Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission

Shaoquan Wu, Jiwu Huang, Senior Member, IEEE, Daren Huang, and Yun Q. Shi, Senior Member, IEEE

Abstract—In this paper, we propose a self-synchronization algorithm for audio watermarking to facilitate assured audio data transmission. The synchronization codes are embedded into audio with the informative data, thus the embedded data have the self-synchronization ability. To achieve robustness, we embed the synchronization codes and the hidden informative data into the low frequency coefficients in DWT (discrete wavelet transform) domain. By exploiting the time-frequency localization characteristics of DWT, the computational load in searching synchronization codes has been dramatically reduced, thus resolving the contending requirements between robustness of hidden data and efficiency of synchronization codes searching. The performance of the proposed scheme in terms of SNR (signal to noise ratio) and BER (bit error rate) is analyzed. An estimation formula that connects SNR with embedding strength has been provided to ensure the transparency of embedded data. BER under Gaussian noise corruption has been estimated to evaluate the performance of the proposed scheme. The experimental results are presented to demonstrate that the embedded data are robust against most common signal processing and attacks, such as Gaussian noise corruption, resampling, requantization, cropping, and MP3 compression.

Index Terms—Audio watermarking, audio data transmission, cropping, robustness, self-synchronization, wavelet.

I. INTRODUCTION

INFORMATION hiding and watermarking [1], [2] play an important role in multimedia security. Compared with watermarking technology in images, audio watermarking receives relatively less attention.

According to IFPI (International Federation of the Phonographic Industry) [3], audio watermarking should meet the following requirements. 1) The watermark should not degrade perception of audio. 2) The algorithm should offer more than 20 dB SNR for watermarked audio versus original audio and 20 bps (bits per-second) data payload for watermark. 3) The watermark should be robust against most common audio processing operations and attacks, such as D/A and A/D conversions, temporal scaling (stretching by ±10%), additive and multiplicative noise corruption, MP3 compression. 4) The watermark should be able to prevent unauthorized detection, removal and embedding, unless the quality of audio becomes very poor. These requirements present great challenges to robust audio watermarking.

Most of the recent watermarking algorithms for audio can be grouped into two categories: algorithms in time domain [4], [5] and algorithms in frequency domain [6]–[8]. The main weaknesses of the existing algorithms include low payload and low robustness, especially, vulnerability to shifting and cropping [9], [10].

Since the watermarked audio is likely to suffer from shifting and cropping (such as editing, signal interruption in wireless communications, packet loss in IP network), it is necessary to introduce synchronization into audio data hiding. Note that the synchronization codes have been embedded into the original audio together with the watermark [8], [11] and that the watermark may be a sequence of ECC (Error Correct Code) coded information bits, referred to as informative data. The synchronization codes are exploited to locate the positions where the watermark is embedded. Huang et al. [8] hid the Bark code in time domain as the synchronization codes while embedding the watermark into DCT (discrete cosine transform) coefficients. Because of the limited embedding strength in time domain, the synchronization codes are not robust enough owing to the imperceptibility constrain. On the other hand, if the synchronization codes are embedded in frequency domain, such as DCT and DFT (discrete Fourier transform) domains, the robustness of the synchronization codes will increase but the computational cost in searching synchronization codes will increase greatly at the same time.

This paper presents an audio watermarking algorithm in orthogonal DWT domain. We embed the synchronization codes with the hidden informative data so that the hidden data have the self-synchronization ability. Both the synchronization codes and informative bits are embedded into the low frequency sub-band coefficients in DWT domain to achieve strong robustness against common signal processing procedures, noise corruption, and attacks. By exploiting the time-frequency localization capability of DWT, the proposed technique reduces computational load occurring in searching synchronization codes dramatically and thus resolves the contending requirements between the robustness and low computational complexity. The experimental results demonstrate that the hidden data are robust against signal processing and attacks, such as Gaussian noise corruption, resampling, requantization, cropping and MP3 compression.

The rest of this paper is organized as follows. Section II introduces the outline of the proposed algorithm, followed by detailed descriptions. In Section III, we focus on the performance analysis of the proposed algorithm. Experimental results are shown and discussed in Section IV. Finally, conclusions are drawn in Section V.

Manuscript received November 13, 2003. This work was supported by NSF of China (60325208, 60172067, 60133020); NSF of Guangdong (04205407); Foundation of Ministry of Education, China; and New Jersey Commission of Science and Technology via NJWINS.

S. Wu, J. Huang, and D. Huang are with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510275, P.R. China (e-mail: wu_shao_quan@hotmail.com; issjhw@zsu.edu.cn; hdren@zsu.edu.cn).

Y. Q. Shi is with the Department of ECE, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: shi@njit.edu).

Digital Object Identifier 10.1109/TBC.2004.838265
The main idea of the proposed algorithm is to segment a long audio data sequence into many sections and then to embed one synchronization code and a portion of to-be embedded informative data into the low frequency subband DWT coefficients of each section. The embedding model is shown in Fig. 1.

In data extraction, we first perform the same DWT as in the embedding on a part of incoming audio data and extract binary data from the DWT coefficients of the low frequency subband. We then search for synchronization codes in the extracted data. This procedure needs to be repeated by shifting the selected segment one sample at a time until a synchronization code is found. The detail is described below. With the position of a synchronization code determined, we can then extract the hidden information bits, which follow the synchronization code. The extracting model is shown in Fig. 2.

There are several advantages for applying DWT to audio watermarking. 1) DWT is known to have the time-frequency localization capability. In Section III-C, it is shown that this characteristic can be used to improve computational efficiency greatly in searching synchronization codes. 2) Variable decomposition levels are available. 3) DWT itself needs a lower computation load compared with DCT and DFT. Specifically, suppose there are $N$ samples in an audio section, the computation load are $O(L \cdot N)$, $O(N \cdot \log_2(N))$, and $O(N \cdot \log_2(N))$ for DWT, DCT, and DFT, respectively, where $L$ is the length of the wavelet filter.

A. Synchronization Codes

The algorithm exploits a PN sequence as a synchronization code. The synchronization code is used to locate the position of hidden informative bits, thus resisting the cropping and shifting attacks.

Suppose $\{a_i\}$ is an original synchronization code and $\{b_i\}$ is an unknown sequence both having the same length. If the number of different bits between $\{a_i\}$ and $\{b_i\}$, when compared bit-by-bit, are less than or equal to a predefined threshold, $c$, the $\{b_i\}$ will be determined as the synchronization code. The analysis of error probability in searching synchronization codes is given in Section III-B.

II. THE PROPOSED ALGORITHM

B. Data Embedding

Before embedding, the synchronization codes and watermark should be arranged into a binary data sequence denoted by $\{m_i\}$ as shown in Fig. 3.

We then split the original composite audio data into proper segments and perform DWT on every segment. The sequence $\{m_i\}$ is embedded successively into the low-frequency subband of segments. The length of audio segment depends on the amount of data that need to be embedded and the number of DWT decomposition levels. It should large enough to accommodate at least one synchronization code and some informative bits.

We choose in our algorithm a watermarking technique in the category of quantization index modulation because of its good robustness and blind nature. By blind, it is meant that no original audio is needed in hidden data extraction. Specifically, the rule for embedding $\{m_i\}$ is as follows [12], [13]:

$$c'_i = \begin{cases} \lfloor c_i/S \rfloor \cdot S + 3S/4 & \text{if } m_i = 1 \\ \lfloor c_i/S \rfloor \cdot S + S/4 & \text{if } m_i = 0 \end{cases}$$

(1)

where $\{c_i\}$ and $\{c'_i\}$ are the DWT coefficients of the low-frequency subband of the original audio data and the corresponding watermarked audio data respectively; $\lfloor \cdot \rfloor$ indicates the floor function; and $S$ denotes the embedding strength. The value of $S$ should be as large as possible under the constraint of imperceptibility.

C. Data Extraction

When extracting the hidden data, we can split the test audio into segments (at least one synchronization code should be included in a segment) and then perform DWT on each segment in the same manner as in embedding. Let $\{c''_i\}$ denote the coefficients of low-frequency subband of each segment, we extract the sequence $\{m''_i\}$ from $\{c''_i\}$ by using the following rule [12]:

$$m''_i = \begin{cases} 1 & \text{if } c''_i - \lfloor c''_i/S \rfloor \cdot S \geq S/2 \\ 0 & \text{if } c''_i - \lfloor c''_i/S \rfloor \cdot S < S/2 \end{cases}$$

(2)
Before extracting the informative watermark, we need to search the synchronization codes in the sequence $m^*_i$ bit by bit. With the found synchronization codes, the embedding locations of the watermark are then determined. Note that if no synchronization code is found, then we need to redo the segmenting as indicated in Fig. 2.

The original coefficients $\{c_i\}$ are not required in the extracting process and thus the algorithm is blind. By (1) and (2), we can know that if $c^*_i = c_i + \Delta$ and $\Delta \in (-S/4, S/4)$, then $m^*_i = m_i$.

### III. Performance Analysis

In this section, we evaluate the performance of the proposed algorithm in terms of data payload or embedding data rate, error probability of synchronization code, SNR with embedding strength $S$, and BER of watermark under Gaussian noise. The BER and SNR are defined as

$$\text{BER} = \frac{\text{Number of error bits}}{\text{Number of total bits}} \times 100\% \quad (3)$$

$$\text{SNR} = -10 \log_{10} \left[ \frac{\sum_i (f'_i - f_i)^2}{\sum_i f_i^2} \right] \quad (4)$$

where $\{f_i\}$ and $\{f'_i\}$ denote the original and modified audio, respectively.

#### A. Data Payload

The data payload refers to the number of bits that are embedded into the audio within a unit of time, measured in the unit of bps (bit per second) and denoted by $B$. Suppose that the sampling rate of audio is $R$ (Hz) and the number of wavelet decomposition levels is $K$. Then the data payload $B$ of this algorithm can be shown as

$$B = \frac{R}{2^K} \text{ bps} \quad (5)$$

#### B. False Positive Error and False Negative Error in Synchronization Code Searching

There are two types of errors in searching synchronization codes, false positive error and false negative error. A false positive error occurs when a synchronization code is detected in the location where no synchronization code is embedded, while a false negative error occurs when an existing synchronization code is missed. When a false positive error occurs, the bits after the locations of the false synchronization code will be regarded as the watermark bits. When a false negative error takes place, some watermark bits will be lost.

The false positive error probability $P_1$ and false negative error probability $P_2$ can be calculated as follows:

$$P_1 = \frac{1}{2^p} \cdot \sum_{k=0}^{p} C_p^k \quad (6)$$

$$P_2 = \sum_{k=1}^{p} C_p^k \cdot (\text{BER})^k \cdot (1 - \text{BER})^{p-k} \quad (7)$$

where $p$ is the length of a synchronization code and $e$ is the threshold introduced in Section II-A. Equation (6) indicates that the false positive error probability $P_1$ is independent on watermark attacks.

#### C. Synchronization Codes in DWT Domain

The synchronization codes can be embedded into either time domain or frequency domain. The advantage of embedding in time domain is its low cost in searching synchronization codes. However, at the same time, the robustness is low due to the limitation of embedding strength under imperceptibility constraint. When the synchronization codes are embedded into frequency domain, such as DFT and DCT domain, the robustness will be improved but the searching cost will greatly increase at the same time.

Since the robustness of synchronization codes is important to watermark extraction, we choose to embed synchronization codes into frequency domain to achieve high robustness. To decrease the searching cost, we embed synchronization codes into DWT domain aiming at utilizing the time-frequency localization capability of DWT to improve the efficiency of searching synchronization codes.

Assume that $\{f_i\}$ is the original audio signal. Without loss of generality, we consider the following two audio Sections A and B with the same length of $M$. The Section B can be regarded as a $2^K$ samples shifted version of Section A:

$$A = [f_0, f_1, \ldots, f_{M-1}],$$

$$B = [f_{2^K}, f_{2^K+1}, \ldots, f_{2^K+M-1}] \quad (8)$$

Let $\{c_{i,A}\}$ and $\{c_{i,B}\}$, respectively, denote the low-frequency subband coefficients of Sections A and B after a $K$-level DWT. For example, if $K = 1$, we have

$$c_{i,A} \equiv c_{i,1} = \sum_{j=0}^{L-1} h_j \cdot f_{2i+j} \quad \text{and}$$

$$c_{i,B} \equiv c_{i,1} = \sum_{j=0}^{L-1} h_j \cdot f_{2i+j+2} \quad (9)$$

where $L$ is the length of the wavelet filters and $\{h_i\}$ denotes the low frequency filter of orthogonal wavelet. Except for some boundary coefficients, it is easy to see:

$$c_{i+1,A} = c_{i,B} \quad (10)$$

where $K = 1$. If $K > 1$, according to the definition of $K$-level DWT and mathematical induction, it is easy to prove that (10) is also valid except for some the boundary coefficients as follows. Assume (10) is valid as $K = n$. Now consider $K = n + 1$. Section $B$ is a $2^{n+1}$ samples shifted version of Section $A$. Let us construct a section, $T = [f_{2^n}, f_{2^n+1}, \ldots, f_{2^n+M-1}]$, which is a $2^n$ samples shifted version of Section $A$ and Section $B$ is a $2^n$ samples shifted version of Section $T$. By the assumption made, we have

$$c_{i+2} = c_{i+1}^T = c_i^B$$

According the definition of K-level DWT we have:

$$c_{i+1}^{n+1,A} = \sum_{j=0}^{L-1} h_j \cdot c_{i+2}^{n,A} = \sum_{j=0}^{L-1} h_j \cdot c_{i+1}^{n,A}$$

$$c_{i+1}^{n+1,B} = \sum_{j=0}^{L-1} h_j \cdot c_{i+2}^{n,B} = c_{i+1}^{n+1,B}$$

Therefore we have $c_{i+1}^{n+1,A} = c_{i+1}^{n+1,B}$. We then show that the (10) is valid for any $K$.
It is known that the number of those boundary coefficients that do not satisfy (10) is less than \( L + 2 \). For a proof, readers are referred to Appendix I.

When searching synchronization codes, we can use (10) to save computational cost. Assume that we have calculated the low frequency subband \( \{ c^A_i \} \) of Section A, for the low frequency subband \( \{ c^B_i \} \) of Section B, we only need to calculate at most \( L + 2 \) boundary coefficients by using (10), instead of \( M/2^K \) coefficients.

In other words, if there is one synchronization code in an audio section in DWT domain with \( M \) samples, it just needs at most \( 2^K \) times sample-by-sample searching to find the synchronization code (refer to Fig. 2). As an example, in our experiments reported in Section IV, \( K = 8 \), \( M = 63 \times 256 = 16128 \). Compared with \( M = 16128 \) times searching for DCT/DFT, we only need \( 2^K = 256 \) times searching. It is observed that a great saving in searching synchronization codes has been achieved. Moreover, the computation cost for DWT itself is lower than that for DCT/DFT as mentioned before. Thus, the proposed method in DWT dramatically save the searching cost compared with methods in DFT and DCT domains.

D. Embedding Strength

Based on the observations made on dozens of different types of audio signal, we can assume that the random variable \( \xi = c_\xi - [c_\xi/S] \cdot S \) obeys uniform distribution on \([0, S]\) where \( c_\xi \) is a DWT coefficient of the low frequency subband used to embed data and \( S \) the embedding strength as introduced in Section II-B.

We rewrite (4) as follows for further discussion. The SNR between the original audio \( F \), \( F = \{ f_j \} \), and the watermarked audio \( F^* \), \( F^* = \{ f_j^* \} \), can be expressed as

\[
\text{SNR} = -10 \log_{10} \left( \frac{\|F - F^*\|_2^2}{\|F\|_2^2} \right)
\]  

(11)

Because we use the orthogonal wavelets in our algorithm and the embedding rule in Section III-B does not change the DWT coefficients of the high frequency subbands of \( F \), we have

\[
\text{SNR} = -10 \log_{10} \left( \frac{\|C - C^\prime\|_2^2}{\|C\|_2^2} \right) = -10 \log_{10} \left( \frac{\|C - C^\prime\|_2^2}{\|F\|_2^2} \right)
\]  

(12)

where \( C = \{ c^A_i \} \) and \( C^\prime = \{ c^B_i \} \) are the coefficients of low frequency subband of \( F \) and \( F^* \) after a \( K \)-level DWT.

By embedding formula shown in (1), and the distribution of random variable \( \xi \) assumed in the beginning of this section, it can be shown that half of coefficients in \( \{ c^A_i - c^B_i \} \) obey uniform distribution in \([0, S/4]\) and others obey uniform distribution in \([S/4, 3S/4]\). Furthermore, the different distribution of binary ‘0’ and ‘1’ in the to-be-embedded data sequence \( \{ m_i \} \) does not affect this observation. For a proof, readers are referred to Appendix II. Let \( \mu \) and \( \nu \) denote the random variables that obey uniform distribution in \([0, S/4]\) and \([S/4, 3S/4]\), respectively. We have

\[
E(\mu^2) = D(\mu) + (E(\mu))^2 = \frac{S^2}{48} \quad \text{and}
\]

\[
E(\nu^2) = \frac{13S^2}{48}
\]

(13)

\[
E((c_i - c_i^*)^2) = \frac{1}{2} E((\mu^2) + (\nu^2)) = \frac{7S^2}{48}
\]

(14)

\[
\|C - C^\prime\|_2^2 = \sum_{i=1}^{N/2^K} (c_i - c_i^*)^2 = \frac{N}{2^K} \cdot \frac{7S^2}{48}
\]

(15)

where \( E \) stands for mean, \( D \) stands for variance, and \( N \) is the length of audio \( F \).

Finally, we have the following equation:

\[
\text{SNR} = -10 \log_{10} \left( \frac{N}{2^K} \cdot \frac{7 \cdot S^2}{48 \cdot \|F\|_2^2} \right)
\]  

(16)

The above equation associates the embedding strength to the SNR. With Formula (16) we can decide the embedding strength \( S \) directly according to SNR without actual embedding attempt.

E. BER Against Gaussian Noise Attack

Assume \( \{ f_k^* \} = \{ f_k^1 \} + \{ n_i \} \), where \( \{ f_k^1 \} \) and \( \{ f_k^* \} \) denote the watermarked audio the corrupted watermarked audio, respectively. \( \{ n_i \} \) is Gaussian noise obeying \( N(0, \delta^2) \). Let \( \{ d_k^i \} = \{ c_k^i - c_k^B_i \} \) be the coefficients of low-frequency subband of \( \{ n_i \} \) after \( K \)-level DWT.

According to the probability theory, we can prove that the low frequency subband of Gaussian noise \( \{ n_i \} \) after \( K \)-level orthogonal DWT is still Gaussian noise.

For example, when \( K = 1 \), we can calculate the mean and variance for \( \{ d_k^i \} \) as follows.

\[
E(\overline{d_k^i}) = E \left( \sum_{j=0}^{L-1} h_j \cdot n_{2i+j} \right) = \sum_{j=0}^{L-1} h_j \cdot E(n_{2i+j})
\]

\[
= 0 \cdot \sum_{j=0}^{L-1} h_j = 0 \cdot \sqrt{2} = 0
\]

(17)

\[
D(\overline{d_k^i}) = D \left( \sum_{j=0}^{L-1} h_j \cdot n_{2i+j} \right) = \sum_{j=0}^{L-1} h_j^2 \cdot D(n_{2i+j})
\]

\[
= \delta^2 \cdot \sum_{j=0}^{L-1} h_j^2 = \delta^2 \cdot 1 = \delta^2
\]

(18)

According the definition of \( K \)-level DWT and mathematical induction, the conclusion is also right when \( K > 1 \). Please note that \( \sum_{j=0}^{L-1} h_j^2 = 1 \) is only valid for orthogonal wavelet bases. For wavelets of other types, the whole analysis procedure is similar, though \( \sum_{j=0}^{L-1} h_j^2 \neq 1 \).
By (2), we know that if \( |d_i^k| = |d_i^{k+1} - d_i^{k'}| > S/4 \), the \( m_i \) could not be extracted exactly, so the BER can be expressed as

\[
\text{BER} = \frac{1}{2\pi} \int_0^{\infty} \frac{1}{\sqrt{2\pi\delta}} e^{-\left(x - \sigma^2/2\right)} \, dx
\]

Equation (19) estimates the algorithm performance and indicates that the BER in watermark detection is independent on the origin audio under additive Gaussian noise corruption.

IV. EXPERIMENTAL RESULTS

We tested our algorithm on two 16-bit signed mono audio signals sampled at 44.1 kHz with the length of about 15 seconds in the WAVE format, denoted by \( \text{march.wav} \) and \( \text{light.wav} \) respectively. The two audio signals have rather different proprieties (marching music and light music). We also tested many other kinds of audio (such as human voice, pop music, and etc.), and the simulation results are similar.

Each synchronization code is composed of an m-sequence with a length of 63 and the watermark is a 256-bit binary sequence. The threshold \( \epsilon \) defined in Section II-A is set as 21. Haar wavelet is applied with eight decomposition levels.

With (5), we can estimate the data payload \( \beta \) as 172 bps satisfying the IFPI requirement described in Section I. It needs an audio section about 1.85 seconds to embed a synchronization code and a watermark. The false positive probability \( P_1 \) of synchronization codes is 0.56% according to (6).

Fig. 4 shows that the experimental and theoretical results of SNR between the origin audio and the watermarked audio agree with one another very well. The theoretical values of SNR are calculated by using (16).

Fig. 5 shows that the experimental and theoretical results of BER after Gaussian noise corruption agree with one another very well. The theoretical values of BER are calculated by using (19), and the experimental values are the same as the values listed in Table I.

In other words, there are in total eight curves in four small diagrams of Figs. 4 and 5. Each small diagram contains two overlapped curves (experimental and theoretical). This indicates the correctness of (16) and (19).
In our following experiments, the values of embedding strength $S$ are fixed as 11000 for march.wav and 3500 for light.wav. The SNRs between the original audios and the watermarked audios are 30.64 dB for march.wav and 29.94 dB for light.wav, respectively, both satisfying the IFPI requirement.

Tables II–V show the test results with Gaussian noise corruption, resampling, requantization, and MP3 compression attacks, respectively. The values of false negative error probability $P_2$ are calculated by using (7).

Through these tables, we can see that the watermarks and synchronization codes are very robust to various attacks. For example, the BERs of extracted data are 24.18% and 10.26% under MP3 compression with the lowest bit rate (32 kbps) for 44.1 kHz audio, respectively. Accordingly, the false negative probabilities are 3.36% and $1.37 \times 10^{-5}$%, respectively.

It is straightforward to compare our algorithm with other proposed algorithms due to the differences of audio samples, watermark imperception, data payload and so on. A rather rough comparison is however given in Table V based on embedding data payload, synchronization, and MP3 compression.

We can see that the data payload of our proposed algorithm is much higher than that of other two algorithms. Therefore the performance of our algorithm can be further improved by reducing data payload (increasing wavelet decomposition level) and then increasing embedding strength on the premise of the same imperception constrain.

To test the ability of resisting cropping, we cut a portion of audio data from the watermarked audio, as shown in Fig. 6. There seven vertical lines in Fig. 6 illustrate seven synchronization codes. The watermarks are between the synchronization codes (vertical lines). The cropped audio includes five synchronization codes and four complete pieces of watermarks. Using the extracting algorithm, we can extract these synchronization codes and watermarks.

We also test the performance of the proposed algorithm with different orthogonal wavelet bases, including Daubechies wavelets, Coiflets wavelets, and Symlets wavelets. The observation is that the choice of different wavelet bases has little effect on the performance of the proposed algorithm. Thus, we exploit the simplest wavelet base, Haar wavelet.

V. CONCLUSION

In this paper, we propose a self-synchronized audio data hiding technique based on DWT. The main contributions are as follows.

Watermark is embedded with synchronization codes and thus the self-synchronized watermark has the ability to resist shifting and cropping.

Synchronization codes and watermarks are embedded into low-frequency subband in DWT domain, thus achieving good robustness performance against common signal processing procedure and noise corruption.

The time-frequency localization capability of DWT is exploited to improve the efficiency in searching synchronization codes.

We provide analytical formula to estimate SNR based on embedding strength $S$. We also estimate the BER of watermark after Gaussian noise corruption. The correctness of these formulas is fully demonstrated by the experiment results.
After conducting and comparing several groups of experiments, we find that the embedding strength $S$ is greatly dependent on the type and magnitudes of the original audio signals. It is not the best choice to use a fixed $S$. Our future works include that how to apply the adaptive embedding [17], [18]. Another consideration of future research is to apply error correct coding technique and interleaving to improve the robustness of hidden data [19], [20].

**APPENDIX I**

Without loss of generality, we suppose that the boundary extension only occurs at the right boundary of audio section and the extension mode is period. As shown in Fig. 1, the Section $B$ is a $2^K$-shifted version of Section $A$.

As shown in Fig. 1, when do the 1st level DWT, according Fig. 1 we can see that the different number of boundary coefficients, $D_1$, is $(2^K + L)/2$. Please note that the 2:1 down sampling is performed when calculating the coefficients.

When do the 2nd level DWT, $D_2 = (D_1 + L)/2$.

...When do the $K$th level DWT, $D_K = (D_{K-1} + L)/2$.

So, for $K$-level DWT, $D_K$ can be calculated as follows:

$$D_K = (((2^K + L)/2 + L)/2 + \cdots + L)/2$$

$$= 1 + \sum_{i=1}^{K} L/2^i < L + 2.$$ 

where $D_n$ is the number of different boundary coefficients between $\{c_i^{nA}\}$ and $\{c_i^{nB}\} (n = 1, 2, \ldots, K)$.

If the extension mode is not period (for example, zero padding), the discussion about $D_K$ is also similar.

**APPENDIX II**

Suppose that the ratio of ‘1’ is $p$ ($0 \leq p \leq 1$) in the sequence $\{m_i\}$, so the ratio of ‘0’ is $(1 - p)$.

Then, we have

$$|c_i - c_d| = \begin{cases} c_i - \left[\frac{c_i}{S}\right] \cdot S + \left\{ \begin{array}{ll} 3S/4 & \text{if } (m_i = 1) \\ S/4 & \text{if } (m_i = 0) \end{array} \right. \\ = \xi - \left\{ \begin{array}{ll} 3S/4 & \text{if } (m_i = 1) \\ S/4 & \text{if } (m_i = 0) \end{array} \right. \\ \end{cases}$$

where $\xi, \xi = c_i - \left[\frac{c_i}{S}\right] \cdot S$, obeys the uniform distribution on $[0, S]$ according to the assumption.

i) Suppose that the probability of $m_i = 1$ is $p$ ($0 \leq p \leq 1$), so

$$X = |\xi - 3S/4| \in \begin{cases} [S/4, 3S/4] & \text{if } \xi \in [0, S/2] \\ [S/4, S] & \text{if } \xi \in [S/2, S] \end{cases}$$

Since $\xi$ obeys the uniform distribution on $[0, S]$, the probability that $X$ obeys uniform distribution on $[S/4, 3S/4]$ is to be $(1/2) \ast p$.

Similarly, the probability that $X$ obeys uniform distribution on $[0, S/4]$ is equal to $(1/2) \ast p$, too.

ii) The probability of $m_i = 0$ is $(1 - p)$

$$X = |\xi - S/4| \in \begin{cases} [0, S/4] & \text{if } \xi \in [0, S/2] \\ [S/4, 3S/4] & \text{if } \xi \in [S/2, S] \end{cases}$$
The probability that $X$ obeys uniform distribution on $[0, S/4]$ is $(1/2) \times (1 - p)$.

The probability that $X$ obeys uniform distribution on $[S/4, 3S/4]$ is $(1/2) \times (1 - p)$.

Finally, the total probability that $X$ obeys uniform distribution on $[S/4, 3S/4]$ is equal to $(1/2) \times p + (1/2) \times (1 - p) = 1/2$.

The total probability that $X$ obeys uniform distribution on $[0, S/4]$ is equal to $1/2$, too.

**REFERENCES**


**Jiwu Huang** received his B.S. degree from Xiadian University, China, in 1982, M.S. degree from Tsinghua University, China, in 1987, and Ph.D. from Institute of Automation, Chinese Academy of Science in 1998. He is currently a professor with the School of Information Science and Technology, Sun Yat-Sen University, P.R. China. He serves as a member of IEEE CASS Technical Committee of Multimedia Systems and Applications. His current research interests include multimedia security and data hiding.

**Daren Huang** is a professor with Sun Yat-sen University. After his graduation from the Mathematics Department of Zhejiang University in 1968, he did his graduate studies there from 1978 to 1981. He became a teacher in the same university upon completion of his studies. From 1985 to 1986, he was a visiting scholar in the Mathematics Department of University of South Carolina in U.S.A. He has published more than 100 research papers in his research areas. His current research interests include image processing and data hiding.

**Shaoquan Wu** received the B.E. degree in Application Mathematics in 1999 from Zhejiang University, China, and is currently working toward the Ph.D. degree in the Department of Computer Science, Sun Yat-Sen University, Guangzhou, China. His research interests include digital watermarking, data hiding and wavelet theory.

**Yun Q. Shi** is currently a professor with the Department of ECE at the New Jersey Institute of Technology. He obtained his B.S. degree and M.S. degree from the Shanghai Jiao Tong University, China, and his M.S. and Ph.D. degrees from the University of Pittsburgh, PA. He has more than 180 research papers in his research areas. He is the chairman of Signal Processing Chapter of IEEE North Jersey Section, a member of IEEE CASS Technical Committee of Digital Signal Processing and Communications as well as Technical Committee of Multimedia Systems and Applications, a member of IEEE SPS Technical Committee of Multimedia Signal Processing, an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II.