

# Robust and Efficient Digital Audio Watermarking Using Audio Content Analysis

Chung-Ping Wu, Po-Chyi Su and C.-C. Jay Kuo  
Media Fair, Inc., 1055 Corporate Center Dr., Ste 580  
Monterey Park, CA 91754

and

Department of Electrical Engineering-Systems  
University of Southern California, Los Angeles, CA 90089-2564

E-mail: {chungpin,pochyisu,cckuo}@sipi.usc.edu

## ABSTRACT

Digital audio watermarking embeds inaudible information into digital audio data for the purposes of copyright protection, ownership verification, covert communication, and/or auxiliary data carrying. In this paper, we first describe the desirable characteristics of digital audio watermarks. Previous work on audio watermarking, which has primarily focused on the inaudibility of the embedded watermark and its robustness against attacks such as compression and noise, is then reviewed. In this research, special attention is paid to the synchronization attack caused by casual audio editing or malicious random cropping, which is a low-cost yet effective attack to watermarking algorithms developed before. A digital audio watermarking scheme of low complexity is proposed in this research as an effective way to deter users from misusing or illegally distributing audio data. The proposed scheme is based on audio content analysis using the wavelet filterbank while the watermark is embedded in the Fourier transform domain. A blind watermark detection technique is developed to identify the embedded watermark under various types of attacks.

**Keywords:** digital watermark, blind watermark detection, audio content analysis, synchronization attack, human auditory system, malicious cropping attack, wavelet

## 1. INTRODUCTION

Digital audio watermarking, the embedding and detection of an imperceptible signal in digital audio data, has received increasing attention recently. Among various different uses of digital audio watermarking, copyright protection is the most highly demanded application. The fast growth of the Internet and the maturity of audio compression techniques enable the promising market of on-line music distribution. However, since the digital technology allows lossless data duplication, illegal copying and distribution would be much easier than before. This concern does make musical creators and distributors hesitant to step into this market quickly. Therefore, the proper content protection technology is the key to the emergence of this new market.

Encryption and watermarking are the two most important content protection techniques. Encryption protects the content from anyone without the proper decryption key. It is useful in protecting the audio data from being intercepted during transmission. However, after the intended receiver decrypts it with the correct key, audio data could be illegally distributed and misused. Watermarks, on the other hand, cannot be removed from audio data even by the intended receiver. The embedded watermark signal permanently remains in audio data after repeated reproduction and redistribution. Thus, this signal could be used to protect the copyright of audio content by playback prohibition, illegal copy source tracing and ownership establishment.

Other applications of digital audio watermarking include data hiding for covert communication, auxiliary data embedding for audio content labeling, and modification detection for authentication. Data hiding can also be used to complement encryption, i.e. enhancing communication security by concealing the existence of sensitive data transmission. Embedded auxiliary data can carry lyrics or descriptions of the carrying audio data, or serve as links to external databases. Disappearance of fragile watermark could indicate unauthorized modifications and be used for content integrity verification.

Different watermarking applications have different sets of requirements. Here, our discussion is focused on copyright protection because it has the most stringent requirement on the watermark's ability to survive intentional

attacks. This is considered as one of the most challenging issues of the watermarking technology today. Users benefit from embedded label data while hackers do not know the existence of hidden communication data. Thus, embedded watermarks in these two applications are generally not subject to malicious attacks.

This paper is organized as follows. The requirements for audio watermarking systems are described in Section 2. Previous work on audio watermarking is reviewed in Section 3. Our current work on salient point extraction and Fourier domain watermarking is presented in Section 4. Experimental results and their analysis are given in Section 5. Finally, concluding remarks are provided in Section 6.

## 2. REQUIREMENTS FOR AUDIO WATERMARKING SYSTEMS

In order for the embedded watermark to effectively protect the copyright of the digital audio data, it has been generally agreed<sup>1-9</sup> that a good watermarking scheme should satisfy the following properties:

1. The embedded watermark should not produce audible distortion to the sound quality of the original audio.
2. The computation required by watermark embedding and detection should be low. The complexity of watermark detection should be especially low to facilitate its integration into consumer electronic products.
3. Watermark detection should be done without referencing the original audio data. This property is known as blind detection.
4. The watermark should be undetectable without prior knowledge of the embedded watermark sequence. This property prevents attackers from reversing the embedding process to remove the watermark.
5. The embedded watermark should be robust against common signal processing attacks such as filtering, resampling and compression.
6. The watermark should survive malicious attacks such as random cropping and noise adding. However, severe attacks that produce annoying noise can be ignored for the survival test.

## 3. PREVIOUS WORK ON AUDIO WATERMARKING

A variety of audio watermarking methods with very different characteristics have been proposed. They will be reviewed in this section.

Early work on audio watermark embedding achieved inaudibility by placing watermark signals in perceptually insignificant regions. One popular choice was the higher frequency region,<sup>10-12</sup> where human sensitivity declines compared to its peak around 1 kHz. In some systems,<sup>10,11</sup> the watermark signal is high-pass filtered before being inserted into the original audio. In another system,<sup>12</sup> the Fourier transform magnitude coefficients over the frequency range from 2.4 kHz to 6.4 kHz are replaced with the watermark sequence. In these systems, inaudibility is further enhanced by only embedding watermarks in audio segments whose low frequency components have a higher energy value. The strong low frequency signals in the original audio could help to mask the embedded high frequency watermark signal.

Another human insensitive domain is the Fourier transform phase coefficients. Human ears are relatively insensitive to phase distortions, and especially lack the ability to perceive the absolute phase value. A scheme<sup>1</sup> proposed to substitute the phase of an initial audio segment with a reference phase that represents the watermark. The phase of subsequent segments is adjusted to preserve the relative phase between segments. In another system,<sup>13</sup> selected Fourier transform phase coefficients in higher frequencies are discarded and new values are assigned based on neighboring reference coefficients. The watermark is represented by the relative phase between selected coefficients and their neighbors. The problem with watermarking schemes that hide watermark signals in perceptually insignificant regions is that they are less robust to signal processing and malicious attacks. Compression algorithms do not preserve these regions well so that malicious hackers could implement stronger attacks in these regions without introducing annoying noise.

Another class of algorithms embed watermarks as echo signals of the original audio. The inaudibility of echo hiding is based on the theory that resonance is so common in our environment that human usually do not perceive it as noise. In these algorithms,<sup>2,14</sup> watermark signals are actually delayed and attenuated versions of the original

signal. The watermark sequence is represented by delay amounts which are retrieved by observing autocorrelation peaks in the time domain<sup>14</sup> or in the cepstrum domain.<sup>2</sup>

Recently, some researchers use a concept borrowed from spread spectrum communication and embed the watermark as pseudo-random noise in the time domain. It is guaranteed by spread spectrum theory that the embedded watermark is statistically undetectable by hackers. Since human ears have different sensitivity to additive noise in different frequency bands, all proposed work uses some filter to spectrally shape the pseudo-random (white) noise and achieve inaudibility. A simple band-pass filter was used in one work,<sup>1</sup> and a nonlinear filter was adopted in another.<sup>4</sup> In yet another system,<sup>15</sup> instead of filtering white noise, a scheme was developed to generate the band-limited pseudo-random watermark signal. The inaudibility of the embedded watermark could be further ensured by utilizing the masking effects of the human auditory system. One system<sup>16,5</sup> used MPEG-I Audio Psychoacoustic Model 1 to spectrally shape the watermark signal while another system<sup>17</sup> used the masking model from MPEG-II AAC. Watermark detection is done by calculating the correlation between the watermarked audio signal and the watermark signal. Armed with the spread spectrum communication theory, this type of watermarking usually survives pretty well under distortions and attacks. However, synchronization is difficult to implement, and its computational cost is high.

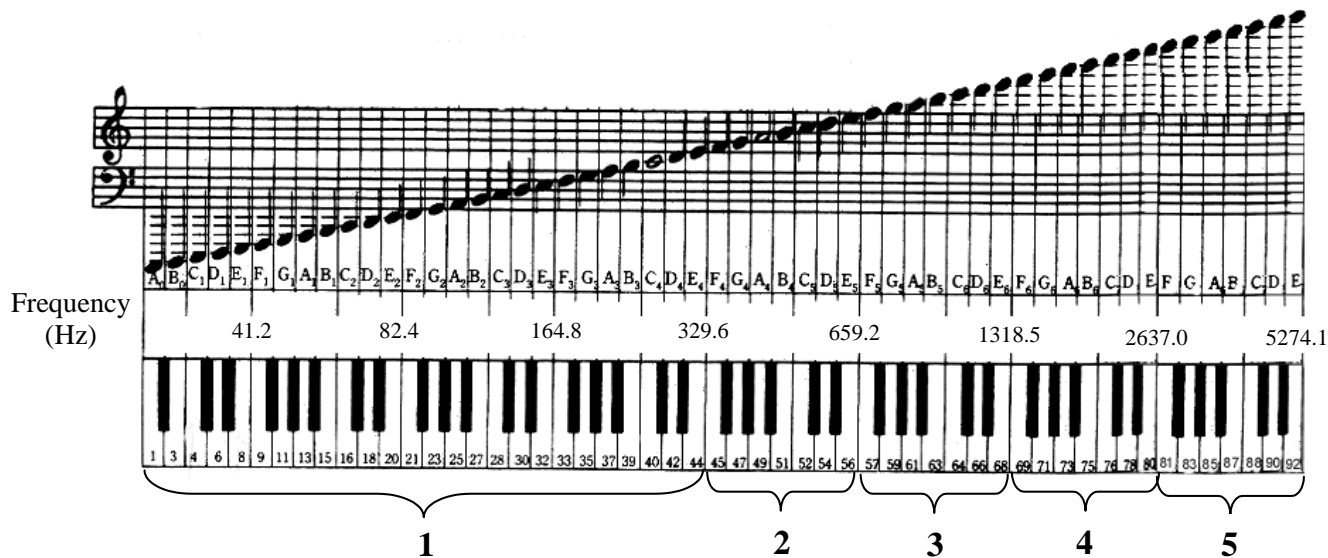
Another trend in digital audio watermarking is to combine watermark embedding with the compression or modulation process. The integration could minimize unfavorable mutual interference between watermarking and compression, especially preventing the watermark from being removed by compression. In one scheme,<sup>18</sup> watermark embedding is performed during vector quantization. The watermark is embedded by changing the selected code vector or changing the distortion weighting factor used in the searching process. The need of the original audio to extract the watermark greatly limits the applications of this scheme. Another algorithm<sup>19</sup> embeds watermark directly in the sigma delta modulation bitstream to eliminate the need of transforming it into PCM data, thereby keeping the computational cost low. This is important to the sigma delta modulation system, where hardware savings is the main goal. In another scheme,<sup>6,20</sup> watermarking is integrated with MPEG-II AAC compression. Watermark is embedded by modifying selected compression coefficients such as the scale factor.

#### 4. PROPOSED ALGORITHM

Although the methods described in section 3 have their own features and properties, they share one common problem. That is, they are vulnerable to the synchronization attack in watermark detection. This problem could be resulted from casual audio editing such as cropping unwanted audio segments or intentional attacks such as randomly deleting or adding samples to watermarked audio data. This *random sample cropping attack* is very effective in interfering with the watermark detection process with respect to the algorithms mentioned above. This attack has a very low computational complexity. Besides, when done correctly, it would not introduce annoying noise to the underlying audio signals. One might argue that such a skillful attack could only be done by a few professionals and not by the majority of consumers. However, once a watermarking method is widely in use, it is almost certain that some professionals would produce and distribute attacking apparatuses so that a majority of common users would be able to perform the skillful attack. One method<sup>5</sup> was proposed to solve the synchronization problem, where an exhaustive search algorithm was used and the original audio signal was required. Consequently, its computational complexity is too high, and the need of original audio for watermark detection greatly limits its applications. Furthermore, it can only handle the casual editing attack, but not the random sample cropping attack.

In this research, we propose a low complexity solution to the synchronization problem caused by both casual and malicious attacks. The solution is composed of a salient point extraction technique and a Fourier transform domain watermark embedding procedure. Salient point extraction through audio content analysis is done during both watermark embedding and detection processes so that synchronization is regained at each salient point. The extraction algorithm is designed such that salient points remain stable after distortion. The Fourier transform domain watermark embedding and detection is adopted since the frequency domain information is less effected by sample cropping in the time domain.

One common characteristic among most existing audio watermarking algorithms is that their watermark is embedded throughout the entire audio signal. However, this may not be the most efficient way to embed and detect watermarks. For a skilled attacker, different amount of attack could be applied to different segments of the audio signal to avoid introducing annoying noise. For example, randomly cropping (deleting) one sample out of every 100 samples in high energy tonal segments of audio signals would produce noticeable noise, but the effect of doing so in



**Figure 1.** Illustration of the correspondence between music notes and frequency values, and the 5-subband partition adopted in this work

low energy segments would be inaudible. Thus, watermarks embedded in highly-attackable areas will face heavier attack and are more likely to be destroyed. The second major contribution of this work is the introduction of “attack-sensitive regions” via audio content analysis. If the watermark is only embedded in attack-sensitive regions where little attack could be applied, the computational complexity of both watermark embedding and detection could be reduced.

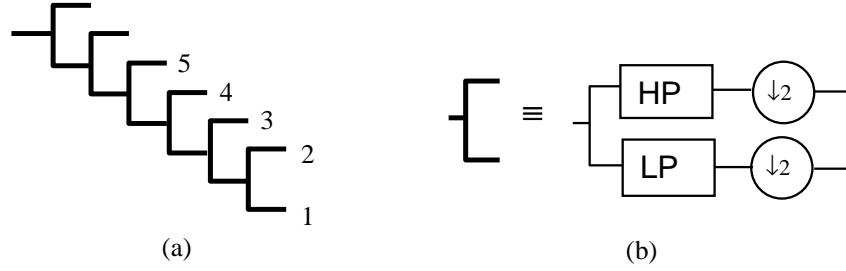
By combining techniques of salient point extraction, attack-sensitive region identification, and Fourier transform domain watermark embedding and detection, we propose a complete audio watermark embedding and detection system for copyright protection. This system satisfies all desired properties of watermark design described earlier. Furthermore, it has a very low computational complexity, and it is robust to casual and intentional synchronization attacks. Although we incorporate the concept of salient point extraction and attack-sensitive regions into our own watermark embedding method here, it is our belief that other watermark embedding algorithms will benefit from the same concepts as well.

#### 4.1. Audio Content Analysis for Watermarking

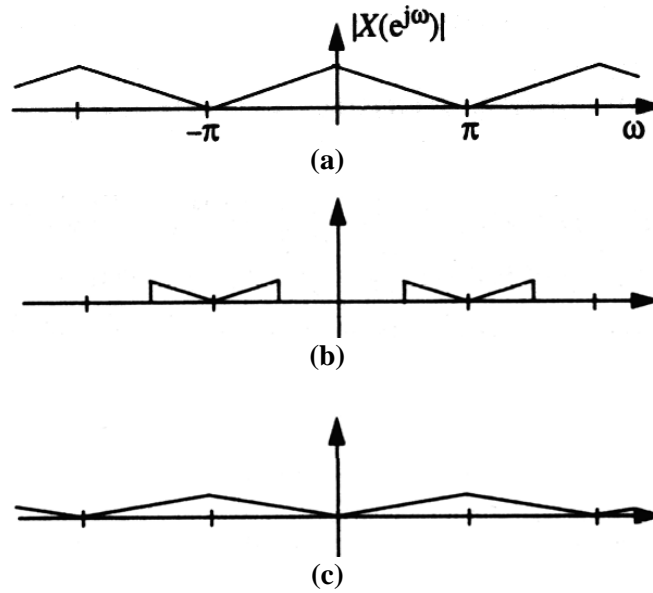
In our system, audio content analysis is performed for the purposes of salient point extraction and attack-sensitive region identification. Salient points in an audio signal allow watermark detection to resynchronize at these locations. Synchronization by salient points has far less complexity than exhaustive search and makes blind watermark detection possible. It should be noted that we do not insert salient points, but extract them from the raw audio via content analysis. This approach has two advantages over explicitly embedding synchronization signals. One is that our content analysis approach does not introduce any distortion to the original audio signal since we do not add anything to it. The other is that the explicitly added synchronization signal is more likely to be taken out by attackers.

A good salient point extraction method should produce approximately the same set of salient points from audio signals before and after attacks such as audio compression, low-pass filtering and noise adding. To achieve this, we extract salient points based on audio features that are sensitive to human ears. In this way, if an attacker wants to destroy these salient points, he/she would have to alter these features and produce noticeable distortions. We choose the energy variation as the main feature for salient point extraction because the associated computational cost is low and alterations in this feature would be audible.

The basic scheme is to extract salient points as locations where the audio signal energy is climbing fast to a peak value. While this approach works well for simple music pieces with few instruments, it has two problems with more



**Figure 2.** A 6-level dyadic wavelet decomposition, where each branch in (a) represents the structure in (b) and outputs are numbered corresponding to subbands in Fig.1

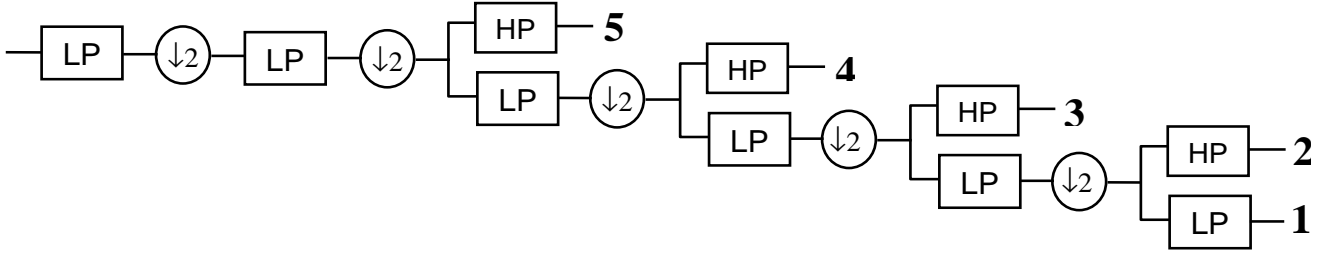


**Figure 3.** The effect of frequency inversion due to downsampling after high-pass filtering: (a) the spectrum before filtering, (b) the spectrum after high-pass filtering, (c) the spectrum after downsampling, where the highest frequency in (a) is now mapped to the lowest frequency.

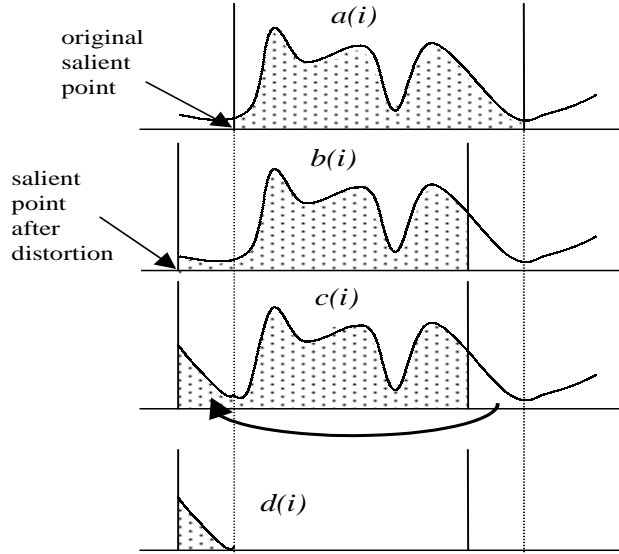
complex music pieces. The first problem is that the overall energy variation becomes ambiguous for complicated music where many instruments are played together. Thus, the stability of salient points decreases. The other problem is that optimal threshold values are different for music pieces with different complexity. While a high threshold value is suitable for music with sharp energy variation, the application of the same value to complex music would yield very few salient points.

Therefore, it is beneficial to parse complex music into several simpler ones so that stability of salient points could be improved and the same threshold could be applied to all music pieces. Complex music is usually composed of instruments whose fundamental frequencies occupy different frequency ranges in order to form harmony. Figure 1 illustrates the correspondence between music notes and frequency values. It also shows the partition in our design, which consists of 5 frequency ranges. Note that the frequency width of each octave is not the same. The frequency intervals in Figure 1 correspond to outputs of a 6-level dyadic wavelet decomposition under a sampling rate of 44.1kHz as shown in Figure 2.

In order to prevent the frequency inversion effect<sup>21</sup> due to the application of downsampling to the output of high-pass filtering as shown in Figure 3, we modify the dyadic wavelet decomposition of Figure 2 into Figure 4 by



**Figure 4.** The modified dyadic wavelet decomposition.



**Figure 5.** The effect of salient point displacement on the discrete Fourier transform domain watermarking.

eliminating the downsampling step after each high-pass filtering. Thus, salient points are extracted separately from each of the 5 outputs in Figure 4.

The procedure of attack-sensitive region identification aims at decreasing the watermark embedding and detection complexity. Thus, it is important that the identification process itself does not require too much computation. In this work, we integrate attack-sensitive region identification process into the salient point extraction process so that almost no extra computation is needed for attack-sensitive region identification. The attack that we are mainly concerned with is the random sample cropping attack. The corresponding attack-sensitive regions is the high energy tonal region. Since salient points chosen with our algorithm are located at positions where the audio signal energy is fast climbing to a peak, the region following each salient point would contain high energy. We simply define this region as the attack-sensitive region, so that no additional computation is needed.

#### 4.2. Fourier Transform Domain Watermark Embedding and Detection

Although salient points are selected to be as stable as possible, it is difficult to get exactly the same salient points after some audio processing such as compression. A certain amount of displacement in the location of salient points is common and should be tolerated. If we embed and detect watermark in the time domain, it is obvious that even a small amount of displacement would have a problem since embedding and detection cannot be synchronized. However, this problem is alleviated by considering the magnitude coefficients of the discrete Fourier transform.

This property is illustrated in Figure 5, where  $a(i)$ ,  $i = 1, \dots, 2^p$ , is the watermarked region. The watermark is embedded in  $|A(k)|$ ,  $k = 1, \dots, 2^p$ , where  $A(k)$  is the discrete Fourier transform coefficient of  $a(i)$ . Suppose that the salient point is displaced in the detection process, and the watermarked region is mistaken to be another region  $b(i)$ ,

$1 \leq i \leq 2^p$ . However, it is a well known property that if  $c(i)$  is formed by moving the right-most part of  $a(i)$  to the left-most part, then  $c(i)$  and  $a(i)$  have identical discrete Fourier transform magnitude coefficients, i.e.

$$|C(k)| = |A(k)|, \quad k = 1, \dots, 2^p, \quad (1)$$

Let us denote the difference between  $b(i)$  and  $c(i)$  with

$$d(i) = c(i) - b(i), \quad i = 1, \dots, 2^p. \quad (2)$$

Then, we have

$$\begin{aligned} |B(k)| &\approx |C(k)| + |D(k)| \\ &= |A(k)| + |D(k)|, \quad k = 1, \dots, 2^p \end{aligned} \quad (3)$$

Thus, from (3), we see that the error caused by the displaced salient point is  $|D(k)|$ . There is no disastrous mis-synchronization effect in the frequency domain. When the displacement amount is small relative to the window size, the energy in  $|D(k)|$  is small.

In order for the embedded watermark to be inaudible, it is common to utilize the temporal and frequency masking effects of the human auditory system (HAS).<sup>1,4,5</sup> Temporal masking refers to the effect that weaker signals immediately before and after a stronger signal may be inaudible while frequency masking refers to the effect that when two signals occur simultaneously and are close together in the frequency, the stronger signal may make the weaker one inaudible.

Since our watermark is only embedded in attack-sensitive regions, which have a high energy value, the temporal masking effect is used. That is, the weak-energy watermark is masked by the high energy audio samples in these regions. To take advantage of the frequency masking effect as well, the proposed scheme only embeds the watermark signal in the magnitude of the discrete Fourier transform coefficients that have large values.

The watermark detection is done by calculating the average correlation coefficient between the watermark sequence and the watermarked audio signal in the Fourier transform domain and comparing it with a threshold. The criterion for selecting the threshold is to minimize the expected cost of detection errors. Note that the cost of miss (i.e. failure to detect when there is a watermark) is different from the cost of false alarm (i.e. claim a detection while there is no watermark). Although these costs vary in different applications, it is generally true that the cost of false alarm is much greater than the cost of miss. The false alarm rate should be extremely low because it undermines the credibility of the watermarking method to prove copyright ownership. In contrast, the constraint on the miss (or failure-to-detect) rate need not be so stringent, since the failure-to-detect rate of 1% or 10% might have a similar effect in scaring people away in illegally copying audio data. To conclude, the detection threshold should be set relatively high to ensure no false detection happens.

## 5. EXPERIMENTAL RESULTS

The inaudible and robust properties of the proposed watermarking scheme are demonstrated with three pieces of audio signals: Piano concerto by Bach with only a single piano, symphony "Bolero" by Ravel with trumpet and drums, and a song with human vocal and complex background music. All signals are sampled at a frequency of 44.1 kHz, and each piece is about 30 seconds long.

### 5.1. Audio content analysis

The effectiveness of the proposed audio content analysis is measured by its ability to extract the same set of salient points from audio signals before and after signal attack and/or processing. An example of the comparison between the salient points extracted from the original and processed files is shown in Table 1. As we can see from this example, almost every salient point is more or less shifted by a few points. However, as explained in Section 4.2, this does not cause a catastrophic effect on watermark detection. Empirically, a displacement of less than 100 points produces very little decrease to the average correlation coefficient in watermark detection. Therefore, it should be viewed as successful salient point extraction. Some salient points may disappear and some may be created after processing. However, again these phenomena only cause a marginal deterioration to detection results.

The success rates of correctly extracted salient points with and without the wavelet filterbank are compared in Table 2. The attack used in Table 2 is MP3 compression/decompression. In our experiments, distortions such as

salient point location extracted from original file	salient point location extracted from distorted file	salient point shift amount between two files	salient point location extracted from original file	salient point location extracted from distorted file	salient point shift amount between two files	salient point location extracted from original file	salient point location extracted from distorted file	salient point shift amount between two files
<b>4401</b>	<b>4557</b>	<b>-156</b>	<b>138454</b>	<b>138342</b>	<b>112</b>	<b>343182</b>	<b>343309</b>	<b>-127</b>
6581	6581	0	144478	144489	-11	<b>347048</b>	<b>347233</b>	<b>-185</b>
<b>14196</b>	<b>none</b>		145827	145823	4	351030	351003	27
<b>14463</b>	<b>none</b>		153485	153484	1	359383	359351	32
19464	19471	-7	185107	185056	49	382173	382186	-13
21092	21063	29	<b>192565</b>	<b>192297</b>	<b>268</b>	384255	384259	-4
28657	28651	6	216786	216784	2	389912	389914	-2
44152	44104	48	224510	224555	-45	391882	391884	-2
59635	59637	-2	232790	232808	-18	397653	397654	-1
91080	91126	-46	242895	242878	17	<b>399407</b>	<b>399526</b>	<b>-119</b>
94883	94879	4	264519	264518	1	<b>406960</b>	<b>none</b>	
98548	98545	3	271803	271803	0	422233	422234	-1
<b>none</b>	<b>105946</b>		273508	273507	1	426680	426682	-2
112475	112471	4	297097	297097	0	<b>429936</b>	<b>none</b>	
127941	127958	-17	304761	304760	1	437456	437473	-17
129319	129315	4	320039	320013	26	444820	444795	25
131028	131025	3	335700	335700	0	460640	460643	-3

**Table 1.** Comparison between salient points extracted from original and processed audio files, where rows printed in the bold type are regarded as failures, and the success rate in this example is 78.5%.

Test audio	Success rate without wavelet filterbank	Success rate using wavelet filterbank	Success rate increase
single piano	83.3%	83.6%	0.3%
drum and trumpet music	71.4%	77.3%	5.9%
vocal with complex background music	63.0%	73.1%	10.1%

**Table 2.** The success rate of correct salient point extraction after the cascade of three MPEG Layer III compression/decompression operations with a bit rate of 64 kbps

additive noise, low-pass filtering and downsampling cause much less salient point displacement than MP3 compression/decompression. It is observed that the more complex the music piece, the lower the success rate. However, the use of wavelet decomposition is most effective in raising the success rate of salient point extraction from complex music.

## 5.2. Watermark Embedding

The quality of the proposed watermarking method is evaluated by using the blind listening test. Listeners are presented with the original and watermarked audio without the knowledge of which one is watermarked. They are asked to tell which one has better sound quality. We do not use the question “whether any differences could be detected between the two audio signals”<sup>5</sup> since people tend to imagine the difference while they actually cannot hear any. In fact, several listeners reported that audio signals were different when the same piece of audio clip was played twice.

Eleven people took the listening test, and the percentage of preferring the original audio to the watermarked audio is given in Table 3. The result shows that about one half of listeners preferred watermarked audio to the original. Therefore, no audible distortion is introduced by the embedded watermark.

Test Audio	Original preferred to watermarked
single piano	45.5%
drum and trumpet music	54.5%
vocal with complex background music	45.5%

**Table 3.** The blind listening test of watermarked audio pieces.

ATTACK	Single piano music	drum and trumpet	Vocal with complex background music
No attack	2.63	2.56	2.17
Additive noise	2.44	2.11	1.70
MPEG compression	2.14	1.98	1.51
Random cropping	2.25	2.08	1.76
Low pass filtering	2.07	1.92	1.71

**Table 4.** The ratio between the correlation peak with the correct user ID and the largest correlation in 1000 random trials.

### 5.3. Blind Watermark Detection

We tested the robustness of the proposed blind watermark retrieval algorithm against several kinds of attacks, including additive noise, MPEG compression, random cropping, low pass filtering, and resampling. The quality of watermark detection is evaluated by the ratio between the correlation value obtained from the correct user ID and the largest correlation obtained from 1000 other random user IDs. The ratio between the correlation value from the correct user ID and the largest correlation obtained from 1000 other random user IDs are summarized for the three test audio pieces in Table 4. Each kind of attack leads to a different amount of decrease in this peak ratio. However, in all cases experimented, the correlation peak of the correct user ID always stands out of the rest correlation peaks.

We have the following observations.

- Additive white noise.  
White noise with 10% of the power of the audio signal is added. Noise of this level is clearly audible, but only causes a moderate decrease in the peak ratio.
- MPEG compression.  
In multimedia applications, lossy compression is a very common procedure to increase transmission and storage efficiency. Some information is thrown away during the compression process, thus creating a potential hazard for watermark detection. To test the robustness of the proposed watermarking approach to lossy compression, the watermarked audio signal is compressed and decompressed by MPEG layer III coder with a bit rate of 64 kbps. As shown in Table 4, this attack is more serious than others. However, the watermark can still be detected correctly.
- Random cropping.  
Randomly cropping one sample out of every 100 samples produces a disastrous synchronization problem for time-domain watermarking methods. However, the correlation peak ratio is only slightly decreased with the proposed method.
- Low pass filtering.  
With watermarks embedded in the frequency domain, low pass filtering with a very low cutoff frequency could effectively eliminate the embedded watermark. However, since our watermark is embedded in the frequency bands with the highest energy, filtering out the inserted watermark also greatly effects the sound quality. In

our experiment, a low pass filter with a cutoff frequency of 4kHz is applied to watermarked audio signals. The loss of high frequency components is clearly audible, but the correlation peak ratio is only decreased around 25%.

As shown in Table 4, the correlation peak ratios after various kinds of attacks are scattered between 1.5 ~ 2.5. These values could be increased if the watermark is embedded and retrieved everywhere in the audio signal, or if the original audio is used in watermark detection. However, the correlation ratio in Table 4 is already high enough for unambiguous watermark detection. The efficiency achieved by blind watermark detection and embedding in attack-sensitive regions only is very important for the practical use of audio watermarks.

## 6. CONCLUSION

The rapid growth of multimedia technologies facilitates the production and transmission of digital media data. It brings us not only opportunities but also challenges to copyright protection. An audio watermarking scheme which meets both the robustness and the low computational complexity requirements via audio content analysis was presented in this paper. The analysis identifies attack-sensitive regions which are suitable for watermark insertion, and provides consistent audio segmentation results before and after attacks. A modified dyadic wavelet filterbank is used to enhance the analysis results for complex music. After audio content analysis, a watermark embedding scheme was developed in the Fourier transform domain that utilizes the temporal and frequency masking effects of the human auditory system. The embedded watermark is inaudible. The impact of malicious random cropping attack on audio watermarking was considered. The proposed watermarking solution, which combines the synchronizing feature of audio content analysis and the time-shift tolerance of Fourier domain watermarking, provides a low complexity solution to this kind of attack. There is work to be done in the near future. For example, we would like to perform more experiments to derive optimal parameters and improve watermark survivability after MP3 compression.

## 7. ACKNOWLEDGEMENTS

This work was completely performed in Media Fair, Inc. when the first two authors were hired as student interns by the company. The support of Media Fair, Inc. of this research is highly appreciated.

## REFERENCES

1. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3-4, pp. 313-336, 1996.
2. D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in *Info Hiding 96*, pp. 295-315, 1996.
3. I. J. Cox, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, 1997.
4. I. Pitas and P. Bassia, "Robust audio watermarking in the time domain," in *EUSIPCO'98*, pp. 25-28, 1998.
5. M. Swanson, B. Zhu, A. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing Journal*, vol. 66, pp. 337-355, 1998.
6. J. Lacy, S. Quackenbush, A. Reibman, and J. Snyder, "Intellectual property protection systems and digital watermarking," *Journal of Optics Express*, vol. 3, pp. 478-484, 12 1998.
7. A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "Dct-based watermark recovering without resorting to the uncorrupted original image," *ICIP*, vol. 1, pp. 520-523, 1997.
8. I. J. Cox and M. L. Miller, "A review of watermarking and the importance of perceptual modeling," in *Proceeding of Electronic Imaging Conference*, pp. 92-99, 1997.
9. M. Swanson, B. Zhu, and A. Tewfik, "Current state of the art, challenges and future directions for audio watermarking," in *IEEE International Conference on Multimedia Computing and Systems*, vol. 1, pp. 19-24, 1999.
10. L. Holt, B. G. Maufe, and A. Wiener, "Encoded marking of a recording signal," *UK Patent GB2196167*, 1988.
11. J. Wolosewicz, "Apparatus and method for encoding and decoding information in audio signals," *US Patent 5,774,452*, 1998.
12. J. F. Tilki and A. A. Beex, "Encoding a hidden digital signature onto an audio signal using psychoacoustic masking," in *The 7th International Conference on Signal Processing Applications and Technology*, pp. 476-480, 1996.

13. J. F. Tilki and A. A. Beex, "Encoding a hidden auxiliary channel onto a digital audio signal using psychoacoustic masking," in *IEEE Southeastcon 97*, pp. 331–333, 1997.
14. R. Petrovic, J. M. Winograd, K. Jemili, and E. Metois, "Apparatus and method for encoding and decoding information in analog signals," *US Patent 5,940,135*, 1999.
15. M. Ikeda, K. Takeda, and F. Itakura, "Audio data hiding by use of band-limited random sequences," in *ISASSP*, vol. 4, pp. 2315–2318, 1999.
16. L. Boney, A. H. Tewfik, and K. N. Hamdy, "Digital watermarks for audio signals," in *IEEE International Conference on Multimedia Computing and Systems*, pp. 473–480, 1996.
17. C. Neubauer, J. Herre, and K. Brandenburg, "Continuous steganographic data transmission using uncompressed audio," in *Info Hiding 98*, pp. 208–217, 1998.
18. T. Moriya, Y. Takashima, T. Nakamura, and N. Iwakami, "Digital watermarking schemes based on vector quantization," in *IEEE Workshop on Speech Coding For Telecommunications*, pp. 95–96, 1997.
19. A. J. Magrath and M. B. Sandler, "Encoding hidden data channels in sigma-delta bitstreams," in *ISCAS '98*, vol. 1, pp. 385–388, 1998.
20. J. Lacy, S. Quackenbush, A. Reibman, D. Shur, and J. Snyder, "On combining watermarking with perceptual coding," in *ICASSP*, vol. 6, pp. 3725–3728, 1998.
21. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Prentice Hall, 1995.