行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　※
※　　以回聲資訊隱藏為基礎的音訊浮水印技術設計(I)　　※
※　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：■個別型計畫　　□整合型計畫

計畫編號：NSC　89-2218-E-009-014

執行期間：　91　年　7　月　1　日至　92　年　7　月　31　日

計畫主持人：劉 啟 民　教 授

共同主持人：

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：交 通 大 學　資 訊 工 程 系

中　華　民　國　93　年　6　月　30　日

以回聲資訊隱藏為基礎的音訊浮水印技術設計(I)

計畫編號：NSC91-2213-E-009-093

執行期限：民 國 91 年 8 月 1 日 起 至 民 國 92 年 7 月 31 日 止

主持人：劉啟民　交通大學資訊工程系

主要參與研究生：李文傑，張子文，許瀚文，楊宗瀚

一、中文摘要

上年度研究進度報告和本年度研究重點

本計畫第一年目標為系統設計，包含以下目標

- 理論分析。
- 系統參數確定以保證音訊品質不遭破壞。
- 系統的程式設計。
- 分析何種音樂特色可加入不影響音樂品質和資料抓取
- 可加入浮水印資料量分析。
- 不可加入浮水印音樂片段偵測。

現此計畫已完成以上目標並進行第二年目標。

本計畫第二年目標為系統改進和新方法，包含以下目標

- 靜音的偵測處理。
- 高浮水印偵測率方法。
- 提出提高浮水印資料量方法。
- 音樂品質評估方法。
- 浮水印資料對壓縮方法的強健性分析。
- 浮水印資料對音效處理方法的強健性分析。

Abstract

This report illustrates the backgrounds of the project in Chapter 1 and provides the research results in the first two years in Chapters 2 and 3. In Chapter 4, we give some experiments we have conducted.

Keywords: Watermark, Audio processing

Introduction

In the last decade, storage of most multimedia data, such as music and pictures, is no longer in analog format, but rather in digital format. For digital data, the improvements in compression techniques distinctly reduce the storage size of these multimedia files. The small size of digital audio files makes it easy to spread music to anywhere in the world through the Internet. Due to the rapid growth of Internet traffic throughout the world over the last few years, the distribution of digital audio has become even easier. For example, nowadays, many digital audio pieces, especially the popular songs, are spreading quickly and easily over the Internet. In addition, e-commerce has allowed trade to occur on a worldwide scale. The simple access and lossless duplication of digital files have already violated the intellectual property of a myriad of original authors and producers. This violation of copyrighted material makes selling digital multimedia products, like e-books or MP3 music, via the Internet a huge risk. The lack of protection from these attacks on intellectual property over the Internet obviously increases the importance of copyright protection [1] [2].

Traditionally, the copyright is declared by putting a special trademark, like a unique type of sticker, on the products or by giving a registration card with the products. However, the traditional methods may not work well in digital world due to the easy reproduction and simple modification characteristics of the digital products. For example, some of the software, which request the user for a registration number to install or to start the program, have been attacked by hackers in order to create an unauthorized, but fully-functional copy that does not require any registration. Thus, the unauthorized software was spreading over the Internet. Software engineers are now searching for methods to avoid destruction of registration information from these types of copyright attack. Currently, the best way to defend against intensive violation of intellectual property, especially to data in digital format, is to hide the authorization information within the software or multimedia files and to make the hidden information imperceptible, which is in the concept of watermark. Some digital audio watermarking techniques were designed to protect the intellectual property of music. The basic idea of the digital audio watermark is to embed some specified information like a personal signature or serial numbers into the digital products before selling them. The watermark can help to prove the ownership of the producers and buyers, to justify the authorization of the products and to avoid the violation of intellectual properties. Thus, selling products, such as music and movies, in digital format through the Internet will no longer threaten the intellectual property [3].

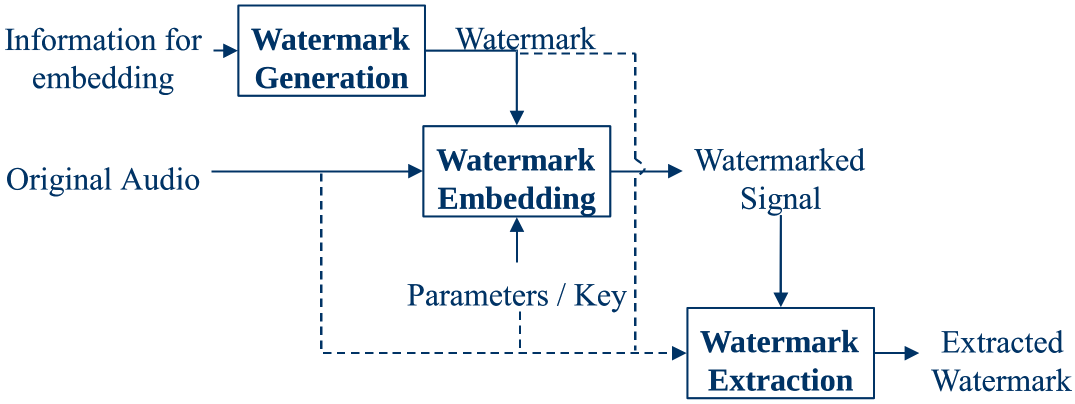


Fig. 1 Architecture of watermark system

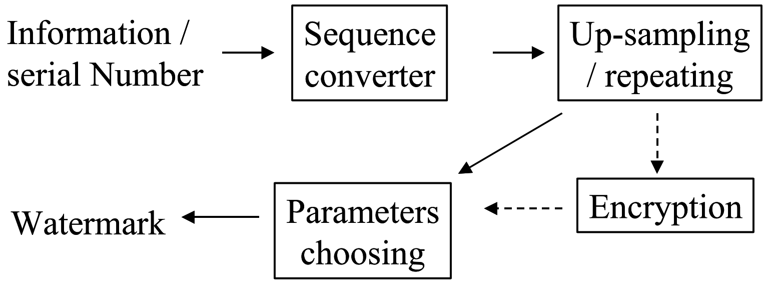


Fig. 2 Block diagram of watermark generation

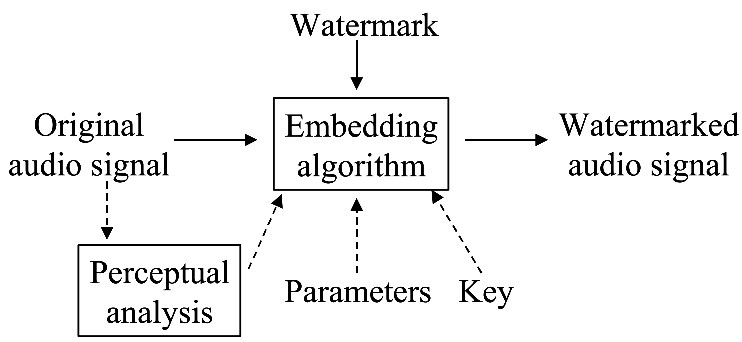


Fig. 3 Block diagram of watermark embedding

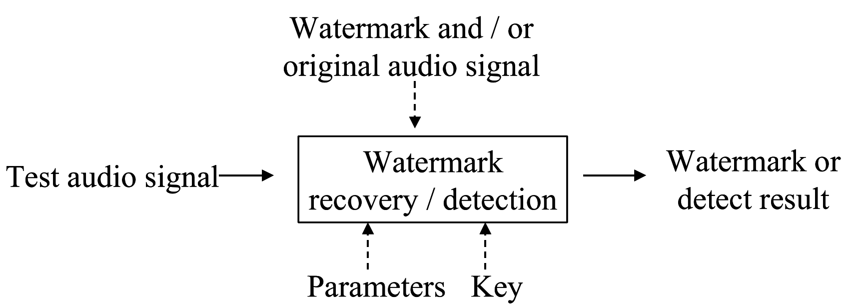


Fig. 4 Block diagram of watermark extraction

1.1 Basic Concepts of Watermark

The architecture of watermark system looks like Fig. 1. The input of the watermark system is the information for embedding and the original audio; the output of the watermarking system is the watermarked signal and the extracted watermark. The watermark data is generated separately before considering the host multimedia file, as illustrated in Fig. 2. The format of watermark data depends on the original format of the information as well as the embedding algorithm, which is adopted the host multimedia file. The watermark is embedded into the multimedia file using specified control parameters and the private key of the technique, which generates a watermarked file as illustrated in Fig. 3 [4]. The embedding algorithm must cope with the variation of the

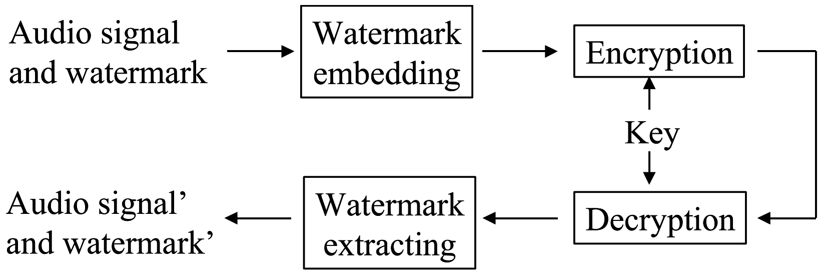


Fig. 5 Flow when watermark and cryptography work together

media data and must carefully ensure that the changed media can still be read, watched or listened without allowing other people to perceive any difference from an ordinary, non-watermarked file.

Authorities may check the embedded information by extracting the watermark within the watermarked file with some statistical techniques. As shown in Fig. 4, the extraction approach, which should approximate to the inverse of the embedding procedure, requires the same control parameters and the private key, which was used in the embedding procedure. After the hidden data has been extracted back from the watermarked audio, the authorized individuals can determine or recognize the owner of this media file [5].

1.1.1 Cryptography and Watermark

Watermark and cryptography appear similar, but the two terms are far from the same. Cryptography and watermark are both used on the protection of some important information, and are applied into the original data that requires protection. However, cryptography is used to protect content from being access by anyone without authorization. For example, an encrypted audio file cannot be listened if the file is not decrypted first. To decrypt the encrypted data, the private key is required to transfer the data back to its original format prior to encryption. Once the data is decrypted, the data is no longer under protection and can thus be accessed by any person [6]. In contrast to cryptography, a watermarked data can be directly access by any person; there is no need to 'un-watermark' the data in advance. In particular, people should not be able to recognize whether the data contains an embedded watermark. This covert method is completely distinct from the cryptography approaches. For instance, one can listen to a watermarked music file without any 'un-watermarking' step and this person would not realize the existence of watermark. While the data is all recovered to the original format after decryption in cryptography, the watermarks are designed to permanently reside within the host data in any condition. Thus, the watermark cannot be removed from the host data even after intentional hacker attacks or repeated reproduction unless one has the private key and the watermarking algorithm has been designed as watermark removable. Some watermark techniques make the mark impossible to remove even if the private key is known; in the other words, once the watermark is added, it is indestructible.

Thus, based on the characteristics of cryptography, encryption is typically used on protecting data from being intercepted during transmission, or preventing unauthorized access of the content. On the other hand, the watermark is generally used to protect copyright and to declare the ownership of the content [7]. The cryptography and watermark can be used together to gain better profit, and the working flow is illustrated in Fig. 5.

1.1.2 Requirements of Audio Watermark

To achieve the goal of copyright protection, an audio watermarking technique has to satisfy several requirements:

- ***Embedding into the audio signal***

The watermark must be embedded directly into the main part of audio signal, not into the header of audio file. Information in the file header may be easily removed or manipulated, but the data embedded in the audio signal would only be destroyed when the audio itself is destroyed [8].

- ***Transparency in perception***

The watermark should be imperceptible to human ears; otherwise, it would affect the audio quality of the watermarked signal and would consequently be detected [9].

- ***Information extractable***

The embedded watermarks are used for ownership verification, so the watermarks must be able to be extracted or to be detected. The algorithm of watermarking technique should not only be able to embed the information but also be able to recover it.

- ***Recovery without referring to the original***

Unless an audio signal can be proven as the original signal by a third party, there is no way to determine whether or not an audio piece is the original. Anyone may generate a fake original, which can extract a different but valid watermark with the watermarked audio piece. Due to the possibility of counterfeit originals, the watermark extracting techniques that depend on the original audio signal are not reliable [10].

- ***Robustness***

  Digital data are easily modified and manipulated using computers and widely available software packages. Several frequently used programs with lossy signal processing functions may be used to modify the watermarked digital audio. Furthermore, a third party may attempt to modify the watermarked signal in order to destroy the embedded data by attacking the file with multiple kinds of audio process operations. A watermarking procedure should be robust to withstand signal manipulation and processing operations on the host data; for example, up/down-sampling, compression, noise, A/D-D/A (A: analog, D: digital) conversions, etc [11].

- ***Security***

  The watermark embedding procedure must be secure. An unauthorized user must not be able to extract the watermark or even to detect the presence of the embedded data. A watermarking scheme is truly secure if knowledge of the exaction algorithm for embedding the data does not help an unauthorized party to detect the presence of the embedded data. This security standard is analogous to the criterion for measuring the security of an encryption method. For further security, the watermark may be encrypted before its insertion into a host signal [8].
- ***Support of multiple watermarks***

An embedding algorithm allowing multiple watermarks to co-exist is more preferable. In some circumstances, it is desirable to embed multiple watermarks into the same host signal, for example, one for the producer and one for the buyer [9].

1.2 Audio Watermarking Techniques

Hiding data in audio signals presents a variety of challenges, due in part to the wide dynamic and differential range of the human auditory system (HAS) as compared to the other senses. The HAS perceives over a range of powers greater than one billion to one and a range of frequencies greater than one thousand to one. Although the human auditory system is sensitive, there are some 'spaces' available in the perception range where data may be hidden. While the HAS has a large dynamic range, it often has a fairly small differential range. For instance, loud sounds tend to mask out the quiet ones nearby. Additionally, while the HAS is sensitive to amplitude and relative phase, it is unable to perceive absolute phase. Finally, there are some environmental distortions that commonly occur and are typically ignored by the listeners in most cases [12].

1.2.1 Main Techniques

Information hiding techniques receives more and more attention from the research community and from the industry in recent years, as Table 1 shows the main driving force is concern over copyright protection [7]. However, most of the works are focused on watermarking techniques of image and video, and only a few audio data hiding schemes have been proposed. This indicates that the amount of information that can be embedded robustly and inaudibly is much lower for audio media than for visual media. An additional problem in audio watermarking is that the HAS is much more sensitive than HVS (human visual system), and that inaudibility of audio is much more difficult to achieve than invisibility of images [13].

Table 1 Number of publications on digital watermarking in the past few years

Reference: [15]

Year	1992	1993	1994	1995	1996	1997	1998
Publications	2	2	4	13	29	64	103

Based on different types of 'spaces' for information hiding, several kinds of watermark approaches have been proposed in recent years:

- ***Low-bit coding***

The first type embeds watermark by replacing the least significant bit (LSB). In low-bit coding technique, watermark data is transferred into a binary stream and the stream is used to replace in series the LSB of each audio sample. This technique may introduce some noise and the embedded watermark is easy to be destroyed [14]. This technique may also be implemented over the frequency domain [16].

- ***Region-based coding***

The second type embeds watermark in some frequency regions of the audio signals. This approach embeds the watermarks into some perceptually insignificant regions, such as the high frequency region, so that the watermark is inaudible. The concentration of hidden information in a small region makes this approach less robust [9].

- ***Phase coding***

The third type embeds the watermark in the Fourier transform phase coefficient. Human ears are relatively insensitive to phase distortions, especially the absolute phase value. A phase coding scheme was proposed to update the phase of an initial audio segment and keep the relative phase of following segments unchanged [14].

- ***Embedding during compression***

Some methods combine watermark embedding with the compression or modulation processes, such as during the vector quantization step in compression procedure [17].

- ***Pseudo-noise coding***

Most techniques add the watermark as pseudo-random noise, some of which are generated based on the spread spectrum concept, in the time domain. Since human ears have different sensitivity to additive noise in different frequency bands, most of the proposed methods in this area use some type of filter to shape the added pseudo-random noise and achieve inaudibility [9] [19].

- ***Echo data hiding***

Still, some approaches embed the watermark as the echo signals of the original audio. The inaudibility of echo data hiding is based on temporal masking effect of the human auditory system. Based on HAS, if the echo is added within a certain range, the human ears cannot distinguish it from the original. The echo is perceived as resonance in our environment. The watermark signals are actually delayed and attenuated versions of the original audio signals in this kind of approaches [12].

1.2.2 Advantages of Echo Data Hiding

Each approach mentioned above has its own pros and cons. The low-bit coding technique is not robust; although this technique provides the highest information

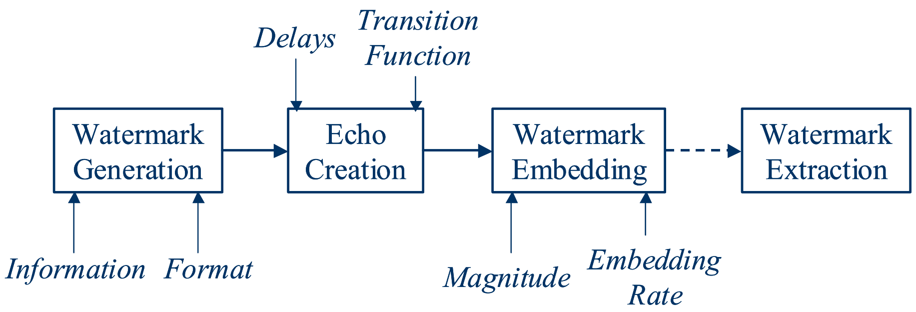


Fig. 6 Basic block diagram of echo data hiding system

transmission rate, the embedded watermark is very fragile. The phase coding approach provides much better robustness, but has a very low information transmission rate since the secret information is encoded only in the first signal segment [18]. The method that embeds the watermark during the compression procedure cannot survive when another kind of audio file format is used. Moreover, embedding the watermark within compression processes is also dependent on compression software produced by particular company. In the algorithm, which hides data in perceptually insignificant regions, the watermark may be easily destroyed when audio compression is performed, especially for low bit rate compression, where the least significant parts of the audio are the first to be removed [9].. One of the most common approaches for hiding data in audio is to introduce the information as additive noise. The main drawback for these noise-adding approaches is that lossy data compression algorithms tend to remove most imperceptible artifacts, including typical low dB noise, and the embedded data would be removed at the same time [12]. Echo data hiding introduces changes to the host audio, which has the characteristic of environmental conditions rather than random noise, thus it is more robust even through any general lossy data compression. However, the available rate to transmit information is a little low in echo data hiding, as compared to some of the other approaches[14] [18].

Based on the tradeoffs between robustness and information transmission rate of watermarks, the echo data hiding technique is the fittest method and will be further studied in this report.

1.3 Design Issues

There are four main procedures for echo data hiding technique: the watermark data generation, the echo creation, the watermark embedding and the watermark extraction, as illustrated in Fig. 6.

In the first stage, the information converting to the embedding watermark must be chosen, and the format of the watermark data, for example a binary stream, must be decided. The embedding information is then converted to the predefined format as the watermark file. The watermark format helps to decide the amount of information can be embedded in per block of the audio signal. For instance, if a binary stream is generated, one single bit is hidden in each block. Different watermark formats may employ different number of digits to encode the information data, which will be embedded into one single block.

In the echo creation stage, the delay values need to be chosen in a specified proper range. The echo signal is then created from a combination of several delay signals. The transition function used for combining the delay signals will influence the audio properties and the extraction accuracy.

During the watermark embedding stage, the embedding control parameters, such as the magnitude of the echo and the embedding rate of the watermark, are

Table 2 Characteristics of watermarking techniques

Reference: [14] [18]

Time / frequency domain embedding	Additive noise?	Robustness	Techniques
Time domain	Yes	Bad	Low-bit coding
	Yes	Middle	Pseudo-noise coding
	No	Good	Echo data hiding
Frequency domain	Yes	Bad	Region-based coding
	No	Good	Phase coding
Time / frequency			Embedding during compression

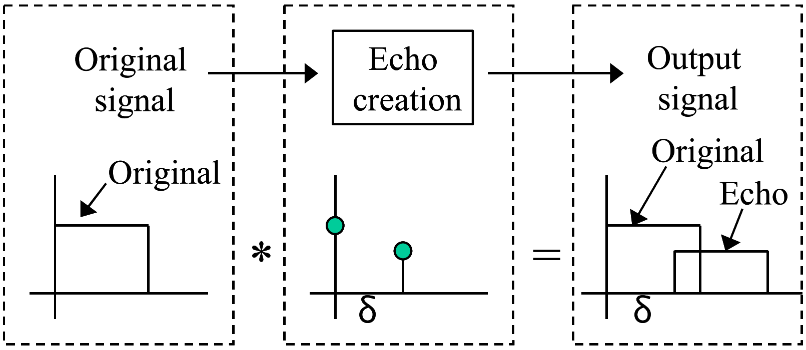


Fig. 7 Example of echo generation

assigned within a specific range. With these control parameters, the echo is embedded into the original audio signal, and then the audio signal becomes an intellectual property protected file. For each audio signal, some of the segments are unsuitable for hiding data since the watermark added in these segments could not be correctly extracted and would decrease the overall watermark recovery accuracy. Also, the block size for watermark embedding is limited by the accuracy rate. There is a tradeoff between the watermark embedding rate and the recovery accuracy rate.

To examine the embedded information within the multimedia file, the extraction stage is needed. The algorithm of watermark extraction in echo data hiding technique is based on the autocorrelation characteristics of the original audio signal and the watermark echo signals. The extraction requires the control parameters, like the delay values and the watermark-embedding rate, as the inputs to check the echo existence of the extraction algorithm. The detection result is later converted into the same format as the watermark information source. The converted value may be compared with the source of embedded data to check the ownership of the audio piece.

Chapter 2 System Design

For the protection of intellectual property, every audio watermarking technique embeds the information into the host audio signal based on some consideration of perceptual characteristics of the human auditory system. To extract the embedded data from the host audio pieces, some statistical approaches were used [20].

Based on the tradeoff between robustness and watermark-embedding rate, a watermarking system for audio signal based on echo data hiding techniques is chosen for study. The echo data hiding techniques employ the temporal masking characteristics of audio combined with imitation of a natural environmental echo to embed the watermark data into the host audio. A statistical method of calculating the autocorrelation of the watermarked signal is used to extract the embedded information.

This chapter begins by explaining several basic concepts used in echo data hiding approach. The following sections discuss each step of the echo data hiding procedure and the corresponding parameters in detail. Based on the different combinations of the control parameters in the procedure, some basic assumptions related to the parameters are discussed at the end of this chapter. Finally, the performances of the echo data hiding system with various control parameters are illustrated for comparison.

2.1 Concepts of Echo Data Hiding

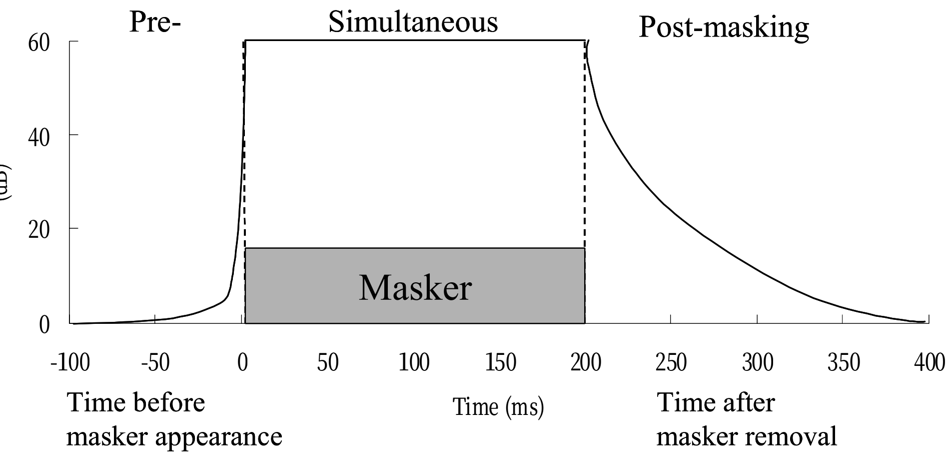
In the echo data hiding technique, we introduce an echo signal into the audio in order to hide the watermark information. The temporal masking property of audio and the simulated echoes are used in watermark embedding, and the statistical method of autocorrelation is used for hidden information extraction [12].

2.1.1 Echo

The definition of ‘echo’ is: “a sound heard again near its source after being reflected” or “a repetition of sound produced by the reflection of sound waves from a wall, mountain, or other obstructing surface” [21]. An echo is, in fact, a natural phenomenon of sound, which is a kind of repeat of the original. When people play music, echoes occur in the room when the sound wave is reflected from the walls or off other objects. People seldom notice the occurrence of echoes in our environment just because they are already accustomed to the existence of echoes. Now a simulation of this natural phenomenon is used for hiding the specific information within an audio signal. To simulate the effect of natural echoes in the original audio, the echo signals are created and embedded directly into the original signal. For example, the echo can be created with a delay time δ and an echo amplitude decay A, as shown in Fig. 7. Thus the echo signal is combined with the original and becomes single one output signal.

2.1.2 Temporal Masking

Watermarks in audio signals are embedded in the positions where human perception would not be able to detect, while the receiver of audio is without doubt the human ear. Through the characterization of human auditory perceptual system, particularly the time-frequency analysis of the capabilities of the inner ear, watermark finds its place to hide in the audio signal. Within the temporal masking effects of audio, the echo data hiding technique conceals the watermark from being heard.



Temporal masking properties of the human ear

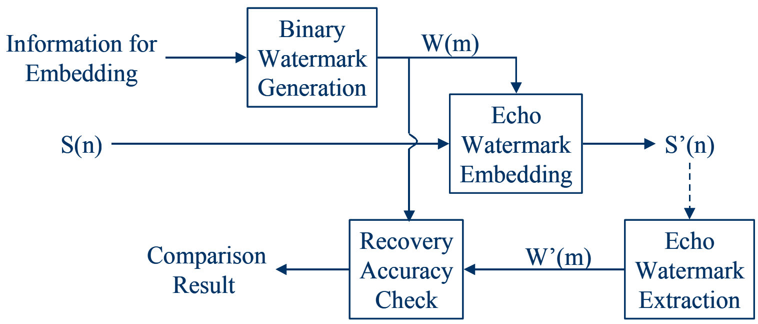


Fig. 9 Original echo data hiding system

The masking phenomena of auditory perception extend in time beyond the window of simultaneous stimulus presentation. In other words, for a masker of finite duration, temporal masking occurs both prior to the masker appearance as well as after the masker removal. Fig. 8 schematically presents the skirts on both sides of the masker. The absolute audibility thresholds for masked sound increase exponentially prior to the occurrence of the masking signal and decrease exponentially posterior to the removal of the masker. While significant pre-masking tends to last only about 1-2 milliseconds, post-masking would extend anywhere from 50 to 300 milliseconds, depending on the strength and duration of the masker. As shown in Fig. 8, the pre-masking decays much more rapidly than the post-masking, so the spotlight of echo data hiding is concentrated on the post-masking region [22] [23].

2.1.3 Autocorrelation

Within a series of audio signals, the echoes appear to be the repetitive signals of the audio that were just played. To determine whether a sequence of signals is in some degree self-repeating, the calculation of its autocorrelation is usually adopted. The autocorrelation of a given signal $x(n)$ is defined as

$$R_{xx}(n) = \sum_{m=-\infty}^{+\infty} x(n+m)x(m). \qquad (1)$$

Let $k = n+m$ and thus $m = k-n$. By replacing m in Eq. (1), the autocorrelation can be equally defined as

$$R_{xx}(n) = \sum_{k=-\infty}^{+\infty} x(k)x(k-n). \quad (2)$$

For each position k in the time domain, Eq. (2) can be used to calculate the corresponding autocorrelation value. When an echo exists at time k_1, the autocorrelation value for the signal that has a time offset of k_1 would be larger than the autocorrelation value for any other signals with an offset of k (where $k \neq k_1$). By determining the maximum autocorrelation value, the embedded echo can be detected. Thus, if any echo was hidden in the audio signals, the autocorrelation can help to detect the offset of the echo at any time. Obviously, if any specific information was hidden inside the audio with echoes, the autocorrelation is helpful for deriving the information and the embedded data in the audio can then be extracted [12].

2.2 Echo Data Hiding System Design

The watermarking system based on echo data hiding is explained in this section. The echo data hiding system can be divided into the following four stages: (1) the watermark data generation; (2) the watermark embedding; (3) the watermark extraction; (4) the check of recovery accuracy. A simple block diagram of the system is illustrated in Fig. 9.

2.2.1 Watermark Data Generation

As Fig. 10 shows, watermark data generation

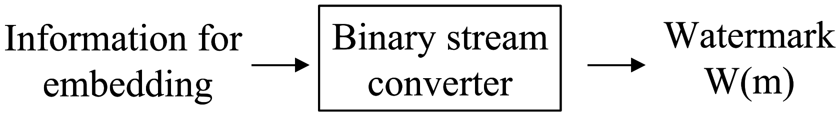


Fig. 10 Watermark data generation in echo data hiding system

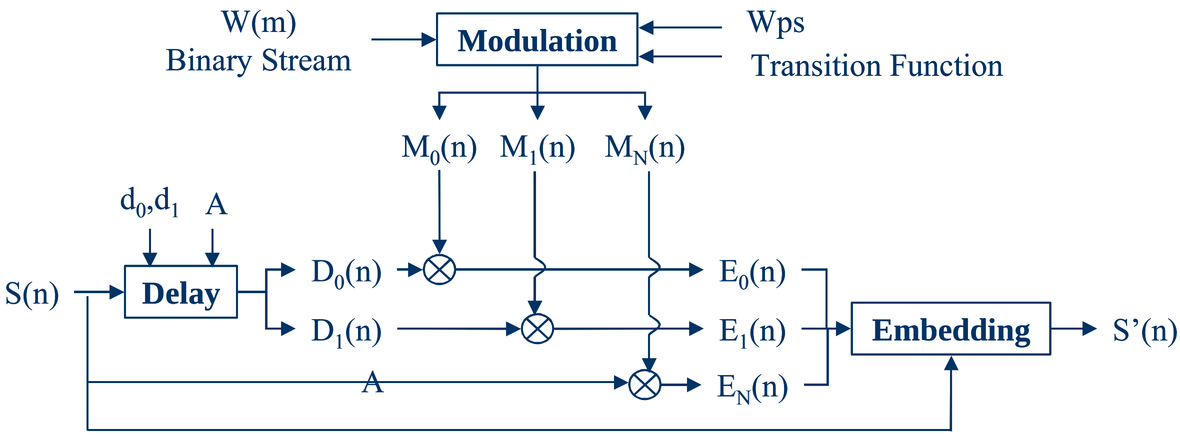


Fig. 12 Echo watermark embedding

involves picking up the data, which is going to be embedded into the host audio, and converting this data into the predefined format of the watermark, where the format is usually a binary stream. For example, a text string "Hello" is converted into the output binary stream as "0100100001100101011011000110110001101111"; or the data of a graph such as Fig. 11 is converted into the output binary stream as "01000111010010010100011000111000 ……".. Converting any data into a binary stream is quite simple, since all of the data is in fact presented as binary code inside the computer. If other watermark formats are adopted, the associated algorithm used in the 'watermark conversion' block should be changed. Moreover, the format of the watermark specifically depends on the watermark embedding method adopted. For example, if the embedding algorithm allows some empty segment without any data to be embedded, the converted binary sequence may become "0100 0111 010 01 000 ……", in which the blank means no watermark bit is hidden in that block. Thus, watermark data generation is pretty simple and corresponds with the binary representation within the computer.

2.2.2 Echo Watermark Embedding

The goal of echo watermark embedding is to create the watermarked audio, which may be distinguished from the original audio by a slight difference, but should not be identified. That is, the audio quality of the watermarked audio must be kept the same as the original audio. The embedding procedure of the echo watermark basically includes two modules: the echo creation stage and the watermark embedding stage as shown in Fig. 6. With the host signals, the echo creation procedure generates the repetitive signals of the original. The created echoes and the inputted watermark data go through the watermark embedding modules to produce the watermarked audio signals.

Fig. 11 Leaf

In echo creation, two important parameters, which indicate the delay values, should be assigned to generate some delayed signals for the creation. Since the watermark data is a binary stream, the two delay values d_0 and d_1, which represents the delays of digit '0' and digit '1' respectively in the binary stream, need to be assigned. In addition, a decay rate 'A', where $0<A<1$, needs to be specified for generating the echo signals as shown in Fig. 13. Assuming that $S(n)$ indicates the host audio signal, the two delay signals using the specified magnitude and decay rate can be represented by $D_0(n)$ and $D_1(n)$ as the following equations:

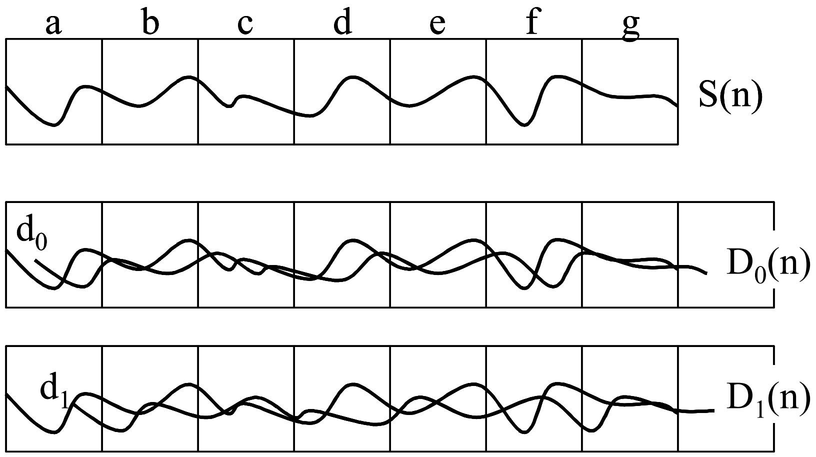


Fig. 13 S(n) and its relationship with D_0(n) and D_1(n)

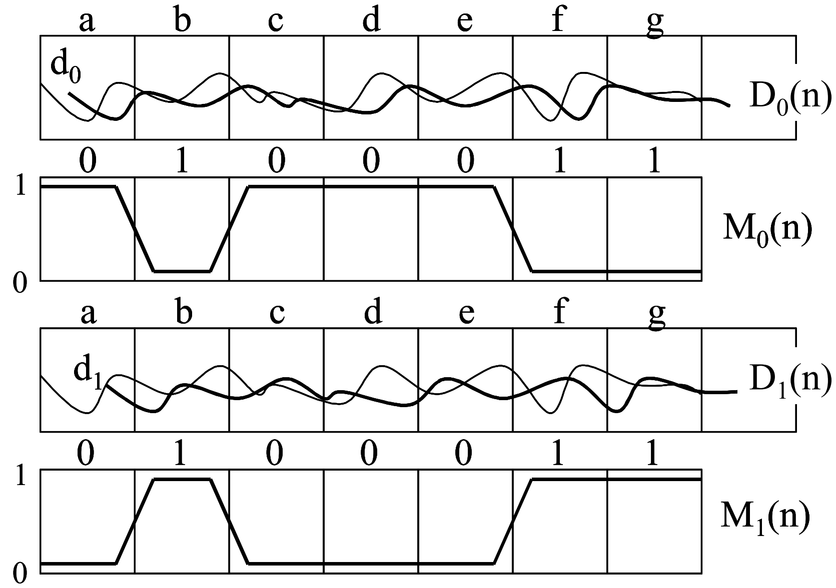


Fig. 14 Modulators and their relationship with delay signals

$$D_0(n) = A * S(n - d_0) \qquad (3)$$

and

$$D_1(n) = A * S(n - d_1)\,. \qquad (4)$$

On the other side, with the given watermark data stream, the watermark is embedded into the host audio with a predefined rate, watermark per second ('*Wps*'). With the rate of *Wps* assigned, two or three modulator streams are generated from the watermark stream in the block of modulation. The values in the modulators should be either one or zero to represent the existence of specific digits, while the values of the modulators in the transition region can be between one and zero. In the echo-watermark embedding module, the two main modulators, indicated as $M_0(n)$ and $M_1(n)$, are used to modulate $D_0(n)$ and $D_1(n)$. When a block without watermark data is allowed, another modulator $M_N(n)$ (where *N* means null, symbolizing 'non-digit') would be adopted. With the modulators, the embedding algorithm can determine whether the specific delay signal is added into the current block. For example, if the digit '1' is hidden in the current block, the corresponding segment of the modulator $M_1(n)$ would have value one, and the other modulators $M_0(n)$ and $M_N(n)$ would have value zero. As shown in Fig. 14, the content in each block of the modulator is highly dependent on the watermark data stream. The block size is selected based on the specified watermark embedding rate and the sampling rate of the original audio, which is given by

$$block\_size = \frac{original\_audio\_sampling\_rate}{Wps}\,. \qquad (5)$$

At the boundaries of each block, a transition function is introduced to avoid a sudden change in magnitude of the embedding signal. The transition function for a smooth change may be either a linear or nonlinear function. For example, the equation

$$M_0(n) + M_1(n) + M_N(n) = 1\,, \qquad (6)$$

makes the magnitudes change linearly, whereas another equation

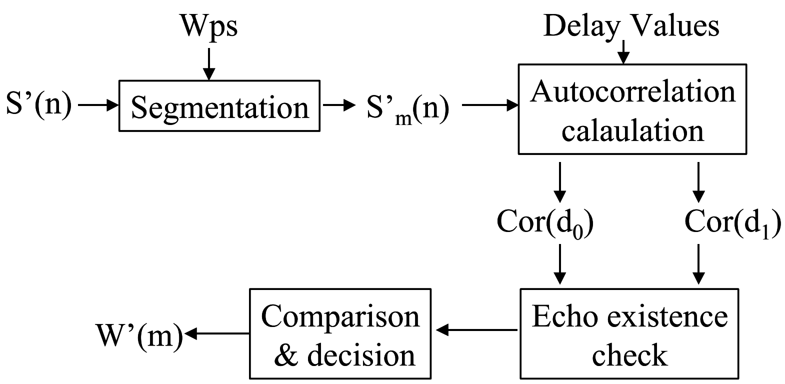


Fig. 15 Echo watermark extraction

$$M_0(n)^2 + M_1(n)^2 + M_N(n)^2 = 1 \quad (7)$$

causes the change in magnitudes of the signal to appear sinusoidal.

With the delay signals and the modulators generated, the embedding signals are generated in the following steps.

Firstly, in the 'modulation' block of Fig. 12, the delay signals are modulated using the modulators, and the embedding signals are generated by

$$E_0(n) = D_0(n) * M_0(n) \quad (8)$$

and

$$E_1(n) = D_1(n) * M_1(n). \quad (9)$$

To cope with the embedding signals within each block, in block which no watermark bit needs to be embedded, a decay of the original signal is modulated as illustrated in Fig. 12, that is,

$$E_N(n) = A * S(n) * M_N(n). \quad (10)$$

Basically, each embedded bit should have magnitude one in the correlating modulator at the time except for the block boundaries condition. Thus, one of the embedding signals, $E_0(n)$, $E_1(n)$ and $E_N(n)$, has non-zero amplitude and the remaining signals are in silence. This relationship is based on the design of modulators that only the specified modulator corresponding to the embedded bit should have a magnitude of one at any time.

Secondly, the embedding signals, $E_0(n)$, $E_1(n)$ and $E_N(n)$, together with the host signal $S(n)$ are the inputs of the 'embedding' block.

Finally, the embedding signals and the original audio are combined to generate the watermarked audio for each embedding block. In the embedding block, we first add $E_0(n)$, $E_1(n)$ and $E_N(n)$ together, which produces the combined embedding signal represented by $E(n)$ as in the following,

$$E(n) = E_0(n) + E_1(n) + E_N(n). \quad (11)$$

With the combined signal $E(n)$, a watermarked audio $S'(n)$ is then constructed simply by adding $S(n)$ and $E(n)$ together as

$$S'(n) = S(n) + E(n). \quad (12)$$

Obviously, the watermarked audio signals may have more energy than the original signals because the addition of echoes. To keep the energy of the watermarked audio signals equivalent with the original audio signals, the normalization of the watermarked audio is made to preserve the energy level of the original audio by the following equation,

$$S'(n) = \frac{S(n) + E(n)}{\sqrt{1 + A^2}}. \quad (13)$$

The normalized parameter in Eq. (13) equals to the summation of squared values of the signals energy [24], which is defined as

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2. \quad (14)$$

Where the value one is used to represent the energy factor of the original and the value A is the energy factor of the embedding audio signal. For a given audio signal, the embedding steps can provide the watermarked audio signal, which can be used or sold as a copyright protected product.

2.2.3 Watermark Extraction

Since the extraction of the watermark is done block by block, the information about the embedding rate (*Wps*) is required in order to segment the watermarked audio signal into blocks. To examine the existence of the specific watermark, the autocorrelation value of each block is calculated. The autocorrelation count is focused on the positions where the echo may be most probably positioned, at times d_0 and d_1. In this report, a normalized version of the autocorrelation matrix in Eq. (2) is used for watermark detection. That is,

$$R_{xx}[n] = \frac{\sum_{k=-\infty}^{\infty} x(k)x(k-n)}{\sum_{k=-\infty}^{\infty} x(k)x(k)}. \quad (15)$$

From Eq. (15), the autocorrelation count at time distance d_0 and d_1 are

$$Cor(d_0) = \frac{\sum S'(n)S'(n-d_0)}{\sum S'(n)S'(n)} \quad (16)$$

and

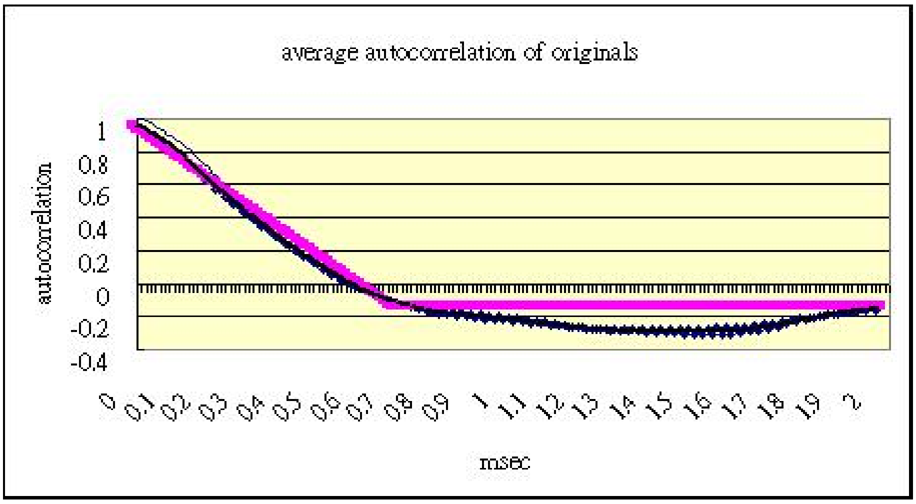


Fig. 16 Autocorrelation experimental result of the originals

$$Cor(d_1) = \frac{\sum S'(n)S'(n-d_1)}{\sum S'(n)S'(n)}, \qquad (17)$$

where n is the time index from the starting position to the end of the current block.

The resulting of the autocorrelation counts, $Cor(d_0)$ and $Cor(d_1)$, are then used for echo existence check prior to the extraction of the watermarked data. This check provides a threshold to determine whether or not the handling block has an echo embedded within it. When no echo exists in the current block, based on Eqs.(10), (11) and (12), the signal $S'(n)$ is

$$S'(n) = \frac{S(n)+E(n)}{\sqrt{1+A^2}} = \frac{S(n)+E_N(n)}{\sqrt{1+A^2}} = \frac{S(n)+AS(n)}{\sqrt{1+A^2}} = \frac{(1+A)S(n)}{\sqrt{1+A^2}}. \qquad (18)$$

Where $E(n)$ is directly replaced by the embedding signal $E_N(n)$, which is a function of $S(n)$, and the signals $E_0(n)$ and $E_1(n)$ are ignored. As shown in the Eq. (18), when no echo exists, the watermarked signal $S'(n)$ becomes a function of $S(n)$ that the characteristics of $S'(n)$ is almost the same as the original signal $S(n)$. That is, based on the direct relationship between $S'(n)$ and $S(n)$ in the absence of an echo, the autocorrelation characteristic of the original signals can be used to define the threshold of echo existence check. For example, several audio streams were tested for its autocorrelation characteristic before watermark embedding, and the average results are shown in Fig. 16. Observing in Fig. 16, the curved line of the autocorrelation value passes through zero point at approximates 0.6msec. Before the crossing at 0.6msec, the curve approximates a linear line that decreases from an initial value of one. Whereas after passing through the zero point, the curve looks like a near-sine wave. The threshold of echo existence check is usually used when the delay of echo is larger than 0.7msec (this is discussed with control parameters). For the echo existence check, the threshold may be simply assigned the maximum autocorrelation value of the original signal in the segment, which at time larger than 0.7msec. For further precision of the autocorrelation value threshold, a sine function can be assigned for the threshold definition, such as

$$\text{Threshold} = B * \sin(\alpha\, t + \theta) + \beta, \qquad (19)$$

where B is the amplitude of the sine function, α is the vibration cycle of the sine wave, β determines the shift of this autocorrelation curve, and θ defines the time offset. The parameters in Eq. (19) are in fact signal dependent. In other words, the threshold would not possess the same value when different signals are tested, but rather the parameters should be re-assigned every time based on the characteristics of the different audio pieces.

After passing the echo existence check, the blocks that have been verified as echo embedded must be further categorized to establish whether there exists a 'zero' echo or a 'one' echo. Determining the echo delay value involves a simple comparison of the autocorrelation values. The delay time, which produces the greater autocorrelation value, represents the proper echo value. For example, if $Cor(d_0) > Cor(d_1)$ in a segment, then the embedded bit is digit '0' in that block, and vice versa.

The steps shown in Fig. 15 are repeated block by block until all of the watermarked information has been extracted.

2.2.4 Recovery Accuracy Check

The final part of the echo watermarking procedures is to check the extracted data and to determine whether it is equivalent to the particular information by comparing these two data stream. This check is called watermark recovery accuracy check, where the recovery accuracy rate is defined as

$$recovery\_accuracy\_rate = \frac{number\_of\_bits\_correctly\_extracted}{number\_of\_bits\_embedded} * 100\% . \qquad (20)$$

The recovery accuracy rate hardly reaches 100% due to occasional errors, which occur when extracting the watermark. The reasons for these recovery errors are discussed in the following sections. Since the recovery cannot be 100%, an appropriate accuracy threshold is needed to decide whether the particular watermarking data and the extraction data are equivalent.

2.3 System Control Parameters

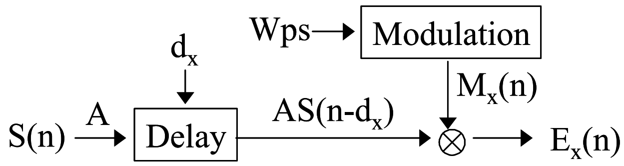


Fig. 17 Control parameters and system architecture

The quality of a watermarking technique is primarily measured by the audio quality, or by the inaudibility of the watermark, as well as by the recovery accuracy rate of the embedded information. Both the inaudibility and accuracy rate are tightly coupled to the assignments of the control parameters on echo magnitude (*A*), watermark-embedding rate (*Wps*) and delay for echoes (d_0 and d_1). In this section, the influence of each parameter on the watermark quality is discussed. In advance of the discussion on these parameters, the relationships of the signals within each procedure must be deduced.

2.3.1 System Architecture

The system architecture and its relationship with the control parameters are shown in Fig. 17. While *S(n)* represents the original host signal, the delayed signals are defined in Eqs. (3) and (4), and the modulated echo signals are defined as Eqs. (8), (9) and (10).

Substituting Eqs. (3) and (4) into Eqs. (8) and (9) yields

$$E_0(n) = AS(n-d_0) * M_0(n) \quad (21)$$

and

$$E_1(n) = AS(n-d_1) * M_1(n). \quad (22)$$

Substitute Eqs. (10), (21) and (22) into Eq. (11) can obtain the combined embedding signal *E(n)* as

$$E(n) = AS(n-d_0)*M_0(n) + AS(n-d_1)*M_1(n) + AS(n)*M_N(n) . \quad (23)$$

Replacing the *E(n)* term in Eq. (13) with the above expression yields

$$S(n) = \frac{S(n)+AS(n-d_0)*M_0(n)+AS(n-d_1)*M_1(n)+AS(n)*M_N(n)}{\sqrt{1+A^2}}. \quad (24)$$

In fact, only one of the embedding signals, $E_0(n)$, $E_1(n)$ and $E_N(n)$, should have a non-zero value at any time because only one of the modulators values can have a value of one in any block, when the block boundary condition is not considered. When the embedded bit is digit ‘0’, *S’(n)* would simplify to

$$S'(n) = \frac{S(n) + AS(n-d_0)}{\sqrt{1+A^2}}. \quad (25)$$

Likewise, if the embedded watermark bit is digit ‘1’, *S’(n)* is

$$S'(n) = \frac{S(n) + AS(n-d_1)}{\sqrt{1+A^2}}. \quad (26)$$

Otherwise, when the block contains no watermark, *S’(n)* becomes

$$S'(n) = \frac{(1+A)S(n)}{\sqrt{1+A^2}}. \quad (27)$$

Based on the *S’(n)* functions, Eqs. (25), (26) and (27), and the autocorrelation definition functions, Eqs. (15) and (16), the discussion of the parameters influence can now proceed.

2.3.2 Echo Magnitude

The magnitude of echo is assigned for the decay percentage of delay signals. The magnitude value assignment needs to consider both the recovery accuracy rate and the audio quality. When the echo magnitude is too large, it can be directly seen that the audio quality of the watermarked signal would degrade. On the other side, when the echo magnitude is too small, the recovery accuracy rate would become too low for correctly identify the watermarked content. The relationship between the recovery accuracy rate and the echo magnitude is illustrated as follow.

The accuracy of watermark extraction is highly related to the difference between $Cor(d_0)$ and $Cor(d_1)$. In other words, if the theoretical value of $|Cor(d_0) - Cor(d_1)|$ is large, then the actual value would be less influenced by the local characteristics of any single block signal. For example, the autocorrelation value of a small block in the original host audio may have a sudden peak at time d_0, while in reality a watermark digit ‘1’ is to be embedded. After the watermark digit ‘1’ embedded, it is hard to say whether $Cor(d_1)$ would be larger than $Cor(d_0)$ after the echo is added, or if $Cor(d_0)$ would be larger because of the local property of the host audio segment. Further examination of $Cor(d_0)$ and $Cor(d_1)$ is as follows:

Assuming the situation described in the previous paragraph occurs, than the appropriate expression for *S’(n)* comes from Eq. (26) where digit ‘1’ is embedded. Substituting *S’(n)* into the autocorrelation function, then the autocorrelation values of d_0 and d_1 are

$$Cor(d_0) = \frac{\sum(S(n)S(n-d_0) + AS(n)S(n-d_1-d_0) + AS(n-d_1)S(n-d_0) + A^2S(n-d_1)S(n-d_1-d_0))}{\sum(S(n)+AS(n-d_1))^2} \quad (28)$$

and

$$Cor(d_1) = \frac{\sum(S(n)S(n-d_1) + AS(n)S(n-2d_1) + AS^2(n-d_1) + A^2S(n-d_1)S(n-2d_1))}{\sum(S(n)+AS(n-d_1))^2}. \quad (29)$$

In Eq. (28), the value of terms $S(n)*S(n-d_1-d_0)$ and $S(n-d_1)*S(n-d_0)$ can be neglected based on the autocorrelation characteristics, which assumed that the autocorrelation function has peaks only at time d_0 (the original host audio segment characteristic) and at time d_1

(where the echo embedded). Thus, Eq. (28) is reduced to

$$Cor(d_0) = \frac{\sum (S(n)S(n-d_0) + A^2 S(n-d_1)S(n-d_1-d_0))}{\sum (S(n) + AS(n-d_1))^2} \cdot (30)$$

In Eq. (29), under the same reasoning, the term $S(n)*S(n–2d_1)$ can be ignored, and the equation is reduced to

$$Cor(d_1) = \frac{\sum (S(n)S(n-d_1) + AS^2(n-d_1) + A^2 S(n-d_1)S(n-2d_1))}{\sum (S(n) + AS(n-d_1))^2} \cdot$$

(31)

It is still unclear whether the value of $Cor(d_0)$ would be larger than $Cor(d_1)$ when A is small, because the terms $S(n)*S(n–d_0)$ and $S(n)*S(n–d_1)$ are both large. When A is large, the value of $Cor(d_1)$ would certainly increase more quickly than the value of $Cor(d_0)$. By inspection, the sub-terms of $Cor(d_1)$ are dominated by the $A*S^2(n–d_1)$ term instead of the $A^2*S(n–d_1)*S(n–d_1–d_0)$ term, because A is always larger than A^2, since $0<A<1$ is assumed.

An induction from the above condition is made.

Hyporeport 1*: "When a larger echo magnitude is assigned, a higher recovery accuracy rate can be attained, and vice versa."*

This hyporeport is verified with system simulation in Chapter 4, and the proper range for echo magnitude assignment would be given.

2.3.3 Watermark Embedding Rate

The watermark-embedding rate is defined as the number of blocks can be segmented within one single second audio. In other words, Wps is the number of watermarking information units can be embedded in per minute audio signal. The extractability of the hidden information and the information transmission rate need to be considered before assigning the watermark-embedding rate. The upper boundary of the watermark-embedding rate is constrained by the extractability, because the watermark could hardly be correctly extracted when the block size is too small. However, we need a larger watermark-embedding rate to get a higher information transmission rate. The value constrain of the watermark-embedding rate is discussed as follow.

Based on the same situation assumed in last section: There exists an autocorrelation peak at time d_0 in the block, and the embedded watermark digit is '1'. The corresponding expression for $Cor(d_0)$ and $Cor(d_1)$ are defined in Eqs. (30) and (31), and A now has a fixed value. If the difference of $Cor(d_0)$ and $Cor(d_1)$ is calculated term by term, $|Cor(d_0) – Cor(d_1)|$ is

$$\Delta = \frac{\sum (S(n)(S(n-d_1) - S(n-d_0)) + AS^2(n-d_1) + A^2 S(n-d_1)(S(n-2d_1) - S(n-d_1-d_0)))}{\sum (S(n) + AS(n-d_1))^2} \cdot$$

(32)

On average, as a small difference is gathered from one summation term, the large difference can be reached with many terms. In other words, with a larger block size, the difference between the autocorrelation values would be more obvious. The local autocorrelation characteristic of a small block signal may not be representative of the overall characteristic of the entire host signal. When the segment size is larger, there is a higher probability to reduce the random fluctuations in the summation terms that may induce extraction error.

We can make another induction based on the discussion above.

Hyporeport 2*: "When a lower watermark-embedding rate is given, and consequently, larger block size is assigned, a better recovery accuracy rate can be obtained, and vice versa."*

This hyporeport is also verified with system simulation in Chapter 4, and the proper range for watermark-embedding rate assignment would be given.

2.3.4 Delay Values and Delay Distance

When generating the echo signals for a watermark, a varied delay time is assigned for each echo. The influence on recovery error rate caused by delay times is a little different from the one caused by echo magnitude and watermark-embedding rate. This kind of error is much more related on the autocorrelation characteristic of both the local block and the overall watermarked audio signals. As shown in Fig. 16, the autocorrelation value is pretty high when the time distance from the origin is small. This property implies when the delay value is small, the autocorrelation calculation of the watermarked audio will be primarily influenced by the high autocorrelation of the host signal. In other words, there will be some instances when in fact the watermark digit '1' is embedded, but the result $Cor(d_0) > Cor(d_1)$ is obtained for a large autocorrelation value that existed at time d_0. When $d_1>d_0$ is assumed, the extraction algorithm would not be able to determine whether the watermark is digit '0' or digit '1', because the delay d_0 lies within the section where the large original autocorrelation value is initially large. In this situation, although $Cor(d_0)$ is larger than $Cor(d_1)$, the influence by the existing high correlation at time d_0 may overwhelm the correlation that is produced by the echo at time d_1.. Many errors may occur by improperly extracting digit '1's as digit '0's. As the value of d_0 decreases, the likelihood of extraction errors increases. Thus, delay values should be constrained by the autocorrelation characteristic of the host audio, and must not be too small in order to avoid the errors caused by high correlation of the original audio. Based on the experimental results shown in Fig. 16, the average allowed value of the shorter delay should be larger than 0.6msec to avoid the influence from the high correlation of the original signal.

Similar problems may also occur when the delay distance, $|d_0 – d_1|$, is too small, for the relationship between the two delay signals is much the same as the one between the delay and the original signals. Examine Eqs. (28) and (29). Assume that $d_0>d_1$ and the embedded bit is '1'. When the distance between d_0 and d_1 is very small, the $S(n)*S(n–d_1–d_0)$ term in $Cor(d_0)$ becomes fairly significant. Although $S(n)*S(n–d_1–d_0)$ is theoretically smaller than $S(n)*S(n–d_1)$, there is still a high autocorrelation at time d_1+d_0 in the watermarked audio due to the echo that exists at time d_1.. This correlation makes $S(n)*S(n–d_1–d_0)$ a large percentage of $S(n)*S(n–d_1)$, which is much greater than zero, and cannot be ignored. Under this situation, if the other additional terms of $Cor(d_0)$, such as $S(n)*S(n–d_0)$, are not small either, an extraction error would likely occur. The difference between $Cor(d_0)$ and $Cor(d_1)$ would not be distinct enough due to the correlation between the two autocorrelation functions.

From this analysis, two assumptions are made based on the small delay distance problem described above.

Hyporeport 3*: "An extraction error is more likely to occur when the smaller delay lies too close to the original signal."*

Hyporeport 4*: "Better recovery accuracy rate can be reached when a longer delay distance is assigned, and vice versa."*

The two hypotheses about delay are verified with system simulation in Chapter 4, and the proper range for delay value assignment and the suitable distance between the delays would be given.

2.3.5 Analysis Based on Parameters

As mentioned before, the recovery accuracy rate and the quality of the watermarked audio are the most important properties of a watermarking technique design. While the accuracy rate of extraction is highly dependent on the assignment of the control parameters in the echo data hiding techniques, the combinations of various control parameters assignment would obtain quite different results. In order to achieve a desired recovery accuracy rate, some proper combinations of the control parameters must be selected. In practical application, the highest recovery accuracy rate is not always required, and we may only wish for an acceptable accuracy with a better audio quality or a higher watermark-embedding rate. Thus, when a parameter is assigned for high accuracy rate, the others can be selected from a much larger range and still obtain the acceptable recovery accuracy rate.

From *Hyporeport 1* and *Hyporeport 2*, we can make another assumption:

Hyporeport 5*: "When a larger magnitude is assigned, a higher watermark embedding rate (smaller block size) is allowed for an acceptable recovery accuracy rate." In other words, "Lower watermark embedding rate (larger block size) is needed to achieve an acceptable recovery accuracy rate when smaller magnitude is given."*

This assumption is useful for us to adjust the control parameters when only a certain level of accuracy rate is required. For example, if we increase A, the recovery accuracy rate will consequently increase. But if this increased recovery accuracy rate is not needed, we can then instead increase the Wps for higher watermark-embedding rate, which is another important property of the watermarking techniques. Or on the opposite side, when a low Wps is given as the high recovery accuracy is reached, a smaller A can be assigned to make the audio quality of watermarked audio better.

Based on *Hyporeport 1*, *Hyporeport 2* and *Hyporeport 4*, the additional correlations between the parameters can be deduced.

Hyporeport 6*: "When a longer delay distance is chosen, a smaller magnitude is allowed for an acceptable recovery accuracy rate."*

Hyporeport 7*: "When a larger delay distance is given, a suitable recovery accuracy rate is achieved with a higher watermark embedding rate."*

The hypotheses above give more flexibility to the assignment ranges of the parameters in echo data hiding techniques, and they serve as the foundation for some of the modification techniques we would like to propose in the next chapter.

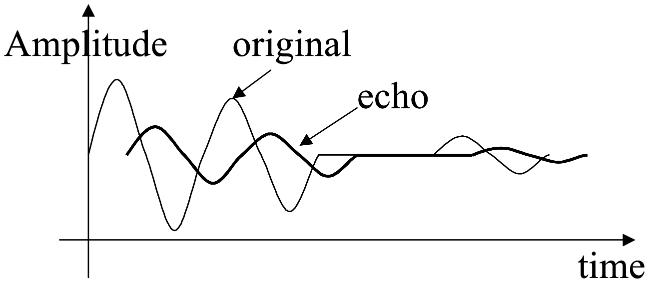


Fig. 18 Echo makes no difference in silence

Chapter 3 System Modification

Several problems exist in the echo data hiding techniques. One disadvantage of this watermarking method is the echo signal cannot exist in the segments of silence. Any decay of magnitude zero is still zero, there is no use trying to add an echo in the silent parts. Thus, the echo watermark embedded in the silent regions cannot be correctly extracted and would increase the recovery error rate. Another weakness of the echo data hiding techniques is its low information transmission rate comparing with some other approaches. In the previous system design, only one bit of watermark information (‘0’ or ‘1’) can be added in a single block of the host audio. Generally speaking, in echo data hiding, only about two to 64 blocks can be segmented for echo embedding in each second of the audio signal, while 44.1k bits of data per second is allowed in low-bit coding approaches [14]. In a word, the watermark-embedding rate of the echo data hiding system is too low for information transmission, and we wish to increase the information transmission rate.

In this chapter, we will focus on these two problems of the echo data hiding techniques mentioned above in hopes of improving the echo data hiding techniques as a better watermarking system.

3.1 Bypassing Silence

Since embedding echoes in the silent or small energy blocks is useless for extraction of the hidden echo data, skipping these blocks may be the only choice. To skip the silent or small energy blocks, the system of echo data hiding is modified. The system modification is adding a procedure of energy check, in which a threshold is assigned for bypassing the silent blocks and the small energy blocks.

3.1.1 Causes of Extraction Errors

It can be directly observed that the added echo makes no difference on the signal magnitude in silent segments, as Fig. 18 shows. The extraction error occurs when a digit misread is made in watermark extraction procedure or when the embedded digit could not be decided by autocorrelation check; that is, the added echo makes no difference.

Proof that an echo could not exist in the silent segments can be easily demonstrated. A signal with an echo can be represented as

$$S'(n) = S(n) + AS(n-d)\,. \qquad (33)$$

When $S(n)$ in the segment is zero, $S(n-d)$ is also zero in the same segment. The extraction of watermark data depends on the difference between the autocorrelation values of the watermarked signal at the two assigned delay positions. As long as there is no difference between $Cor(d_0)$ and $Cor(d_1)$ in the silent segments, for the signals are both zero, it is useless to embed data into these silent segments.

Now let us consider the segments with small energy. If the magnitude of the sample is only about one-quantization-step large, which is represented as ‘one’ after quantization, adding a small echo on the sample would not make any difference in the resultant sample magnitude, which is still ‘one’.. The digital audio signals are usually quantized before storage. Performing quantization means replacing the signal magnitude with its nearest integer. When embedding an echo, for any integer K,

$$\text{nearest\_int}(K + AK) = K \qquad \text{when} \quad 0 < A < \frac{1}{2K} \qquad (34)$$

and

$$\text{nearest\_int}(K + AK) \geq K + 1 \qquad \text{when} \quad \frac{1}{2K} \leq A < 1 \,. \qquad (35)$$

From Eqs. (34) and (35), in order to make a difference in magnitude after data embedding, A must be larger than 1/2 when K=1, and be bigger than 1/4 when K=2, etc. That is to say, when the amplitude of a sample is K and an echo with scaling factor A is added, the value of $A*K$ must be larger than 0.5 to make a difference in the

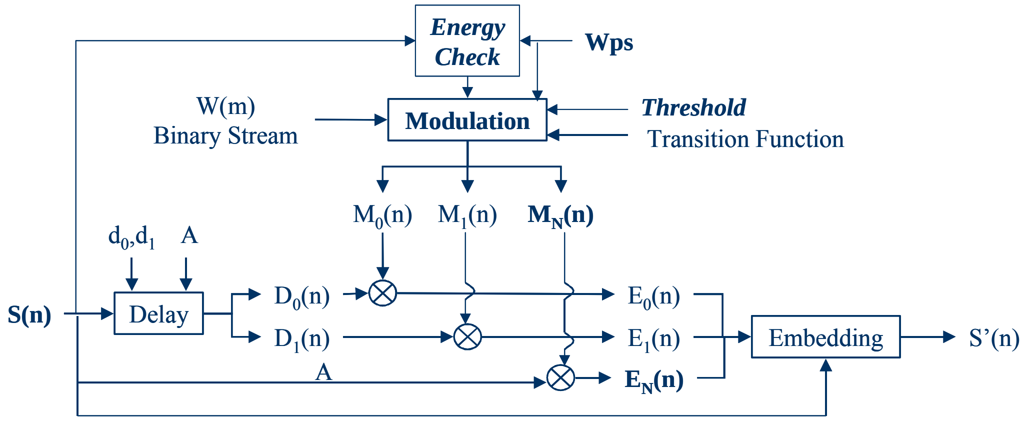


Fig. 19 Watermark embedding procedure with energy check module

resulting amplitude, when the amplitude is always rounded off to the nearest integer. The effect of this quantization error causes reduction of the watermark extraction accuracy rate.

To avoid the recovery error caused by quantization or by silent segments, the average energy of each block should be checked before the watermark embedding, and the watermark data should not be embedded into these segments. So, we decided to add an energy check module into the echo data hiding system to avoid the watermark being embedded into any of the unsuitable block.

3.1.2 Modified System I

The echo data hiding system is modified to skip the blocks unsuitable for echo embedding, that is, the silent or small energy blocks. To modify the system, an energy check module is added into the embedding stage of the watermark system, and before the modulation module. The modulation module is also modified, by creating a block without watermark embedding, which is decided by an energy threshold check. In order not to embed any data in the block, the modulator $M_N(n)$ is assigned as a magnitude of one in the segments that have an average energy below the assigned threshold, while the other modulators are assigned a magnitude of zero. The average energy of a block is defined as

$$Average\_Energy = \frac{\sum S^2(n)}{segment\_size}. \qquad (36)$$

The average energy of each block is calculated, and the result, which varies with different watermark embedding rates, is sent into the modulation procedure. Then a proper threshold is defined for the creation of modulators, as Fig. 19 illustrates. The energy check of audio signal is combined with the watermark data stream for modulators generation, where the none-watermark-embedding blocks must be allowed. All of the following procedures remain unchanged and are not influenced. There is no need to do the energy check in the extraction stage, because the segments without echo embedding would not have passed the echo existence check.

3.1.3 Theoretical Analysis of Improvement

The improvement made by the modified echo data hiding system is the decrease of error extraction rate. The error extraction of the echo watermark is caused by two major reasons: one is the original audio autocorrelation characteristic of the block, and the other is the echo cannot work in the small energy or the silent blocks, especially when audio compression is processed. The former may be improved by increasing the magnitude of the echoes or by lowering the watermark-embedding rate, as discussed in the previous chapter. The latter is now avoided by skipping the watermark embedding in the silent or small energy blocks.

After bypassing the blocks with an average energy beneath the threshold, the recovery accuracy rate should be increased. The percentage increase in the accuracy rate should be linearly dependent on the percentage of small energy blocks in the host audio signal. Although the watermark bit can be transmitted is somewhat decreased when some of the blocks are unable to embed any watermark data after the energy threshold check, it does not really matter because these blocks are hardly any good for correctly extracting the watermark. The increase in the accuracy rate would be more than compensate the loss of embedding rate for the whole system, while the total performance is measured as

$$Performance = information\_transmission\_rate \times recovery\_accuracy\_rate. \qquad (37)$$

where the information transmission rate, which means the bits of information transmitted per minute, equals to the watermark embedding rate under binary watermark representation system.

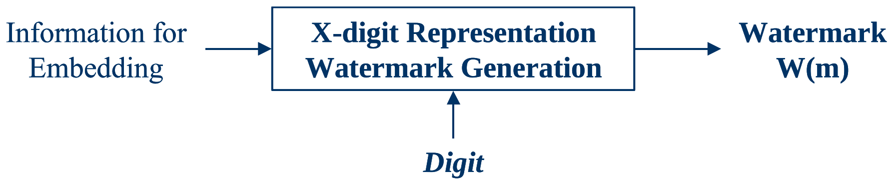


Fig. 20 Various formats watermark creation

The threshold of the energy check must be carefully selected to achieve better system performance, which should be about two or three quantization steps large. Furthermore, the energy check module helps to increase the recovery accuracy after audio compression. Because when audio compression is performed, the small energy blocks are usually the first destroyed parts because they were not originally audible and whether the energy existed or not made little difference. As long as the small energy signal cannot survive through audio compression processing, the echo watermark embedded on the signal cannot survive either. When performing audio compression, the energy threshold can be assigned according to the psychoacoustics model and the quantization step size of compression algorithms. Thus, the signal energy check procedure with threshold adopted to varying algorithms could improve the recovery accuracy rate of the echo data hiding system during audio compression. By increasing the embedded watermark survival percentage, the robustness of echo data hiding system is improved.

3.2 Multiple Bits in Single Block

One of the major disadvantages of the echo data hiding techniques as compared with other watermarking approaches is the relatively low information transmission rate. The information transmission rate of the original echo data hiding techniques, which is equal to the watermark-embedding rate, is constrained by the requirement of a sufficiently long block for a bit of watermark to be embedded as an echo and to survive. To add an echo in the signal, the segment must be long enough for covering both the delay and transition parts to exist, as well as have enough samples for correctly extracting the embedded watermark, which is discussed in prior chapter. To divide more blocks in a fixed length of audio signal is almost impossible because the limit mentioned above. To increase the information transmission rate, embedding more data into a single block seems like a good idea. If more than one bit of data can be embedded into a segment, the information transmission rate would undoubtedly increase.

3.2.1 Trinary System or More

The information transmission rate is the amount of watermark information can be transmitted per minute, which is controlled by the watermark embedding rate and the number of bits embedded in each block, as

$$information\_transmission\_rate = Wps * bits\_per\_block. \quad (38)$$

As long as the *Wps* is constrained by the minimum allowable block size, while embedding only a single bit of data in a block is not adequate, inserting multiple bits into the audio stream has been tried. Originally a binary stream format is used for the watermark data, where a unit of the stream represents one single bit of data. To satisfy the need of a higher information transmission rate, the trinary format is a possible solution. The trinary representation system uses not only digits ‘0’ and ‘1’ as in the binary format, but also uses digit ‘2’ to represent the data. For example, the text string “ABC” is represented as “010000010100001001000011” in binary format, which requires 24 units, while using the trinary format, it requires only 15 units to present the text string, as “021020211002111”. Table 3 shows the number of bits can be represented by one unit of varying digit representation system. Using more digits to represent the data, more bits of data can be represented in each single unit; the relationship is given by

$$no\_of\_digit = 2^x, \quad (39)$$

where x is the number of bits can be represented by a unit in the representation system, and

$$x = \ln_2(no\_of\_digit). \quad (40)$$

Besides the binary format, more digits can be used for representing the data and thereby increase the information transmission rate. Although more digits can represent more data in a block, there are some limitations on increasing the number of digits used in the watermark format in echo data hiding techniques. For example, the delay distance must be larger than a threshold to achieve an acceptable accuracy rate; using more digits means more echoes with different delays are needed, and the requirement of block size becomes even larger. Also, a smaller magnitude must be used for the echoes for the delay signals to be masked by the original. Using too many or too few digits in the watermark data format would not necessarily improve the system performance. A suitable balancing point needs to be determined.

3.2.2 Modified System II

Change of the watermark representation format requires many procedures related to the watermark format in the echo data hiding system to be modified. Thus, data can be represented by various representation systems, and be embedded into the host audio signals. The watermark generation procedure is adjusted to handle different digit representations, as Fig. 20 illustrates. The output watermark stream is formatted to the assigned representation system. For example, if a trinary system is assigned, the watermark data stream is a combination of zeros, ones and twos. The change of watermark format also influences the echo data embedding and extracting procedures.

The number of echoes generated in the embedding procedure is increased with the number of digits used for the watermark data, one echo per digit. For example, when the trinary system is used, digits ‘0’, ‘1’ and ‘2’ are mapped into the echoes with delay values d_0, d_1 and d_2. For each of these delay values, a corresponding echo signal is generated, and a corresponding modulator is created. The watermarked signal output is a combination of these embedding signals, while the reminder of the algorithm is kept approximately the same. The whole system is illustrated in Fig. 21. The delay signals $D_x(n)$ are defined as

$$D_x(n) = A * S(n - d_x), \qquad (41)$$

where x is the index from zero to one less than the number of digits used for the watermark. The embedding signals are defined based on the delay signals and the modulators as

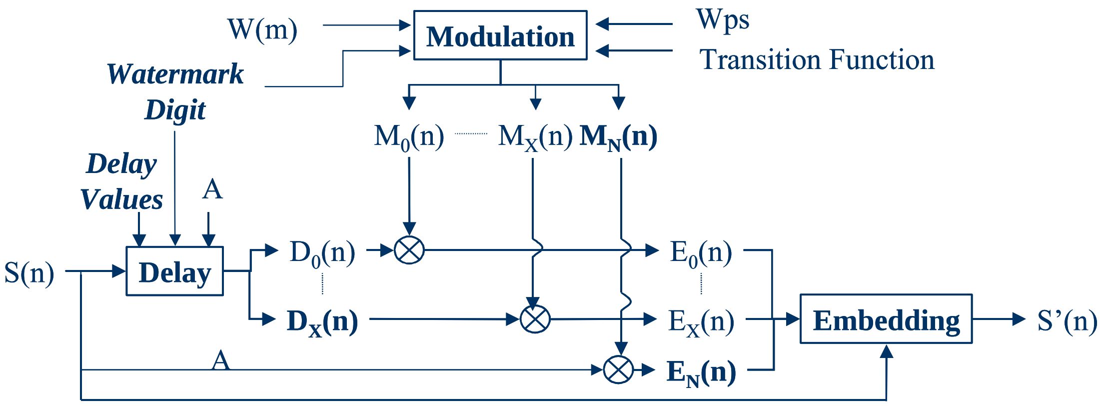


Fig. 21 Watermark embedding procedure with multiple echoes

Table 3 Number of bits represented by single unit in x-digit system

No. of Digits	2	3	4	5	6	7	8
Bits in a Unit	1	1.585	2	2.323	2.585	2.807	3

$$E_x(n) = D_x(n) * M_x(n)\,, \quad (42)$$

where x is the integer index from zero to one less than digit-number and $E_N(n)$ has the same definition as Eq. (10). Based on the embedding signals, the watermarked signal is given by

$$S'(n) = \frac{S(n) + E_N(n) + \sum_{x=0}^{digits-1} E_x(n)}{\sqrt{1+A^2}}\,. \quad (43)$$

The watermark extraction procedure must also be modified after watermark representation system changed. More autocorrelation values need to be calculated, and more terms need to be compared in order to determine the location of the echo, as Fig. 22 shows. The autocorrelation calculation is given by

embedded signal would be distinguished from the original signal. The delay values are limited by minimum allowable delay distance, and a sufficient distance is needed in order to correctly extract the watermark. To satisfy the demand of delay distance, the largest delay value for the multiple echo data hiding system would not be small, and this long delay restricts the allowable magnitude for the echoes. With a smaller signal magnitude, the proper range for watermark-embedding rate assignment may also be reduced.

These influences operate in a vicious circle. The information transmission rate is increased, by using more digits to represent the watermark data, and more echo signals are consequently generated. More echoes limit the range for assigning delay values, thus limiting the maximum allowable magnitude of the embedding signals. However, a smaller echo magnitude infers a lower recovery accuracy rate can be obtained. Thus decreases

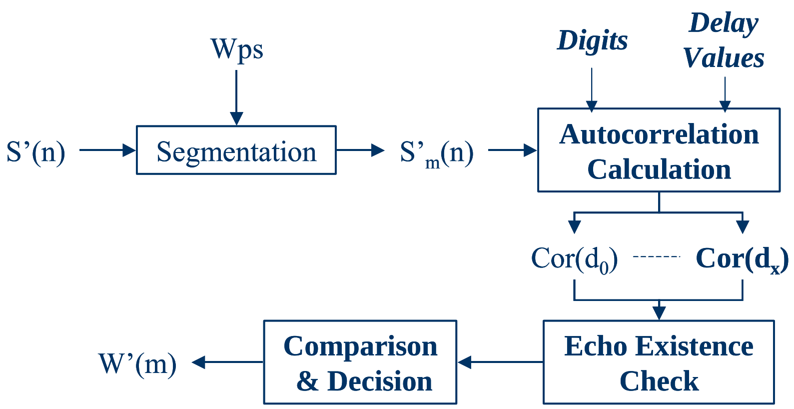


Fig. 22 Watermark extraction procedure with multiple echoes

$$Cor(d_x) = \frac{\sum S'(n)S'(n-d_x)}{\sum S'(n)S'(n)} \qquad 0 \le x \le digit\_no - 1\,.$$

(44)

The algorithm for the watermark extraction procedure in multiple bits echo data hiding is approximately the same as the procedure in binary approach. The procedure of recovery accuracy check is unchanged. This modified echo data hiding system can adopt various watermark representation formats.

3.2.3 Theoretical Analysis of Improvement

In the modified system, the design of multiple bits echo embedding is to improve the information transmission rate of the system. The influences caused by assigning different parameters that were discussed in the previous chapter also exist in the multiple bits echo data hiding system. The magnitude of the echo is constrained by the audio temporal masking effect. The longer the delay, the smaller the magnitude can be; otherwise the

the watermark-embedding rate while the acceptable recovery accuracy rate is required, based on the parameter assignments tradeoff between the watermark embedding rate and the echo magnitude. Because the watermark-embedding rate and the recovery accuracy rate are linear dependent on the information transmission rate, the information rate is decreased after lower *Wps* or smaller A is assigned. After the cycle is complete, the degree of increase rate of information transmission is a big question. The performance of watermarking system is defined as

$$performane = informatio\_transmision\_rate * recovery\_acuracy\_rate\,,$$

(45)

where the information transmission rate is defined as Eq. (38). There is an obvious tradeoff between the *Wps* and the number of digits used in the watermark data, and an optimal point for the overall system performance must be found.

In the next chapter, all of the experimental results are shown and discussed. First, the results about parameter properties are analyzed. Second, some robustness tests were performed to examine the watermarked audio generated by the echo data hiding system. Then the modified system with energy threshold was conducted to verify its efficiency. Finally, multiple echo data hiding systems were examined with various watermark formats. The overall performance of each echo data hiding system is compared against the others.

Chapter 4 Experimental Results and Analyses

In this chapter, all of the experimental results are shown. Each hyporeport proposed in the prior chapters is verified with experiments. From the experimental results, the proper ranges of control parameters assignment are suggested. The robustness of the embedded watermarks in the echo data hiding system is also tested, which is focused on the noise distortion and the audio compression. Both of the modified systems, bypassing silence and multiple bits in single block, are set up and the performances are examined on increase of the recovery accuracy rate or raise of the information transmission rate. System performance of the modified systems are also analyzed and evaluated.

4.1 Experimental Environment

Before the experiments were conducted, the audio pieces as testees needed to be selected, and the format of the audio files had to be decided. The criteria of assigning the values of the control parameters for experiments are defined. The recovery accuracy rate of the acceptable performance is also clarified.

All of the experiments are made on Windows 98 operating system, with YMH XG/128 4C sound card. The application software Cool Edit 2000 is used for the audio signal processing, including the digital recording and the MP3/WAV format conversion, and the WMA/WAV format conversion is done with the software Winamp, version 2.50.

4.1.1 Audio Files for Experiment

In order to minimize the degradation of audio signal due to the quantization error and to maintain a high audio quality, the audio pieces used for embedding watermarks are in 16-bit format, and the sampling rate are 44.1kHz. This resolution (16-bit) is the usual audio resolution used in ordinary music CDs (compact discs), such as the WAV (waveform) format. Sixteen-bit linear quantization introduces a negligible amount of signal distortion [12]. In fact, the PCM (Pulse Code Module) format files, which are converted directly from the WAV files, are used in the experiments, because the PCM file stores each audio sample as a 16-bit signed integer and is more convenient for magnitude calculation in the embedding and extraction procedures.

The pre-selected audio pieces were chosen to represent the typical types of audio signals. Six pieces of audio were chosen as:

1. Melodie in F Op. 31, by Anton Rubinstein, from Klassik Träumereien
2. Piano Concerto No. 1 in B Flat Minor, Op. 23, by Tchaikovsky, from the Classical Collection
3. Kanon und Gigue, by Johann Pachelbel, from Klassik Träumereien
4. Turkish March, by Beethoven, from The Best of Piano I
5. Famous People: Albert Einstein, from Studio Classroom Basic article reading CD, May 1997
6. Violinkonzert Nr. 3 Adagio, by Wolfgang Amadeus Mozart, from Klassik Träumereien

Trying to include various kind of music, the pieces of cello, piano, and violin were selected to representing different waveform of various instruments; and pieces of solo, quartet, and symphony were chosen for comparison of varying number of players. These audio pieces can represent several typical kinds of music. The piece of canon was especially selected because of its self-repetition characteristic, which may become a confusion of echo watermark extraction. Still a segment of reading was chosen to represent the human voice. These selected audio pieces are testees of every experiments described in the following sections. Each of the selected audio pieces is about two minutes long to gather enough blocks for watermark embedding and for the statistical calculation of recovery accuracy rate when the watermark-embedding rate is low.

4.1.2 Masking Test

When the temporal masking effect of audio is considered, the magnitude of signal that can be masked is decreased from the time masker removed. Before the watermark is embedded, a masking threshold test is made to determine what magnitude can be added into the audio with the assigned delay time. This is an objective test, which is just for reference, and the allowable magnitude is determined by human ear.

In the auditory threshold test of maskee, the embedded signals are echoes of the original signal. The added echoes are generated with various delay times and different magnitudes. At each delay time tested, from 0.25 to 5msec with step 0.25msec, the highest allowable magnitude of the masked echo is decided. The test result is shown in Fig. 23. From the test result, the highest allowable magnitude of the embedding echo at 1.5msec delay is about 0.75, and the one at 3msec delay is about 0.55. When two echo signals are generated, the larger delay offset is usually at the time segments between one and two millisecond, and the allowable echo magnitude is between 0.8 and 0.6. When three or more echoes are generated, the largest delay offset is usually positioned in the range from 2.5 to 3.5msec, and the allowable echo magnitude is about 0.45 to 0.65. When the value of magnitude needs to be fixed in some experiments, it is assigned a value in the range from 0.5 to 0.75, based on the observations above.

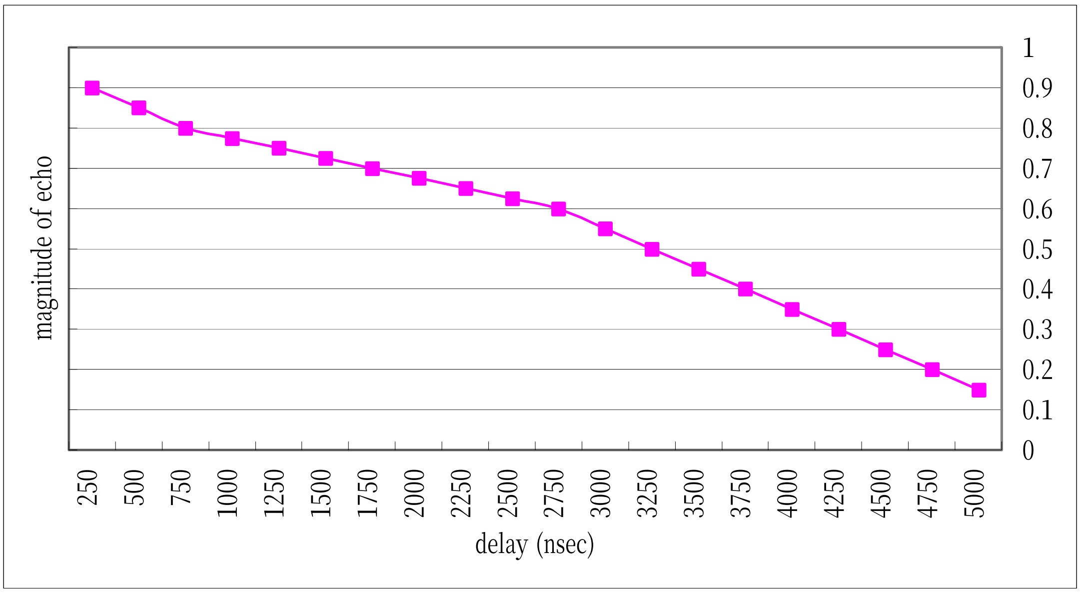


Fig. 23 The largest maskable echo magnitude at different delays

The audio signals embedding single echo are created, with varying magnitude (from 0.05 to 0.95, step 0.05) and varied delay (from 250 to 5000 nanosecond, step 250nsec). The highest magnitude is picked as where noise could not be detected by the testers' ear.

4.1.3 Definitions

There are several major control parameters in the echo data hiding system, which have been defined in prior chapters.

- ***Magnitude (A)***

The magnitude is the decay of echo signal amplitude compared to the original signal amplitude. A value larger than 0 and smaller than 1 is required for A, where the former permits the existence of echo and the latter makes the echo smaller than the original and could be masked.

- ***Watermark embedding rate (Wps)***

The watermark-embedding rate (*Wps*) is defined as the number of units that can be embedded in each second of audio segment. The minimum allowable block size for correctly watermark extraction is the upper bound of the *Wps* assignment.

- ***Delay values***

The delay values of echoes are assigned based on the time unit of the audio temporal masking, millisecond, and are converted to the number of samples of the corresponding audio signal. For example, if a one-millisecond delay is assigned and the sampling rate of the audio is 44.1kHz per second, then 44 samples are offset for the delay signal generating.

Robustness tests are used to examine the performance of the watermarking techniques. Under varying parameters given, the robustness of watermarks are tested and compared. The two major ways used for robustness tests are noise addition and compression processing. For the noise tolerance tests, the digital watermarked signals are converted into analog form by playing the audio and then converted to the digital format again by recording it into the computer. In the experimental process, the watermarked signals are outputted from the sound card through a one-meter long audio line (D/A) and input into the sound card again (A/D) from this audio line. This 'D to A to D' test is to simulate the condition of recording the broadcasted audio signals and saving them in digital format, which frequently happens.

For audio compression tolerance tests, the watermarked signals are compressed into MP3 (MPEG-1 audio layer 3) format and WMA (Windows Media Audio) format, which are two of the most popular audio compression formats. The compressed audio signals are used as the inputs of the watermark extraction procedure of the robustness tests and the recovery accuracy rates are examined. The recovery rates after robustness tests are compared with the accuracy rates of the watermarked audio before noise addition or compression process. The surviving percentages of the watermarked information with varying control parameters assigned in the echo data hiding technique are the grading of the parameters.

For easier verification of the quality of extracted watermark information, 0.75 is defined as the acceptable recovery accuracy rate. A combination of control parameters with recovery rate above 0.75 is considered as suitable value assignment.

4.1.4 Comparison of Audio Pieces

To determine the overall proper value ranges for the control parameters assignments, the difference between the experimental results of various audio pieces is analyzed. Several combinations of control parameters were assigned for the performance comparison of various audio pieces, as illustrated in Fig. 24. From the

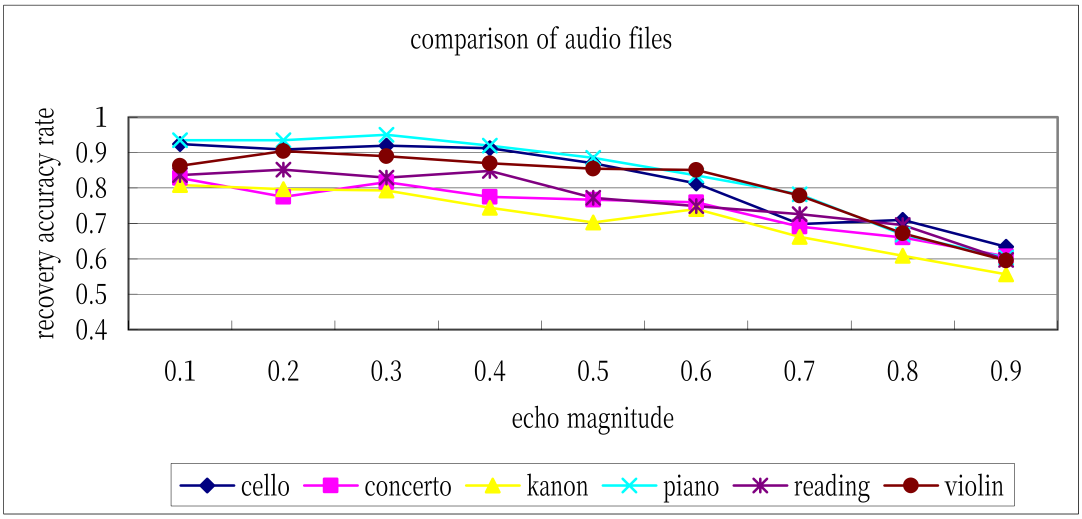


Fig. 24 The recovery accuracy rate of different audio pieces

A=0.1~0.9, step 0.1; Wps=2; d_0=0.8msec; d_1=1.2msec

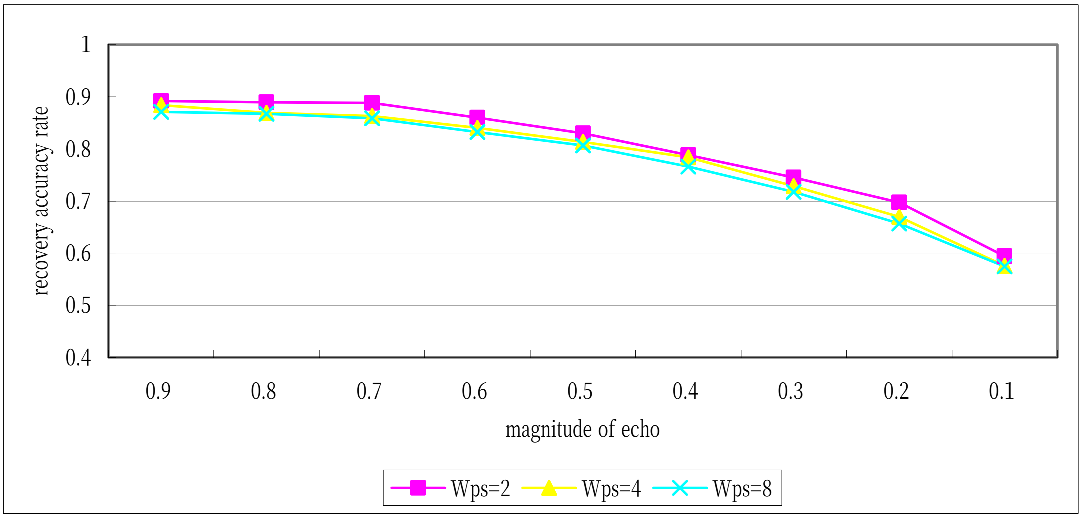


Fig. 25 Watermark recovery accuracy rate related to varying magnitudes

A=0.1~0.9, step 0.1; Wps=2, 4, 8; d_0=0.8msec; d_1=1.1, 1.2, 1.3msec

experimental results, the recovery accuracy rate of 'kanon' is the lowest one, which may be influenced by its self-repeating characteristic. The highest recovery accuracy rate occurs on the audio piece 'piano', which has a pretty pure waveform based on the characteristic of the instrument, while 'concerto' has the second low recovery accuracy rate because of its complex music waveform. Though the recovery accuracy rates of these audio pieces are somewhat different, the variance of the values is smaller than 0.004 under each tested echo magnitude [25]. Since the variance of these audio pieces is small, the average value of all the recovery accuracy rate of these pieces is used in the fallowing experiments.

4.2 Parameters Analyses

The control parameters of the echo data hiding techniques influence the percentage of watermark survival very much. Based on the hypotheses made in Chapter 2, the tests with various combinations of control parameters (magnitude, watermark-embedding rate, delay times) are made in series on the simulation system. The watermark data used in the experiments are random generated in each test.

3.2.1 Parameter Ranges

First, the control parameter magnitude (*A*) was examined. The assigned value of A varies from 0.1 to 0.9, with step 0.1. Several fixed value of other parameters are given for the tests. In Fig. 25, the change of recovery accuracy rate generated from the varying magnitudes is shown, where the *Wps* was assigned as fixed values 2, 4 and 8, and the results shown are the average values of several different delay distance 0.3, 0.4 and 0.5msec, while a fixed d_0=0.8msec was given.

Let us recall *Hyporeport 1*, which is about varied magnitudes, at the end of Chapter Two. In Fig. 25, an

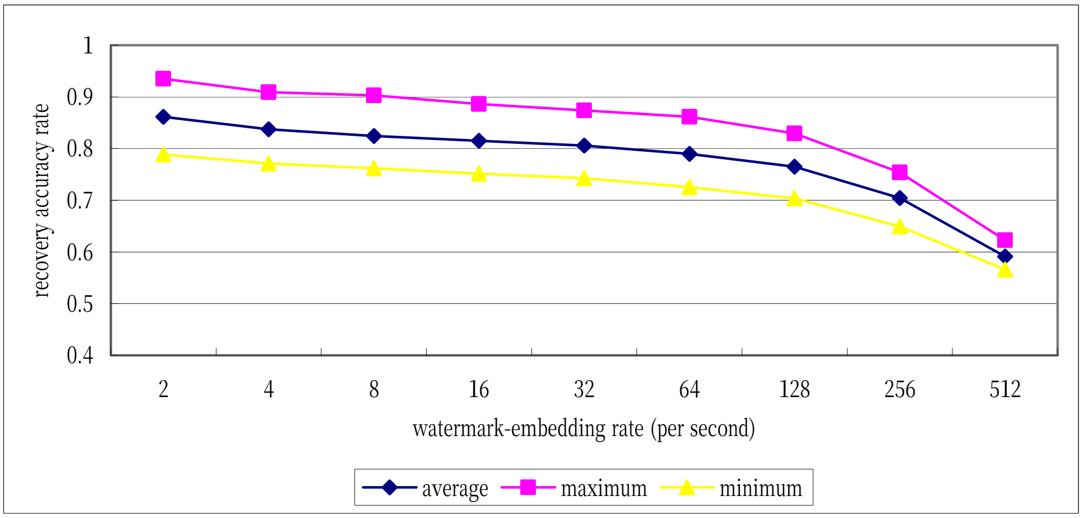


Fig. 26 Relationship between the recovery accuracy rate and varying watermark-embedding rate

A=0.65, 0.75; Wps=2~512, step 2 time; d_0=0.8msec; d_1=1.2msec

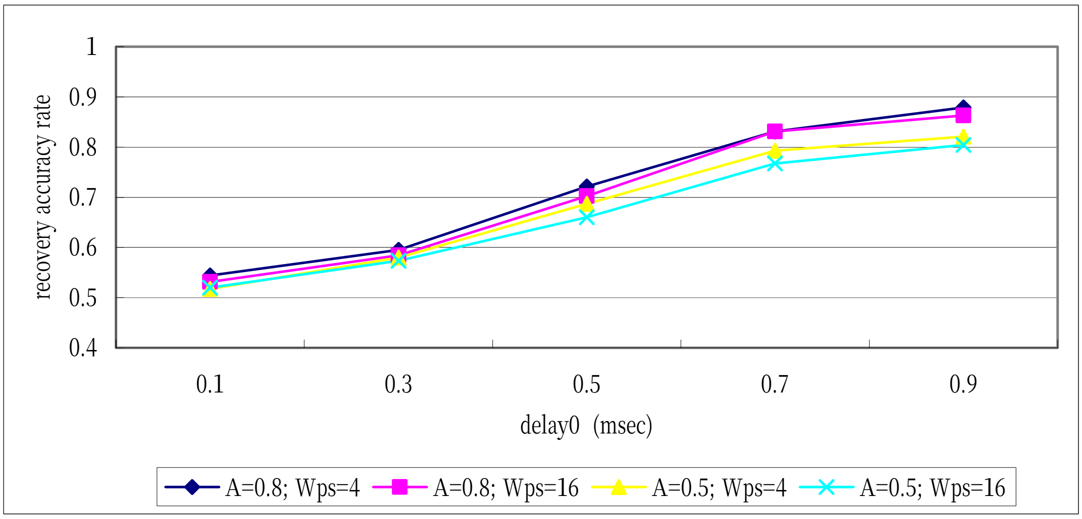


Fig. 27 Relationship between value of d_0 (the smaller delay) and the accuracy rate

A=0.5, 0.8; Wps=4, 16; d_0=0.1~0.9msec, step 0.2msec; delay distance= 0.5msec

overall trend of the experimental results is the recovery accuracy rate decreases as the echo magnitude decreases, which verified what the hyporeport said. The decrease of accuracy rate is not very obvious when the echo magnitude is large (in the range between 0.7 and 0.9). This effect shows there is no need to assign a very large magnitude, so assigning *A* as high as 0.7 or 0.8 is quite enough. Since the acceptable recovery rate is defined as 0.75 (70% correction), the smallest allowable magnitude is approximately 0.3, as shows in Fig. 25.. The echo with even smaller magnitude may not be correctly extracted. From the experimental result, assigning the magnitude a value between 0.3 and 0.8 is recommended in the echo data hiding system.

Second, the watermark-embedding rate (*Wps*) was examined. A sequence of values, 2, 4, 8, … 256, was assigned for the tests of watermark-embedding rate. Several specified values of magnitudes and delay times were assigned. The change of recovery accuracy rate with varying *Wps* is examined with *A*=0.75 and 0.65, and with d_0=0.8msec, d_1=1.2msec. *Hyporeport 2* in Chapter 2 stated that the recovery accuracy rate decreases as the watermark-embedding rate increases. The decreasing tendency of the accuracy rate as the *Wps* increased is clearly shown in Fig. 26. The decrease of the recovery accuracy rate is more obvious when higher *Wps* is given. This effect implies when low *Wps* (in the range between 2 and 32) is assigned, there would not be as large difference between the recovery results. From the test results shown in Fig. 26, the assignment of the watermark-embedding rate is suggested to be in the scope from 2 to 128, which has an accuracy rate better than the predefined acceptable recovery accuracy rate.

Third, the influence from assigning varying delay times is considered. Several different distances of the smaller delay (d_0) from the original were tested. The distance between d_0 and the zero point of time domain ($|d_0–0|$) was given values that varied from 0.1 to 0.8msec.

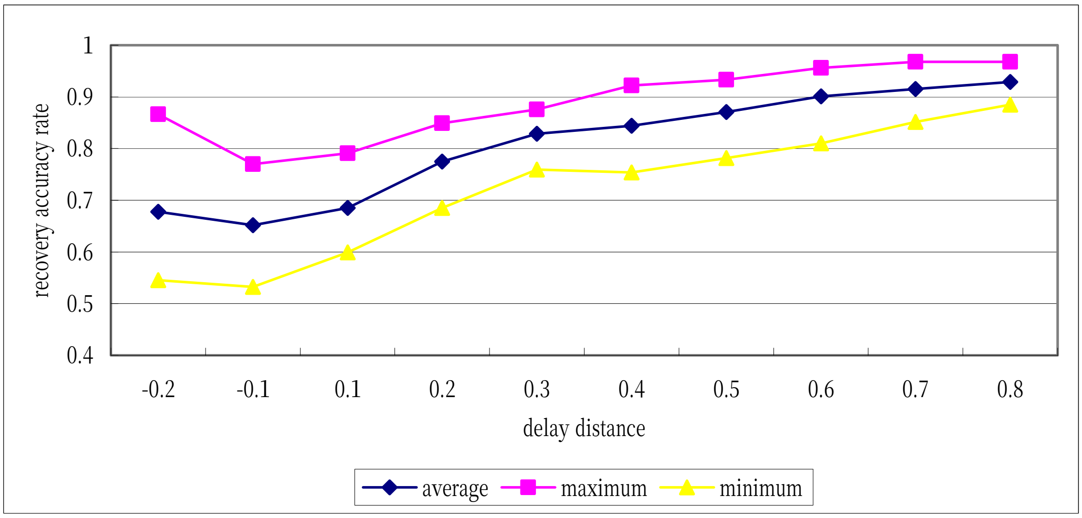


Fig. 28 Relationship between the delay distance and the recovery accuracy rate

A=0.75, 0.6; Wps=2; d_0=0.8msec; d_1=0.6~1.6msec, step 0.1msec, but $d_1 \neq 0.8$msec.

Fig. 27 shows the experimental results of varying delay distances from the original signal. The distance between the delays was fixed at 0.5msec, that is $|d_1-d_0|$=0.5msec. The magnitudes of the echoes were fixed at A=0.8 and 0.5, and the watermark-embedding rates were assigned as Wps=4 and 16 in these experiments. The trend of increasing the recovery accuracy rate as the length between the original and the smaller delay increases can be easily told from Fig. 27. This effect is based on the autocorrelation characteristics of the audio as discussed in prior chapter. Based on the experimental results, to achieve the acceptable recovery rate, the offset of smaller delay needs to be at least 0.6msec from the original.

Besides the influences of distance between the original signal and the echo signal with smaller delay, the influences of distance between the delays also need to be considered. Based on the experimental results in the prior section, value of d_0 was assigned as 0.8msec while d_1 was given values varied from 0.6 to 1.6msec, with step size 0.1msec. The other control parameters are given as fixed values, 0.75 and 0.6 for the magnitude, and the watermark-embedding rate is two in the tests with varying delay distance. While *Hyporeport 4* implies the accuracy rate of recovery would increase as the distance between delays increase, the experimental results match this assumption. As Fig. 28 illustrated, the accuracy rate increases obviously with the increasing delay distance from $|d_1-d_0|$=0.1 to 0.4msec. The delay distance does not make as much difference on the recovery rate when the distance is larger then 0.4msec. To reach the acceptable recovery rate, 0.75, the delay distance should be kept larger than 0.3msec based on the experimental results.

In this section, varied values of the control parameters are examined to find out the best suit range for the control parameters assignment. For echo magnitude, the value range from 0.3 to 0.8 is suggested; and for watermark embedding rate, the scope of 2 to 128 is recommended. The delay values should be at least 0.6msec away from the original while keeping a distance larger than 0.3msec between each other.

4.2.2 Correlation of Parameters

There exist some correlations between the control parameters under the satisfaction of acceptable recovery accuracy rate. Three of the correlations are examined; including the relationship between echo magnitude and watermark embedding rate based on *Hyporeport 5*; the relationship between magnitude and delay distance of echo signals based on *Hyporeport 6*; and the relationship between watermark embedding rate and delay distance based on *Hyporeport 7*. In the experiments, these control parameters are assigned values in the suggested proper range, and the selected parameter combinations have the accuracy rates of approximately 0.8, which is a little larger than the predefined acceptable recovery accuracy rate to avoid the influence of variance between various audio signals.

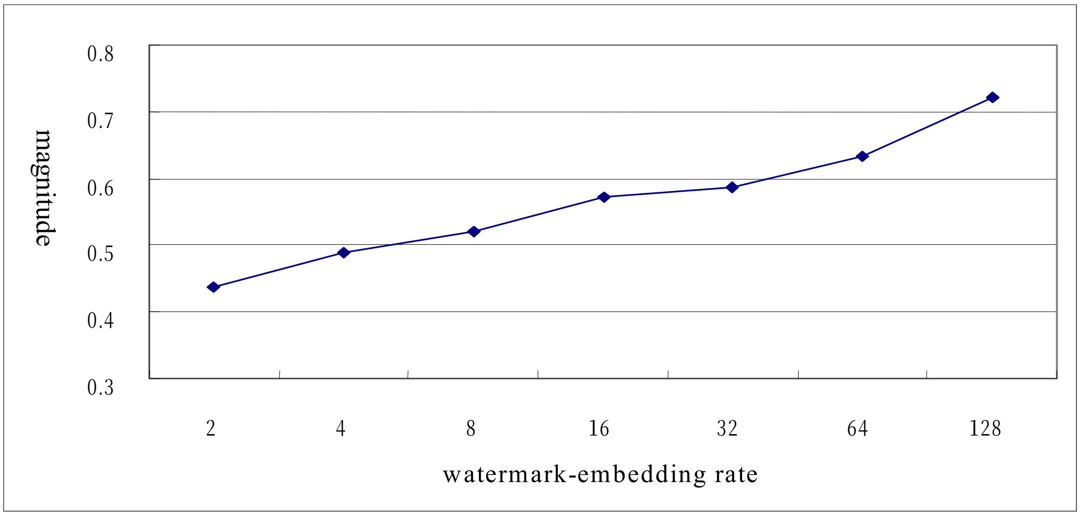


Fig. 29 Tradeoff between magnitude and watermark-embedding rate

A=0.3~0.8, step 0.1; Wps=2~128, step 2 time; d_0=0.8msec, d_1=1.3, 1.5msec, and d_0=1msec, d_1=1.3, 1.5msec.

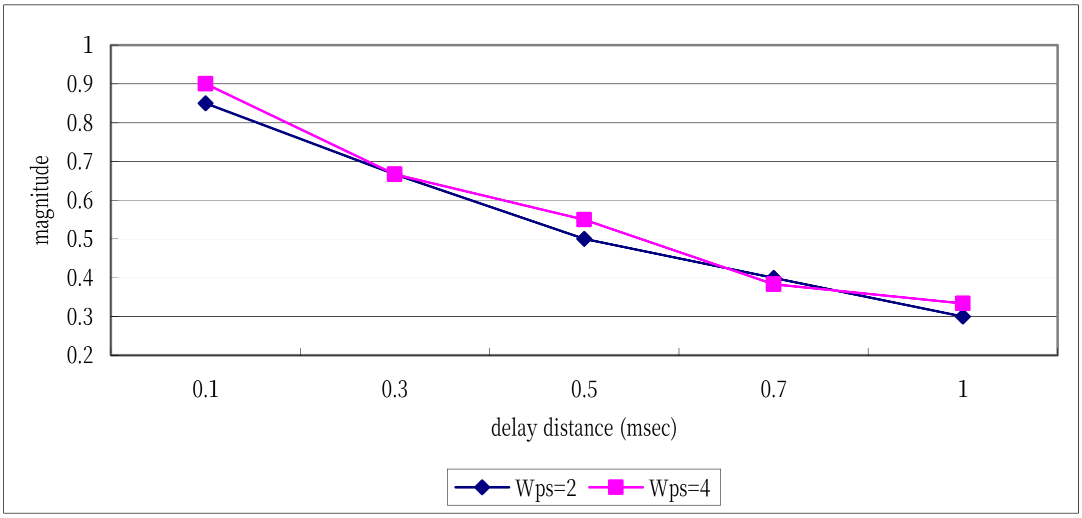


Fig. 30 Tradeoff between the delay distance and the magnitude of echo

A=0.3~0.8, step 0.1; Wps=2, 4; d_0=0.8msec, d_1=0.9, 1.1, 1.3, 1.5, 1.8msec

The relationship of the magnitude and the watermark-embedding rate was considered under the range of $0.3 \leq A \leq 0.8$ and $2 \leq Wps \leq 128$, and the delay distance was given as 0.3, 0.5 and 0.7msec. To achieve the acceptable recovery accuracy rate, the requirement of *A* and *Wps* pairs were determined, as Fig. 29 illustrated. When the *Wps* was given as 128, the magnitude had to be larger then 0.7 to achieve the accuracy rate 0.8; while the magnitude can be as small as 0.4 when the *Wps* is 2. This result verified what *Hyporeport 5* said, when assigning a higher watermark embedding rate, the magnitude of echoes needs to be larger. This experimental result also gives a reference for parameter assignment, where the minimum magnitude for each watermark embedding rate or the maximum watermark-embedding rate for each magnitude is shown.

There also exist some tradeoffs between the delay distance and the echo magnitude, and between the delay distance and the watermark embedding rate under certain performance requirements while the recovery accuracy rate is larger than 0.8. In the experiments, the magnitudes were assigned in the range $0.3 \leq A \leq 0.8$, and the embedding rates were given in the range $2 \leq Wps \leq 128$, while the delay distance varies from 0.1 to 1msec.

The experimental results are shown in Fig. 30 and Fig. 31. The trend of the curves is very clear in the correlation tests of the magnitude and the delay distance, the former can be smaller as the latter is increased while keeping the recovery accuracy rate at the same level, which verifies *Hyporeport 6*. The relationship between the watermark embedding rate and the delay distance of echo signals is also obvious, although the curve is not as linear as the previous experiment. The flat stages of the curve occur because the *Wps* were assigned as a geometric series, where the appearance of not enough precision can be improved by giving values of the *Wps* as an arithmetic

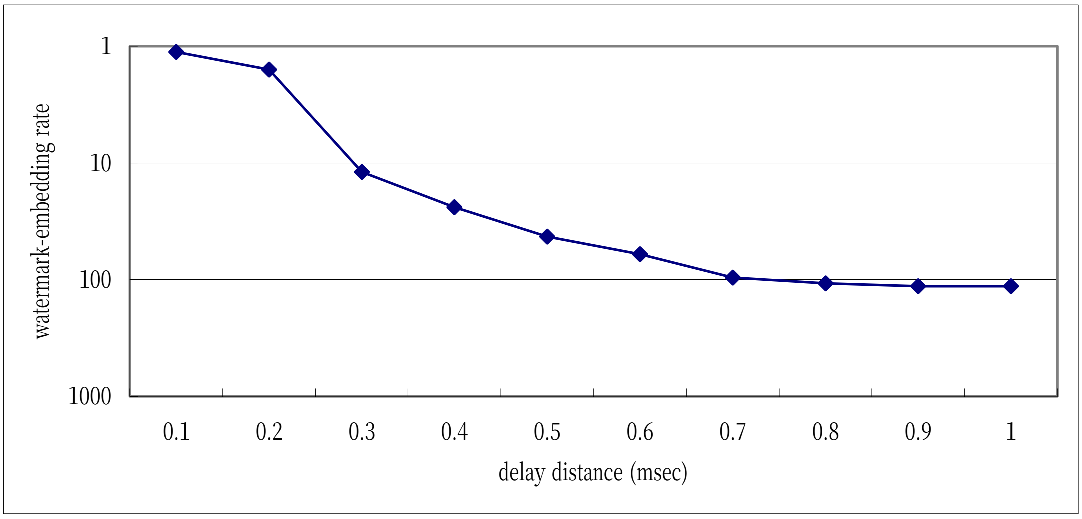


Fig. 31 Tradeoff between the delay distance and watermark-embedding rate

A=0.7; Wps=2~256, step 2 time; d_0=0.8, 1msec, d_1=1.1~1.9msec, step 0.1msec

progression. Based on the experimental results, the watermark-embedding rate can be larger when the delay distance is longer under the same recovery accuracy rate, and *Hyporeport 7* is verified. The results in Fig. 30 and Fig. 31 also give references for the parameter assignments.

These experimental results make the parameter assignment easier. The parameter assignment of a system can be done when considering the recovery accuracy rate, information transmission rate. The audio quality also needs to be considered. An echo magnitude as 0.7 is usually allowed when the largest delay value is smaller than 1.3msec, and a magnitude near 0.5 is usually allowed when the delay value is near 2.5msec.

Although a tradeoff between the recovery accuracy rate and the number of digit for watermark representation must be made, the multiple-bits echo data hiding system does work.

References

[1] Hal Berghel, "Digital Watermarking Makes Its Mark", Mixed Media, September 1998

[2] Minerva M. Yeung, "Digital Watermarking", Communications of the ACM, Vol. 41, No. 7, pp.31-33, July 1998

[3] John M. Acken, "How Watermarking Adds Value to Digital Content", Communications of the ACM, Vol. 41, No. 7, pp.75-77, July 1998

[4] Scott Craver, Boon-Lock Yeo, and Minerva Yeung, "Technical Trials and Legal Tribulations", Communications of the ACM, Vol. 41, No. 7, pp.45-54, July 1998

[5] Nasir Memon and Ping Wah Wong, "Protecting Digital Media Content", Communications of the ACM, Vol. 41, No. 7, pp.35-43, July 1998

[6] Douglas R. Stinson, Cryptography – Theory and Practice, CRC Press, United States of America, 1995

[7] Fabien A. P. Petitcolas, Ross J. Anderson and Markus G. Kuhn, "Information Hiding – A survey", Proceedings of the IEEE, Vol. 87, No.7, pp.1062-1075, IEEE, 1999

[8] Mitchell D. Swanson, Bin Zhu, Ahmed H. Tewfik, Laurence Boney, "Robust Audio Watermarking Using Perceptual Masking", Signal Processing 66 (1998), pp.337-355, Elsevier Science, 1998

[9] Laurence Boney, Ahmed H. Tewfik, Khaled N. Hamdy, "Digital Watermarks for Audio Signals", Proceedings of MULTIMEDIA '96, pp.473-480, IEEE, 1996

[10] F. Bartolini, G. Bini, V. Cappellini, A. Fringuelli, G. Meucci, A. Piva, M. Barni, "Enforcement of Copyright Laws for Multimedia Through Blind, Detectable, Reversible Watermarking", pp.199-203, IEEE, 1999

[11] Ingemar J. Cox, Joe Kilian, Tom Leighton and Talal Shamoon, "A Secure, Robust Watermark for Multimedia", Workshop on Information Hiding, pp.1-16, Newton Institute, Univ. of Cambridge, May 1996

[12] Daniel Gruhl, Anthony Lu, Walter Bender, "Echo Hiding", Information Hiding: First International Workshop, pp.295-315, Cambridge, U.K., May 30 – June 1, 1996

[13] Frank Hartung, Martin Kutter, "Multimedia Watermarking Techniques", Proceedings of the IEEE, Vol. 87, No. 7, pp.1079-1107, IEEE, 1999

[14] W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for data hiding", IBM Systems Journal, Vol. 35, Nos. 3&4, pp.313-336, IBM, 1996

[15] S. Roche and J. L. Dugelay, "Image watermarking based on the fractal transform", Proc. Workshop Multimedia Signal Processing, pp.358-363, Los Angeles, CA, 1998

[16] Chia-Jen Wang and Lieg-Hwei Chen, "A Study on Data Hiding in Audio Signals", The 13th IPPR Conference on Computer Vision, Graphics and Image Processing, pp.14-21, 2000

[17] T. Moriya, Y. Takashima, T. Nakamura and N. Iwakami, "Digital watermarking schemes based on vector quantization", IEEE Workshop on Speech Coding for Telecommunications, pp.95-96, IEEE, 1997

[18] Stefan Katzenbeisser, Fabien A. P. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, United States of America, 2000

[19] C. Neubauer, J. Herre, "Digital Watermarking and its Influence on Audio Quality", AES

[20] Mitchell D. Swanson, Mei Kobayashi, and Ahmed H. Tewfik, "Multimedia Data-Embedding and Watermark Technologies", Proceedings of the IEEE, Vol. 86, No. 6, IEEE, June 1998

[21] Dictionary in Lycos Research Center, "echo", Lycos Home > Reference > Research Center, http://www.lycos.com/reference/

[22] Ted Painter, Andreas Spanias, "Perceptual Coding of Digital Audio", Proceedings of the IEEE, Vol.88, No.4, pp.451-513, IEEE, 2000

[23] E. Ambikairajah, A. G. Davis and W. T. K. Wong, "Auditory Masking and MPEG-1 Audio Compression", Electronics & Communication Engineering Journal, pp.165-175, August 1997

[24] Alan V. Oppenheim, Ronald W. Schafer, Discrete-time Signal Processing, second edition, Prentice Hall Signal Processing Series, United States of America, 1998

[25] Saeed Chahramani, Fundamentals of Probability, Prentice Hall International Editions, United States of America, 1996

[26] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems", Lecture Notes in Computer Science, Vol. 1525, Berlin, Germany, Springer Verlag, pp.218-238, 1998