DIGITAL AUDIO WATERMARKING IN THE CEPSTRUM DOMAIN

Sang-Kwang Lee and Yo-Sung Ho
Kwangju Institute of Science and Technology
1 Oryong-dong, Puk-gu, Kwangju, 506-712, Korea

Abstract

In this paper, we propose a digital audio watermarking technique in the cepstrum domain. We insert a digital watermark into the cepstral components of the audio signal using a technique analogous to spread spectrum communications, hiding a narrow band signal in a wideband channel. In our method, we use pseudo-random sequences to watermark the audio signal. The watermark is then weighted in the cepstrum domain according to the distribution of cepstral coefficients and the frequency masking characteristics of human auditory system. Watermark embedding minimizes the audibility of the watermark signal. The embedded watermark is robust to multiple watermarks, MPEG audio coding and additive noise.

1. Introduction

In recent days, there has been significant interest in watermarking. This is primarily motivated by a need for providing copyright protection to digital contents, such as audio, image and video. Digital representation of copyrighted materials offers various advantages; however, the fact that an unlimited number of perfect copies can be illegally produced is a serious threat to the right of content owners. Watermarking is used for owner identification, royalty payments, and authentication by determining whether the data has been altered in any manner from its original form.

A watermark must be embedded in the data in such a way that it is imperceptible by the user. Moreover, the watermark should be inaudible or statistical invisible to prevent unauthorized detection and removal. The watermark should also have similar compression characteristics as the original signal and be robust to any manipulations or signal processing operations on the host data, e.g., filtering, compression, resampling, requantization, cropping, noise, A/D-D/A conversions, etc. The watermark should also be embedded directly in the data, not in the header, and be self-clocking for ease of detection in the presence of cropping and time-scale change operations. The watermark should be characteristic of the author, but a private should not be able to detect the watermark by comparing several signals belonging to the same author. The signal should be degraded when the watermark is removed through any unauthorized means.

Compared to image and video signals, audio signals are represented by much less samples per time interval. This alone indicates that the amount of information that can be embedded robustly and inaudibly is much lower than for visual signal. An additional problem in audio watermarking is that the human audible system (HAS) is much more sensitive than the human visual system (HVS), and that inaudibility is much more difficult to achieve than invisibility for images [1].

Boney *et al.* [2] proposed a spread-spectrum approach for audio watermarking. They used a pseudo-random sequence that is filtered in several stages in order to exploit long-term and short-term masking effects of HAS. Bassia and Pitas [3] applied a very straightforward time-domain spread spectrum watermarking technique to audio signals. They reported robustness against audio compression, filtering and resampling.

In this paper, we investigate a spread spectrum technique to insert a watermark into the cepstral component of the audio signal, taking into account the characteristics of the human audio perception. Our watermark is a perceptually inaudible modification of the audio signal, based on the distribution of cepstral coefficients and the cepstral masking of HAS.

Section 2 of this paper introduces basics of watermarking. Section 3 discusses watermark embedding. Section 4 covers watermark detection, and our conclusions are given in Section 5.

Original manuscript received June 19, 2000

2. Basics of Watermarking

The basic idea of watermarking is to add a watermark signal into the host data to be watermarked such that the watermark signal is unobtrusive and secure in the signal mixture, but can partly or fully be recovered from the signal mixture later on if the correct cryptographically secure key is used.

In order to ensure imperceptibility of the modification caused by watermark embedding, we use a perceptibility criterion of some sort. This can be implicit or explicit, fixed or adaptive to host data. As a consequence of the required imperceptibility, the individual samples (e.g., pixels or transform coefficients) that are involved in watermark embedding can only be modified by an amount relatively small to their average amplitude.

In order to ensure robustness despite of the small allowed changes, we usually distribute the watermark information redundantly over many samples (e.g., pixels) of the host data. Therefore, the recovery is more robust if more watermarked data are available for recovery.

In general, watermark systems use one or more cryptographically secure keys to ensure security against manipulation and erasure of the watermark.

There are three main issues for the design of a watermarking system.

ISSUE 1: Design of a watermark signal W to be added to the host signal. Typically, the watermark signal depends on a key κ and watermark information I into which it is embedded

$$W = f_0(I, \kappa) \tag{1}$$

ISSUE 2: Design of an embedding method itself that incorporates the watermark signal W into the host data X yielding watermarked data Y.

$$Y = f_1(X, W) \tag{2}$$

ISSUE 3: Design of a corresponding extraction method that recovers the watermark information from the signal mixture using the key and with help of the original data

$$\hat{I} = g(X, Y, \kappa) \tag{3}$$

or without the original data

$$\hat{I} = g(Y, \kappa) \tag{4}$$

The first two issues, watermark signal design and watermark signal embedding, are often regarded as one, especially when the embedded watermark is adaptive to the host signal.

A generic watermarking scheme for the embedding process is shown in Figure 1.

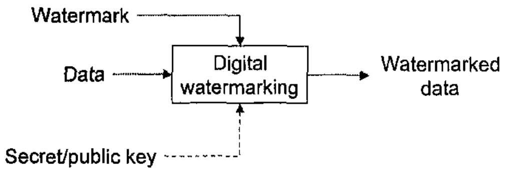

Figure 1. Digital Watermarking Scheme

The input to the scheme is the watermark, the host data, and an optional public or secret key. Depending on applications, the host data may be uncompressed or compressed; however, most proposed methods work on uncompressed data. The watermark can be of any nature, such as a number, a text, or an image. The secret or public key is used to enforce security. If the watermark is not to be read by unauthorized parties, a key can be used to protect the watermark. In combination with a secret or public key, watermarking techniques are usually referred to as secret and public watermarking techniques, respectively. The output of the watermarking scheme are the modified, i.e., watermarked, data.

A generic process for watermark recovery is depicted in Figure 2.

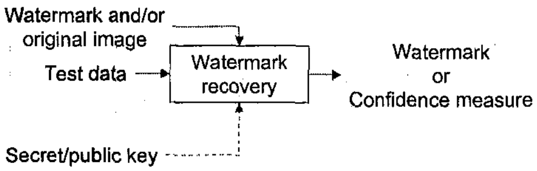
Watermark and/or original image
Test data
Watermark recovery
Watermark or Confidence measure
Secret/public key

Figure 2. Watermark Recovery Scheme

Inputs to the scheme are the watermarked data, the secret or public key, and the original data or the original watermark depending on the method. The output of the watermark recovery process is either the recovered watermark or some kind of confidence measure indicating how likely it is for the given watermark at the input to be present in the data under inspection.

Many watermarking schemes employ ideas borrowed from spread spectrum communications. They embed a watermark by adding pseudo-random sequences of low amplitude to the host data. The specific pseudo-random sequences can be detected using a correlation receiver or a matched filter. If the added pseudo-random sequence is chosen appropriately, the probability of false alarm is very low.

3. Watermark Embedding

3.1 Spread Spectrum Watermarking

Frequency-domain watermarking was first introduced by Boland *et al.* [4] and Cox *et al.* [5], who independently developed perceptually adaptive methods based on modulation. Cox *et al.* drew parallels between their technique and spread-spectrum communication since the watermark is spread over a set of visually important frequency components.

The watermark consists of a sequence of N samples $w(n)$ with a given statistical distribution, such as a normal distribution $N(0,1)$ with zero mean and a variance of one. The watermark is inserted into the signal $x(n)$ to produce the watermarked signal $y(n)$. Cox *et al.* proposed three techniques for watermark insertion

$$y(n) = x(n) + \alpha w(n) \qquad (5)$$

$$y(n) = x(n)(1 + \alpha w(n)) \qquad (6)$$

$$y(n) = x(n)e^{\alpha w(n)} \qquad (7)$$

where α determines the watermark strength, and $x(n)$ is a perceptually significant spectral component. Equation (5) is only suitable if the values of $x(n)$ do not vary too much. Equation (6) and Equation (7) give similar results for small values of $\alpha w(n)$. For positive $x(n)$, Equation (7) may even be viewed as an application of Equation (5) where the logarithm of the original value is used. In most cases, Equation (6) is employed. The scheme can be generalized by introducing multiple scaling parameters $\alpha(n)$ so as to adapt to different spectral components and thus to reduce perceptual artifacts.

3.2 Complex Cepstrum

Consider a sequence $x(n)$ whose Fourier transfrom is $X(\omega)$. As shown in Figure 3, the complex cepstrum $c_x(n)$ of the sequence $x(n)$ is defined as the inverse Fourier transform of $C_x(\omega)$, where

$$C_x(\omega) = \ln X(\omega) = \sum\nolimits_n c_x(n)e^{-j\omega n} \qquad (8)$$

i.e., $c_x(n)$ is the sequence obtained by the inverse Fourier transform of $\ln X(\omega)$ [6, 7].

The principal advantage of cepstral coefficients is that they are generally decorrelated. Besides, high-order cepstra are numerically quite small and have a wide range of variances when going from low to high-order cepstral coefficients [8]. Figure 3 shows a typical distribution of the complex cepstrum.

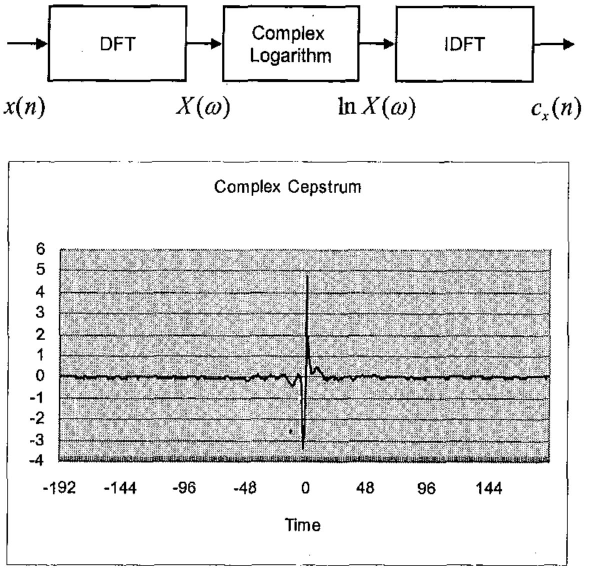

Figure 3. Complex Cepstrum

3.3 Frequency Masking

We can employ a psychoacoustic model to analyze the audio signal and to compute the amount of noise masking available as a function of frequency. The masking ability of a given signal component depends on its frequency position and its loudness. The encoder uses this information to decide how good the input audio signal with its limited number of code bits is.

The MPEG audio standard [9] provides two implementations of the psychoacoustic model. Psychoacoustic model 1 is less complex than psychoacoustic model 2, and has more compromises to simplify the computation. Either model works well for any coding layer.

Below is a general outline of the basic steps involved in the psychoacoustic model.

STEP 1: Align audio data in time. There is one psychoacoustic evaluation per frame. The audio data sent to the psychoacoustic model must be concurrent with the audio data to be coded. The psychoacoustic model must account for both the delay of the audio data through the filter bank and a data offset so that the relevant data is centered within the psychoacoustic analysis window.

STEP 2: Convert audio data to a frequency-domain representation. The psychoacoustic model should use a separate, independent, time-to-frequency mapping instead of the polyphase filter bank because it needs a fine frequency resolution to calculate the masking thresholds accurately. Both psychoacoustic models use a Fourier transform for this mapping. The standard Hanning weighting, applied to the audio data before the Fourier transform, conditions the data to reduce the edge effect of the transform window.

STEP 3: Process spectral values in grouping related to critical bandwidths. Both models process frequency values in perceptual quanta to simplify the psychoacoustic calculation.

STEP 4: Separate spectral values into tonal and non-tonal components. Both models identify and separate tonal and noise-like components of the audio signal because the masking ability for the two types of the signal is different.

STEP 5: Apply a spreading function. The masking ability of a given signal spreads across its surrounding critical band. The model determines the noise masking threshold by applying an empirically determined masking (model 1) or spreading function (model 2) to the signal component.

STEP 6: Set a lower bound for threshold values. Both models include an empirically determined absolute masking threshold and the threshold in quiet. This threshold is the lower bound on the audibility of sound.

STEP 7: Find the masking threshold for each subband. Both psychoacoustic models calculate masking thresholds of higher frequency resolution than that provided by the polyphase filter bank. Both models must derive a subband threshold value from a multitude of masking thresholds computed for frequencies within that subband.

STEP 8: Calculate the signal-to-mask ratio (SMR). The psychoacoustic model computes the SMR as the ratio of the signal energy within the subband to the minimum masking threshold for that subband. The model passes this value to the bit allocation section of the encoder.

3.4 Watermark Embedding

In our watermark embedding scheme, the watermark is spread over several cpestral components so that the energy in any component is very small and certainly undetectable. Spreading the watermark throughout the audio cepstrum ensures high security against unintentional or intentional attack. The watermarked signal does not produce any perceptual distortion.

Let us assume an audio signal of N samples $x(n)$ and a pseudo-random sequence $w(n)$. The watermarked sample $y(n)$ is represented as

$$y(n) = f(x(n), w(n)) \tag{9}$$

where $f(\cdot)$ is a function to embed a watermark. In order to embed a watermark in the complex cepstrum domain, we modify Equation (9) as

$$c_y(n) = c_x(n) + \alpha w(n) \tag{10}$$

where $c_y(n)$ is the complex cepstrum of $y(n)$. We can view multiple scaling parameters α as a relative measure of how much one must alter $w(n)$ to change the perceptual quality of $x(n)$. A large value of α means that one can perceptually "get away" with altering $x(n)$ by a large factor without degrading the document.

3.5 Multiple Scaling Parameters

A single scaling parameter α may not be applicable for perturbing all values of $x(n)$, since different spectral components may exhibit more or less tolerance to modification. Therefore, we can use multiple scaling parameters $\alpha(n)$.

In order to define multiple scaling parameters, we exploit the distribution of cepstral coefficients and the frequency masking characteristics, described in Section 3.2 and Section 3.3. When the difference between adjacent cepstral coefficients is small, the watermarked signal can be noticeably different from the original signal. In order to obtain the first scaling parameter, we exploit the wide range of variances when going from low to high-order cepstral coefficients. We define the first scaling parameter as

$$\begin{aligned} \alpha_1(n) &= c, \quad \text{for } \alpha(n) - \alpha(n-1) > \gamma \\ &= 0, \quad \text{for } \alpha(n) - \alpha(n-1) \le \gamma \end{aligned} \tag{11}$$

where γ is a threshold value.

In order to take advantage of the frequency masking characteristics of HAS, we compute the cepstral signal-to-mask ratio of a signal from the corresponding cepstra. We define the second scaling parameter as

$$\alpha_2(n) = 1 + \text{Cepstal Signal-to-Mask Ratio} \tag{12}$$

Figure 4 shows the cepstral signal-to-mask ratio.

Finally, the multiple scaling parameters are defined as multiplication of two scaling components

$$\alpha(n) = \alpha_1(n)\alpha_2(n) \tag{13}$$

and, the watermarked signal in the complex cepstrum domain is

$$c_y(n) = c_x(n) + \alpha(n)\, w(n) \tag{14}$$

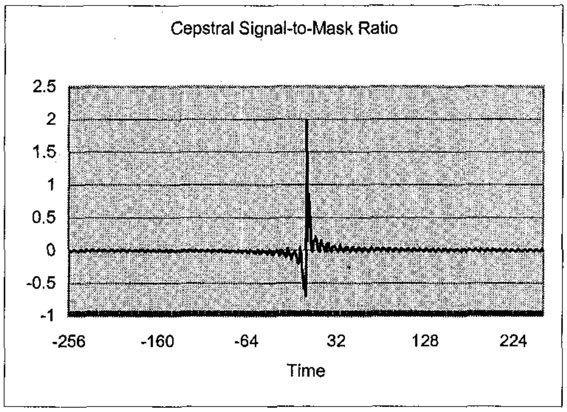

Figure 4. Cepstral Signal-to-Mask Ratio

Figure 5 shows the watermark insertion process. ⊕ corresponds to the embedding algorithm and ⊗ to the weighting of the watermark by the information of the human auditory system.

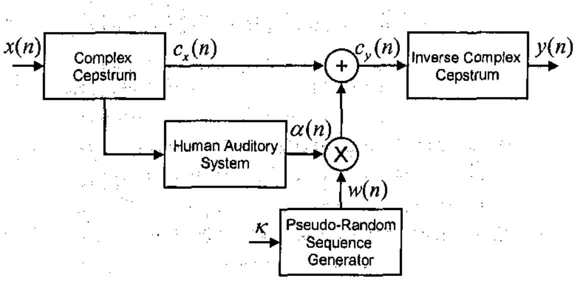

Figure 5. Watermark Embedding Algorithm

4. Watermark Detection

To verify the presence of the watermark, we measure the cross-correlation between the recovered watermark $w^*(n)$ and the original watermark $w(n)$. The cross-correlation is defined as

$$\begin{aligned} r(m) &= \sum_{n=1}^{N} w^*(n)w(n+m) \\ &= \sum_{n=1}^{N} \alpha(n)w(n)w(n+m),\ 0 \le m \le N-1 \end{aligned} \tag{15}$$

where

$$w^*(n) = c_y(n) - c_x(n) = \alpha(n)w(n) \tag{16}$$

Ideally, a pseudo-random sequence should have an autocorrelation function that has correlation properties similar to the white noise [10]. We define the watermark detector response as

$$d = \max|r(m)|,\ 0 \le m \le N-1 \tag{17}$$

4.1 Uniqueness of Watermark

Figure 6 shows the response of the watermark detector to 1000 randomly generated watermarks. This figure demonstrates that only one match of the watermark is present in the watermarked signal. The watermark detector response due to the correct watermark is very strong related to other responses to incorrect watermarks, suggesting that the algorithm has a very low rate of false alarm.

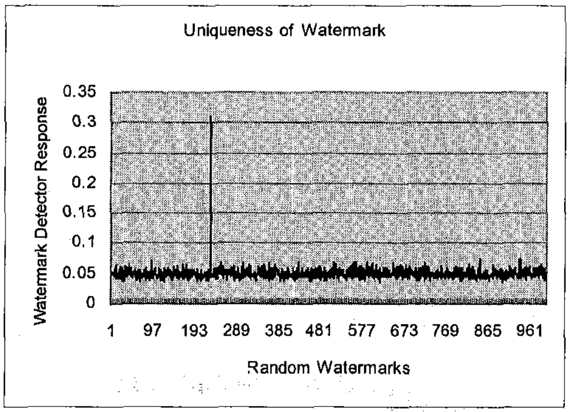

Figure 6. Uniqueness of Watermark

4.2 Robustness to Multiple Watermarks

Figure 7 shows a watermark detector response after four successive watermarking operations, i.e. the original signal is watermarked, the watermarked signal is watermarked again, etc. This may be considered another form of the attack and we can expect that significant signal degradation eventually occurs as the process is repeated.

On the other hand, there are many instances where it is useful to add multiple watermarks to a signal. For example, there may be multiple authors for a piece of music, each with his/her own unique identification. When we detect a specific watermark, the other watermarks are considered to be noise.

Figure 7 shows the response of the detector to 1000 randomly generated watermarks, including four watermarks present in the audio signal. Four spikes clearly indicate the presence of the four watermarks. This result demonstrates that successive watermarking does not interfere with the watermark detection process.

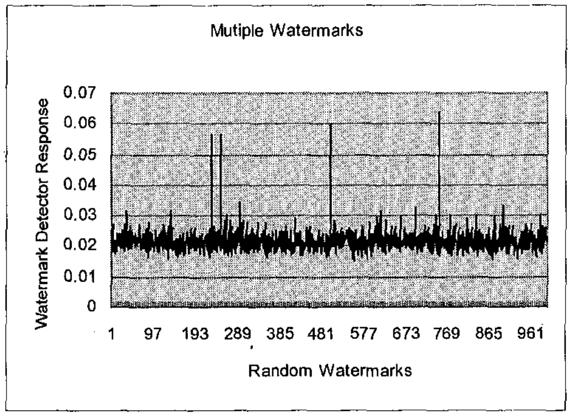

Figure 7. Detection in Multiple Watermarks

4.3 Robustness to MPEG-1 Audio Coding

Robustness of the watermark technique described above, has been tested using Layer III coding of the MPEG-1 audio standard (MP3). Several signed 16-bit stereo 44,100 Hz watermarked signals were encoded at several different bitrates. The watermark survives through the encoding-decoding process, as shown in Figure 8. Watermarking detection after decompression indicates a slight increase of the watermark detector response.

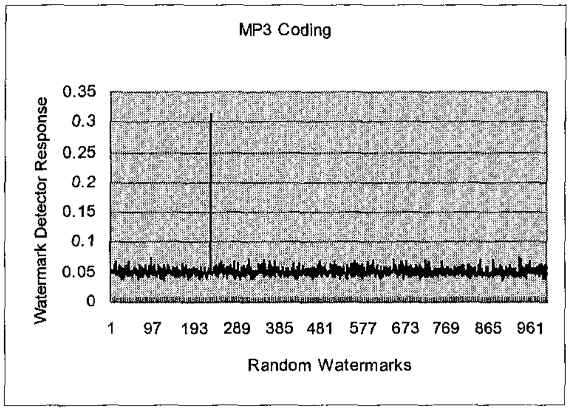

Figure 8. Detection in MP3 Coding

4.4 Robustness in the Presence of Additive Noise

Robustness of the proposed watermark algorithm was studied under additive noise conditions. Figure 9 shows the response of the watermark detector for the watermarked signal containing additive noise. Figure 9 indicates a slight decrease of the watermark detector response.

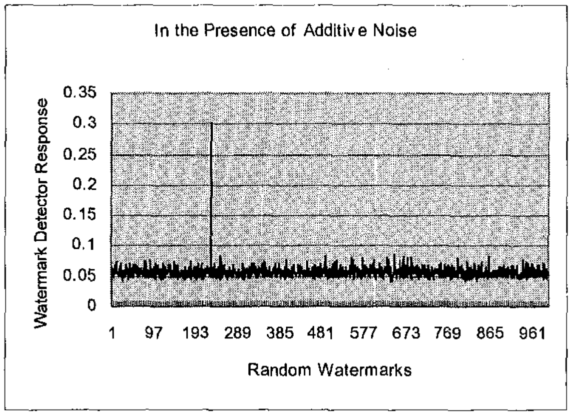

Figure 9. Detection in Additive Noise

5. Conclusions

In this paper, we propose a new algorithm for digital audio watermarking in the cepstrum domain. We insert a watermark into the cepstral components of the audio signal using a spread spectrum technique. Our watermarking scheme exploits the distribution of the cepstral coefficients and the cepstral signal-to-masking ratio. Our watermark is imperceptibly embedded into the audio signal and easy to detect by the author due to the correlation properties of pseudo-random sequences. Our results show that our watermarking scheme is robust to multiple watermarks, lossy coding/decoding and additive noise.

Acknowledgment

This work was supported in part by the Korea Science and Engineering Foundation (KOSEF) through the Ultra-Fast Fiber-Optic Networks (UFON) Research Center at Kwangju Institute of Science and Technology (K-JIST), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. of IEEE*, vol. 87, no. 7, July 1999.

[2] L. Boney, A. Twefik and K. Hamdy, "Digital watermarks for audio signals," *Europ. Signal Processing Conf.*, Trieste, Italy, Sept. 1996.

[3] P. Bassia and I. Pitas, "Robust audio watermarking in the time domain," *Europ. Signal Processing Conf.*, Rhodes, Greece, Sept. 1998.

[4] F. Boland, J. Ruanaidh and W. Dowling, "Watermarking digital image for copyright protection," *Int. Conf. Image Processing and Its Applications*, vol. 410, Edinburgh, U. K., July 1995.

[5] I. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video," *IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sept. 1996.

[6] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.

[7] J. Proakis and D. Manolakis, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1996

[8] S. Young, D. Kershaw, J. Odell, D, Ollason, V. Valtchev and P. Woodland, *The HTK Book*, Entropy Ltd, 1999.

[9] D. Pan, "A tutorial on MPEG/audio compression," *IEEE Multimedia Journal*, Summer 1995.

[10] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, Reading, Mass., 1994.

Biographies

Sang-Kwang Lee received the B.S. degree in avionics from Hankuk Aviation University, Korea, in 1996, and the M.S. degree in information and communications engineering from Kwanju Institute Science and Technology (K-JIST), Kwangju, Korea, in 1998. He is currently working toward the Ph.D. degree in Information and Communications Department of K-JIST. His research interests include digital video and audio coding, digital watermarking, and multimedia communication systems.

Yo-Sung Ho received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, Korea, in 1981 and 1983, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990. He joined ETRI (Electronics and Telecommunications Research Institute), Daejon, Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, New York, where he was involved in development of the Advanced Digital High-Definition Television (AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korean DBS digital television and high-definition television systems. Since 1995, he has been with Kwangju Institute of Science and Technology (K-JIST), where he is currently Associate Professor of Information and Communications Department. His research interests include digital video and audio coding, data recovery and restoration, advanced coding techniques, and content-based signal representation and processing, multimedia communication systems.