each segment is embedded with one-bit watermark datum.

**Echo watermarking by using the analysis-by-synthesis approach:** This stage is broken down to details in Fig. 2. With the analysis-by-synthesis approach, each watermark datum is embedded in each segment. Steps of such procedure are as described below:

Step 1. Generating kernels: If the embedding watermark datum is the binary “ONE”, the positive echo with decay rate of $G_1$ is placed at the time delay of $\delta_1$, while the negative echo with a decay rate of $G_0$ is placed at the time delay of $\delta_0$. The decay rate represents the magnitude ratio of the echo signal over the original signal. Convolution of the original signal and kernel “ONE” function would generate these two echo signals. The function of kernel “ONE” is shown in Fig. 3(a). Contrarily, if the embedding watermark datum is the binary “ZERO”, then the negative echo with a decay rate $G_1$ and the positive echo with a decay rate $G_0$ are placed at the time delays of $\delta_1$ and $\delta_0$ respectively. Convolution of the original signal and kernel “ZERO” function would generate these two echo signals. The function of kernel “ZERO” is as shown in Fig. 3(b), with the initial decay rates of $G_1$ and $G_0$ set at zero.

Step 2. Echo encoding: The echo watermarking method embeds data by adding echoes to the original signals. With the different types of watermark data available, original signals would convolute with the appropriate kernel functions accordingly. Watermarked audio signals can be obtained via the echo encoding process, as seen in Fig. 4.

Step 3. Watermark checking: This step is performed similar to the decoder’s. Watermark data embedded earlier by the encoder are extracted here to match with the original watermark data. Such action is to ensure that the decoder would accurately recover the original watermark data at the end. This is also the essence of using the analysis-by-synthesis approach in the proposed watermarking method. In this step, the conditions of two different types of audio signals are looked at – audio signals with embedded watermarks itself and after MP3 attacks. Now, we define the autocorrelation value of the signal’s cepstrum as follows [2, 9]:

$$Mag(x) = |\text{complex}(F(x))|^2.$$ (1)

First, the autocorrelation values of the watermarked audio signals at the two time delays, $\delta_1$ and $\delta_0$, are computed. Next, the difference between the autocorrelation values of these two time delays is compared to the threshold value defined below.

$$\text{Threshold} = \frac{|Mag_{G_1}(x) - Mag_{G_0}(x)|}{\text{Mag}_{G_1}(x) + \text{Mag}_{G_0}(x)} < 0.25,$$ (2)

where $|\cdot|$ represents the absolute value of a number and $Mag_{G_0}(\cdot)$ indicates the autocorrelation value of the signal’s cepstrum at the position “A.”

**Fig. 1 The block diagram of the encoder.**

**Fig. 2 The block diagram of the echo watermarking using the analysis-by-synthesis approach.**

**Fig. 3 The impulse responses of the echo functions: (a) “ZERO” kernel. (b) “ONE” kernel.**

**Fig. 4 Echo encoding process.**

If the autocorrelation difference is less than the threshold value, then the recovered watermark data is not confirmed and decay rates would be corrected. If the difference is...
greater than the threshold value, then the watermark data could be determined pending on the autocorrelation values of the two time delay positions. If the autocorrelation value at position $\delta_1$ is greater than the autocorrelation value at position $\delta_0$, watermark data is set as bit “ONE”. If not, the watermark data would set to bit “ZERO”. Once the system retrieves the recovered watermark data, this value would be compared to the original embedded watermark data again. Any inconsistency between the two would result in modifying the values of the decay rates. 

Step 4. Modifying decay rates: If the threshold requirement or the recovered watermark datum differs from the expected value, the decay rate is modified. If the original watermark datum is set as bit “ONE”, then the decay rate $\gamma_1$ at a position $\delta_1$ would increase while the decay rate $\gamma_0$ at a position $\delta_0$ would decrease. On the other hand, if the original watermark datum is bit “ZERO,” then the decay rate $\gamma_1$ at a position $\delta_1$ would decrease while the decay rate $\gamma_0$ at a position $\delta_0$ would increase. After the decay rates are adjusted, go back to Step 2 to derive a new watermarked audio signal segment until all signals are processed.

**Segment recombination:** Upon embedding the appropriate watermark data in all audio segments, these embedded signals are recombined to form a complete set of watermarked audio signals.

**2.2. Decoder design**

Figure 5 displays the block diagram of the decoder. First, watermarked audio signals are sectioned into many segments. Then autocorrelation values of the cepstrum for each segment are computed at two different time delays. If the value of position $\delta_1$ is greater than that of position $\delta_0$, the extracted watermark datum is considered as a “ONE”. Otherwise, the datum would be assigned with “ZERO”.

**Fig. 5 The block diagram of the proposed decoder.**

**3. Experiments and results**

The proposed method uses four audio pieces as the host signals. The total length of each signal has 4096000 samples. Each sample is denoted with 16 bits, with the sampling rate of 44.1 kHz. The time delay positions of “ZERO” and “ONE” are located at the 9th and 26th samplings, approximately 0.2ms and 0.6ms, respectively. Original audio signals are partitioned into segments, with each segment having 1024 samples, around 23ms. Performance comparisons of the proposed and conventional method developed by Bender et al. are done based on the following conditions:

1. Immediately decoding after encoding (closed-loop).
2. Decoding after the MPEG encoding/decoding: Encoding/decoding was performed using the MPEG-1 audio layer III.
3. Decoding after quantification: After quantification, each sample of the watermarked audio signal is quantified to 9 bits from the original 16 bits.
4. Decoding after adding random noises: After adding random noises, watermarked audio signals would have a SNR ratio of 20dB between noises and original audio signals.
5. Re-sampling: After downsampling by a factor of 2, watermarked audio signals recover by the way of linear interpolation.

Hence the Recovery Accuracy (RA) is as defined below:

$$\text{RA} = \frac{\text{number of bits correctly decoded}}{\text{number of bits encoded}} \times 100\% \quad (3)$$

After encoding the test audio signals by ways of the proposed and the conventional echo watermarking methods, and running them through different forms of attacks, the recovery accuracies are listed in Fig. 6. While Fig. 7 shows a comparison between decay rates and segment sampling numbers. However with the proposed method stating that each audio segment is to have two different decay rates at two time delays, it is only fair to set the following decay rate definition for each audio segment:

$$\text{decay rate} = |p_0| + |p_1|. \quad (4)$$

The chart in Fig. 8 compares the Signal-to-Noise Ratio (SNR) of watermarked and original audio signals, as demonstrated under the proposed and conventional methods. In most circumstances, SNR can be used to evaluate audio signal quality. The definition of such quality measure is set out below:

$$\text{SNR} = 10 \log_{10} \left( \frac{\sum_{n=-N}^{N} (x(n) - \hat{x}(n))^2}{\sum_{n=-N}^{N} \hat{x}(n)^2} \right) \quad (5)$$

Where $x(n)$ is the original audio signal and $\hat{x}(n)$ as the watermarked audio signals.

In echo hiding, a decision must be made between inaudibility and robustness. The trade-off of using a large decay rate to achieve higher recovery accuracy is to sacrifice the audio quality. However, as the chart reveals in Fig. 7, the proposed method has lower average decay rates than those of the conventional method. Such is possible since the analysys-by-synthesis approach is adopted in the proposed method to avoid over-degrading the decay rates in accordance to the characteristics of each audio signal segment. This would help to improve audio quality, as the signal-to-noise ratios would be shown in Fig. 8. With the proposed method, watermarked signals would have a SNR of 4dB higher than that of the conventional method.

With the proposed method, decay rates are adjusted to ensure that the autocorrelation values at two predefined locations are consistent with the embedded watermarking data. However, sometimes a decay rate is increased far too much to preserve audio quality. This case occurs when the autocorrelation value of the cepstrum at an echo position is much smaller than that at the other position. To avoid this case, negative echoes are added as well as positive echoes. In the adaptation process, when the positive decay rate is increased at one position, the negative echo is also increased at the other position. Adding a negative echo would lower
the autocorrelation value of the cepstrum at the particular location and thus could control the positive and negative decay rates to stay within a reasonable range. As for audio quality, the original audio with negative echoes may descend the amplitudes of its low-frequency bands to yield sharper and high-pitched sounds [10]. In addition, many decay rates have values of zero, which implies that these audio segments have the same quality as the original audio signals. This validates the goal of this work – to use the analysis-by-synthesis approach to embed watermark data without adding any other signal.

Furthermore, comparing to the conventional method, the proposed method not only has a better audio quality with its regards to the average decay rate and SNR performances, but it is also more robust. Data from Fig. 6 indicate the high recovery accuracy rate of the proposed method under various attacks, which is much better than that of the conventional method. This exceptional performance is attributed to the analysis-by-synthesis approach adapted in the proposed method, in which the decay rates are adjusted accordingly – appropriate echo signals are given in accordance to the different echo characteristics to keep the robustness of all the signals on a par. Theoretically, the recovery accuracy under the closed loop and MP3 scenarios should be 100%. But in actuality, it may have been that the characteristics of the audio signals are prohibiting any embedding of the watermarking data, the recovery accuracy could not achieve its potential.

![Fig. 6 Comparisons of the recovery accuracies of the proposed and conventional echo watermarking methods.](image)

![Fig. 7 Comparisons of the decay rates.](image)

![Fig. 8 Comparisons of the signal to noise ratios.](image)

4. Conclusion

In this work, an echo watermarking method with the analysis-by-synthesis approach is proposed to preserve audio quality and to increase recovery accuracy. In the process of embedding watermarks, the decay rate of each signal segment is adapted by using the positive and negative echoes to minimize the echo decay rate required. Such would also minimize the impacts that embedded echo signals have on the original signals. As for the widely used MP3 compression and its impact on the watermarking signals, the proposed method will take it into account and is able to accurately extract the watermarks after MP3 compression. Simulation results reflect the fact that the proposed method is of a better method than the conventional method, as verified with the values of the embedded echo amplitude and SNR. Last but not least, the proposed method is much more robust than the conventional method against various attacks. With all these evidences, it is fair to say that the proposed echo watermarking method can be widely used in various industrial applications.

5. References


