Audio Watermarking through Modification of Tonal Maskers

Hee Suk Lee and Woo Sun Lee

Watermarking has become a technology of choice for a broad range of multimedia copyright protection applications. This paper proposes an audio watermarking scheme that uses the modified tonal masker as an embedding carrier for imperceptible and robust audio watermarking. The method of embedding is to select one of the tonal maskers using a secret key, and to then modify the frequency signals that consist of the tonal masker without changing the sound pressure level. The modified tonal masker can be found using the same secret key without the original sound, and the embedded information can be extracted. The results show that the frequency signals are stable enough to keep embedded watermarks against various common signal processing types, while at the same time the proposed scheme has a robust performance.

Keywords: Audio watermarking, psychoacoustic model, information hiding.

Manuscript received Feb. 23, 2005; revised June 3, 2005.

This research was financially supported by Changwon National University in 2003.

Hee Suk Lee (phone: +82 55 279 7420, email: lhs@changwon.ac.kr) and Woo Sun Lee (email: wslee@changwon.ac.kr) are with Department of Computer Engineering, Changwon National University, Changwon, Korea.

I. Introduction

A number of different techniques for watermarking have been researched and studied, but often they have been found to suffer a rather high degree of loss in the application of various common signal processing types. In particular, signal processing for audio contents tends to alter a great deal of the audio signal values, requiring a more robust watermarking technique to overcome this problem.

Watermarking techniques have recently been proposed that make use of human auditory characteristics and embed a watermark that is larger in size and yet is imperceptible. Some examples are echo hiding [1], phase coding [1], or spread spectrum watermarking that uses a psychoacoustic model [2]-[4]. Echo hiding has proven to be robust against a variety of signal processing types including compression, noise insertion, or band-pass filtering, while hiding watermarks that are not unpleasant to the ear. Nevertheless, using this technique, watermarks can be detected and deleted through echo cancellation by an unauthorized person. Phase coding is effective in terms of sound quality, while the phases are susceptible to signal processing like MPEG coding [5]. Spread spectrum (SS) watermarking embeds spread spectrum signals in the shape of the masking threshold of the cover audio [2]-[4]. It can perform imperceptible embedding; however, it requires precise synchronization before correlation to extract the watermark. Moreover, SS watermarking techniques that extract watermarks without the original audio usually show a significant watermark loss when the watermarked audio is applied to band-pass filtering, MP3 compression, or re-sampling [4].

This paper proposes a new audio watermarking scheme that uses the characteristics of a tonal masker to provide a more robust method of watermarking while at the same time reducing sound distortion due to the watermarking process.

A tonal component may be defined as a pure tone. For the purposes of this study, however, it is defined more specifically as described in the literature of the MPEG psychoacoustic model [6], [7]. According to this model, a tone is no more than two local maxima that represent each Bark in the frequency signal energies between 11.6 and 23.2 ms. A tone functions as a tonal masker together with the adjacent frequency signals and also masks neighboring signals that have relatively low energy. These tonal maskers are important factors in identifying the masking threshold in MPEG coding as well as critical elements that offer clues for the perceptual differentiation of an audio signal. In the case of embedding a watermark in parts that are not critical perceptually, this watermark can be more susceptible to various signal processing types and therefore can be easily lost. For a more robust watermarking, the watermark needs to be embedded in those parts that are perceptually important [8], and the characteristics of tonal maskers are highly suitable for this purpose.

In section II, on the basis of previous psychoacoustic research and the course calculating a tonal masker in the MPEG model, we explain that it is possible to imperceptibly modify a tonal masker. We also provide analyses showing that this modification can maintain its robustness even after various types of signal processing. Through an explanation and analysis, we confirm that the modified tonal maskers function as a suitable watermark carrier. Section III includes a more detailed proposal for watermarking through the modification of tonal maskers, and section IV shows the experimental results comparing the proposed method to an existing spread spectrum watermarking method using linear prediction filtering [4]. Section V concludes this paper.

II. Characteristics of Tonal Masker as Watermark Carrier

This chapter explains that the modified tonal maskers can function as embedding carriers for imperceptible and robust audio watermarking.

Tonal maskers in the MPEG psychoacoustic model are calculated using (1) through (4), which is done in order to find the sound pressure level (SPL) of a tonal masker, where n is the number of samples in a frame, $w(n)$ is a window function, and $s(n)$ is the n-th audio sample of an audio frame. A spectral component, $F(k)$, is obtained using a Fourier transform, and $P(k)$ is the SPL of $F(k)$ [6], [7].

$$F(k) = \sum_{n=0}^{N-1} w(n)s(n)e^{-j\frac{2\pi kn}{N}}, \quad 0 \le k \le N \tag{1}$$

$$P(k) = 10\log_{10} |F(k)|^2 \text{ (dB)}, \quad 0 \le k \le \frac{N}{2} \tag{2}$$

$$S_t = \{P(k) | P(k) > P(k \pm 1) \text{ and } P(k) > P(k \pm \Delta k) + 7(\text{dB})\} \tag{3}$$

$$\Delta k \in \begin{cases} \{2\}, & \text{if } \ 2 < k < 63 \ (0.086 - 0.271\,\text{kHz}) \\ \{2,3\}, & \text{if } \ 63 \le k < 127 \ (0.271 - 5.469\,\text{kHz}) \\ \{2,3,...,6\}, & \text{if } \ 127 \le k < 255 \ (5.469 - 10.981\,\text{kHz}) \\ \{2,3,...,12\}, & \text{if } \ 255 \le k \le 500 \ (10.981 - 21.533\,\text{kHz}) \end{cases}$$

$$P_{TM}(k) = 10\log_{10} \sum_{j=-1}^{1} 10^{0.1P(k+j)} (dB) \tag{4}$$

The set of tones in a frame, S_t, is defined by (3). From (3), a tone means a frequency signal greater than 7 dB compared to neighboring frequencies within the 0.5 to 1.0 Bark width. With the tones determined by (3), (4) calculates the SPL of the tonal masker, $P_{TM}(k)$. Figure 1 shows the power spectrum of one audio frame calculated by (2) and the SPL of a tonal masker calculated by (3) and (4).

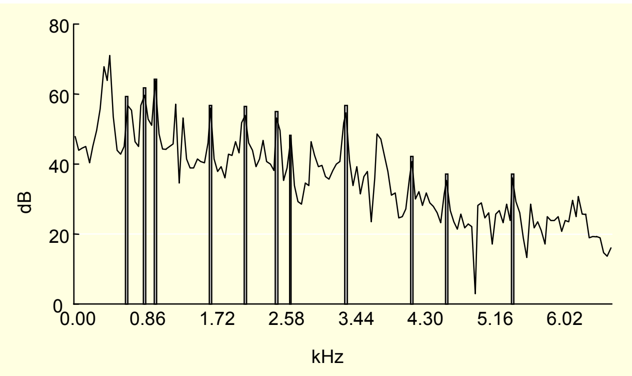

Fig. 1. Power spectrum and tonal maskers (omission of high frequency bands).

Equation (4), which calculates the SPL of a tonal masker, can be redefined as (5).

$$P_{TM}(k) = 10\log_{10} \sum_{j=-1}^{1} |F(k+j)|^2 (dB). \tag{5}$$

Equation (5) shows that a tonal masker is calculated as a log-scaled sum of the tone energy and the energies of the two adjacent frequency signals in the power spectrum (in this study, the square of an amplitude is referred to as "energy"). In other words, the SPL of a tonal masker is the accumulation of, with the same weight of the energies of the tone, its left and right frequency signals. According to this model, the tonal masker to be perceived as the same sound can be produced by changing the energies of the frequency signals located to the left and right of the tone as in (6), while maintaining the conditions delineated in (3) for a tone.

$$P_{TM}(k) = 10\log_{10}(\,|F(k)|^2 + (|F(k-1)|^2 + \Delta e) + (|F(k+1)|^2 - \Delta e)\ )\ (dB) \quad (6)$$

In this model [6], [7], the interval between the left and right frequency signals of a tone is 0.086 kHz, which can be interpreted as a bandwidth of about 0.5 Bark in the 1 kHz frequency range.

The following shows that, according to existing studies, the human ear perceives the modified sound as in (6) as the same sound. As seen in Figs. 2 and 3 [8], which show the results of a psychoacoustic study, the human auditory system perceives such modification to be of the same loudness. Figure 2 shows the SPL of band-pass noises with different bandwidths, which produce the same perceptible loudness. This figure shows the relationship between the bandwidths and the SPL of those noises that are perceived to be of the same loudness through experimenting with the uniform exciting noises from band-pass filters centered on 1 kHz with different narrow bandwidths. The human auditory system identifies these noises as having the same loudness if the sum of the SPL of the frequency signals of a noise is equal to the sum of another. Figure 3 shows what dB of a 1 kHz tone (sine wave) is perceived as having the same loudness as the uniform exciting noise on a band-pass filter centered on 1 kHz according to the change in bandwidth while adjusting the sum of the SPL of band-passed signals to a constant of 20, 40, 60, and 80 dB. Figure 3 shows that, as long as the sum of the SPL of the signals with less than 0.16 kHz bandwidths remains the same, the human auditory system perceives them to be the same sounds as the 1 kHz tones that have the same SPL. Here, 0.16 kHz corresponds to the bandwidth of one Bark in approximately a 1 kHz frequency range.

According to the results of the existing studies, like those in Figs. 2 and 3, the modification in (6) maintains the same sum

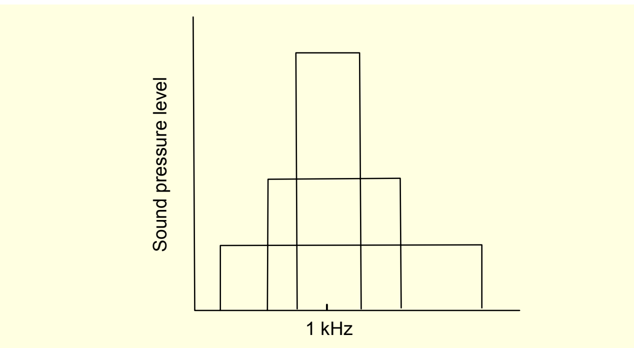

Fig. 2. Sound pressure level of band-pass noises for different bandwidths producing the same perceptible loudness.

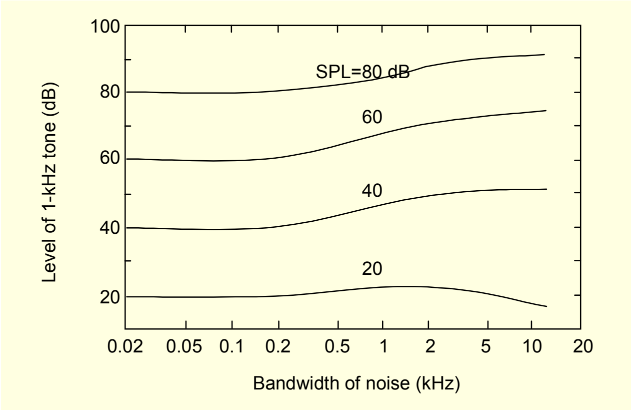

Fig. 3. Sound pressure level of a 1 kHz tone judged to be as loud as a band-pass noise at a center frequency of 1 kHz. The parameter is the sum of the SPL of the band-pass noise.

of the SPL within the 0.086 kHz bandwidth, and therefore the modifications to the sound are undetectable. Additionally, it is hard for the human auditory system to distinguish the change in auditory pitch due to the modification of frequency signals within the bandwidth of 0.5 Bark, as in (6) [8].

The results of the existing studies mentioned here attest that the modified tonal masker in (6) can be one of the carriers of imperceptible watermarking.

This section investigates whether the modification in (6) can be utilized in robust watermark embedding. If the signals comprising a tonal masker show a very small change after signal processing is applied, it means that it can be used as an aspect of strong watermarking. This study analyzes the variation rates of the following values before and after signal processing to confirm the stability of signals consisting of a tonal masker:

(i) the energy of a tone, E_t,
(ii) the energies of the left and right frequency signals of a tone, E_l and E_r, and
(iii) the energies of the left and right frequency signals of a tone divided by its energy, that is, E_l/E_t and E_r/E_t.

The signal processing applied in this analysis is as follows:

- noise addition with a constant level of 36 dB,
- band-pass filtering with 100 Hz and 6 kHz cutoff frequencies,
- compression of MPEG-1 Audio Layer 3 with a bit rate of 64 kbps,
- re-sampling with 96 kHz,
- echo addition with a delay of 100 ms and a decay of 50 %, and
- equalization with the following characteristics.
  frequency (Hz): 31, 62, 125, 250, 500, 1k, 2k, 4k, 8k, 16k

gain (dB): -6, +6, -6, +6, -6, +6, -6, +6, -6, +6

The variation rate was calculated according to (7) by extracting values E_t, E_l, E_r, E_l/E_t and E_r/E_t before and after signal processing:

$$\Delta c = \frac{x' - x}{x}, \qquad (7)$$

where x is the value before the signal processing and x' is the value after the processing. This experiment used 1024 sample frames of the Allegro from Vivaldi's Four Seasons with 16 bits/sample and a mono channel at a sampling rate of 44.1 kHz, calculating the variation rate by extracting 50,000 values for each of E_t, E_l, E_r, E_l/E_t and E_r/E_t. These values were extracted among pairs of tones to have the same tonal frequencies before and after processing from a frequency range below 6.6 kHz. The range is relatively stable and commonly used in watermarking.

To identify the general variation rate of each value against signal processing, Fig. 4 shows the 10% trimmed mean $\overline{\mu}_{.10}$ and deviation $\overline{\sigma}_{.10}$ of the measured variation rates. The experiment shows variation rates between -0.037 and +0.0313 at maximum, except for the rates of equalization in (i) and (ii) shown in Fig. 4, where the equalization is a type of signal processing by which signal energies of neighboring frequency ranges increase or decrease by the same rate. As a ratio between the energies of the left and right frequency signals and the energy of a tone, (iii) in Fig. 4 shows a small rate of change of 0.004 on average against various types of signal processing including equalization. This result reveals that if the watermark is embedded utilizing the relationship between the energies of the left and right frequency signals and the energy of a tone, the watermark will also sustain minimal loss against signal processing attacks.

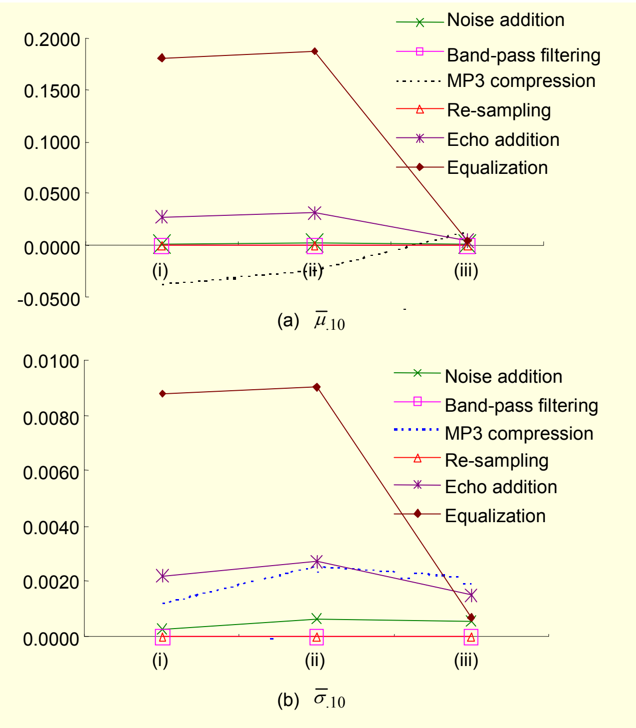

Fig. 4. Variation ratios of the energies of tonal masker's components on various signal processing types. The analyzed data : (i) the energy of a tone, (ii) the energies of the left and right frequency signals of a tone, and (iii) the energies of the left and right frequency signals of a tone divided by the energy of the tone.

Therefore, when a sufficiently sized Δe in (6) is applied to compensate for the change made by these types of signal processing, a modified tonal masker can be a carrier for robust watermarking.

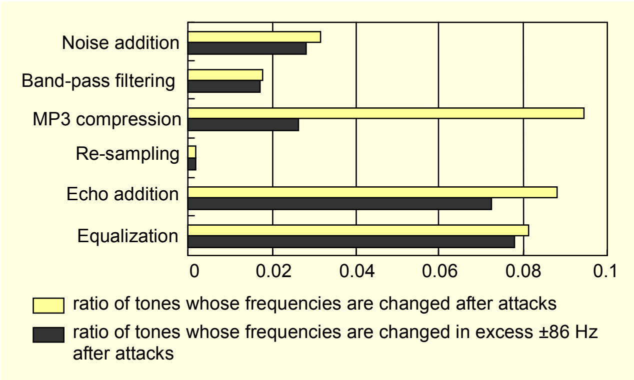

Fig. 5. Ratio of tones whose frequencies are changed to whole tones on various signal processing types.

Figure 5 shows a ratio of tones whose frequencies are changed after various signal processing types to the whole tones. The white bar shows the ratio of tones whose frequencies were created, changed, or deleted to the whole tones after the attacks. The black bar shows the ratio of the tones, excluding the tones whose frequencies were changed within ±86 Hz from the tones shown with the white bar. The black bar indicates that echo addition and equalization changes the tonal frequencies by more than 0.07. And comparing the white and black bars of MP3 compression, such compression changes the tonal frequencies more than 0.068 within ±86 Hz. From this result, we consider that repetitive embedding of a watermark in several frames can complement such variations. If the method to select tones is made under consideration of these changing attributes of tones, watermarking using tonal makers will be sufficiently robust.

So far, we have examined the stability of tonal maskers as watermarking carriers. Next, we look at the capacity of information embedding using this modification. The analysis of the tone distribution of various genres of audio reveals that ten tones on average are distributed in one frame. Figure 6 shows the tone's distribution by Bark in 50 consecutive audio frames chosen from one sample audio file. As shown in this figure, most audio data have more or less equal distribution of tonal maskers in all 25 Barks. These results show that in the audio region, excluding the silence or noise region where there is no tone, there is sufficient space for information to be embedded using the tonal masker modification.

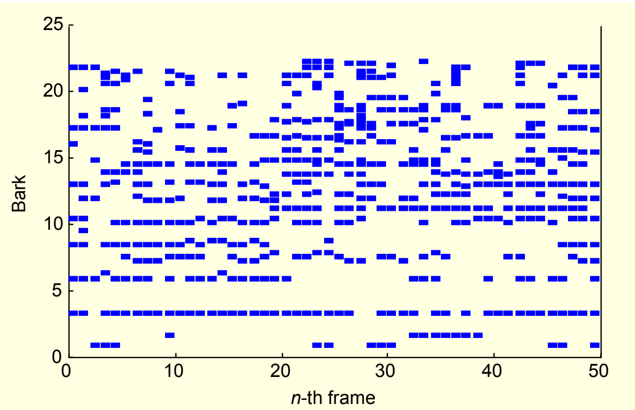

Fig. 6. Distribution of tonal maskers at successive frames.

III. Watermarking Scheme

In this chapter, a new watermarking method is proposed. Figure 7 shows the relationship between three energies of tone, the left frequency signal and the right frequency signal for the embedded binary states, which is defined in (8).

$$\begin{aligned} &\text{State } 0, \quad \text{if } R(E_l, E_r) \le L_1 \\ &\text{State } 1, \quad \text{if } R(E_l, E_r) \ge L_2 \end{aligned} \tag{8}$$

where

$$\begin{aligned} &R(E_l,\ E_r) = \left|E_l - E_r\right|/2, \\ &L_1 = \left(E_t - E_{avg}\right)\alpha, \qquad L_2 = \left(E_t - E_{avg}\right)\beta, \\ &E_{avg} = (E_l + E_r)/2 \end{aligned}$$

and α and β are values determined by the trade-off between sound quality and robustness.

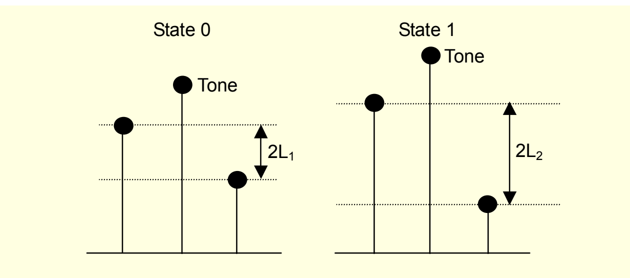

Fig. 7. Binary state to be applied in the watermark embedding.

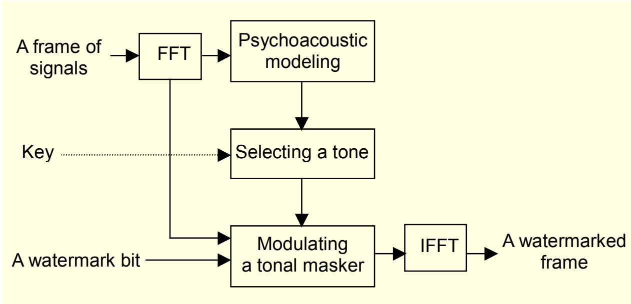

Fig. 8. Watermark embedding at a frame.

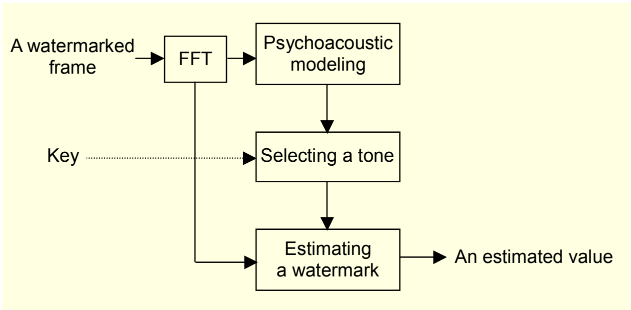

Fig. 9. Watermark extraction at a frame.

Figure 8 outlines the process involved in embedding a single watermark bit into one frame. The detailed procedure is as follows:

Step 1. Construct a frame in n samples and do a fast Fourier transform (FFT) according to (1).

Step 2. Select one tone for watermark encoding according to the procedure described below:

a. Calculate the power spectrum of a frame according to (2).

b. Make a set of tones according to (3).

c. Select a frequency at random using a key in the mid- and low-frequency range. Decide on a direction to discover the tone with the same key (toward a lower or higher frequency). Select the tone to encounter for the first time while searching for it in the decided direction from the selected frequency.

Step 3. Modify the left and right frequencies of the selected tone according to the binary watermark bit to be embedded, as delineated in (9).

$$\begin{aligned} &\text{if } WM = 0 \text{ and } R(E_l, E_r) > L_1 \\ &\qquad E_l = E_{avg} + L_1, \quad E_r = E_{avg} - L_1 \qquad \text{if } E_l \ge E_r \\ &\qquad E_l = E_{avg} - L_1, \quad E_r = E_{avg} + L_1 \qquad \text{if } E_l < E_r \\ &\text{if } WM = 1 \text{ and } R(E_l, E_r) < L_2 \\ &\qquad E_l = E_{avg} + L_2, \quad E_r = E_{avg} - L_2 \qquad \text{if } E_l \ge E_r \\ &\qquad E_l = E_{avg} - L_2, \quad E_r = E_{avg} + L_2 \qquad \text{if } E_l < E_r\,. \end{aligned} \tag{9}$$

Step 4. Perform an inverse FFT (IFFT) for the modified frame.

Watermarks cannot be embedded in either silence or noise regions where there is no tone. To complement the deficiency of such regions, a sequence of watermarks is repeatedly embedded several times.

Figure 9 outlines how to extract one embedded watermark bit from a frame. The tone hiding the watermark can be found using the same key as that used during embedding, and the estimated value Ws can be calculated using (10). The value of Ws is between -1.0 and 1.0. To better understand Ws in (10), Ws for any $R(E_l, E_r)$ is shown in Fig. 10. The estimated value indicates how near the modification of a tonal masker approaches the criteria of the states of 0 and 1.

$$
\begin{aligned}
&Ws = -1.0, \quad \text{if } R(E_l, E_r) \le L_1 \\
&Ws = +1.0, \quad \text{if } R(E_l, E_r) \ge L_2 \\
&Ws = ((R(E_l, E_r) - L_1) - L_3)/L_3, \quad L_3 = (L_2 - L_1)/2 \\
&\qquad \text{if } L_1 \le R(E_l, E_r) \le L_2
\end{aligned} \tag{10}
$$

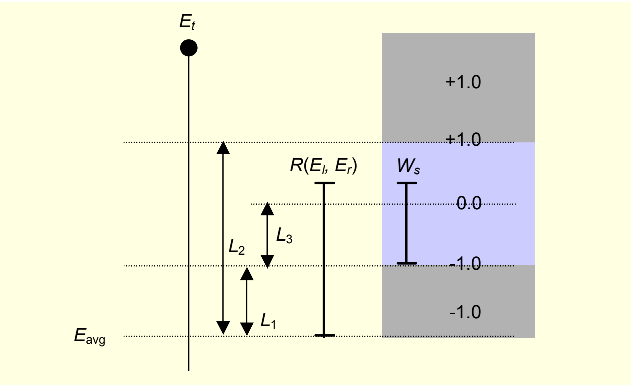

Fig. 10. Example of Ws of $R(E_l, E_r)$.

After repeatedly extracting these estimated values, the embedded watermark sequence is extracted by (11). If the information of N bits (a watermark sequence) is repetitively embedded k times, kN frames are required. In (11), Ws_m is an estimated value calculated from the m-th frame, A_j indicates the average of Ws_m corresponding to the j-th bit of a watermark sequence, T is a threshold, and WO_j is the j-th bit of the watermark sequence finally determined. When A_j is between $+|T|$ and $-|T|$, it indicates a failure to extract the j-th bit of the watermark sequence.

$$
\begin{aligned}
&WO_j = 1, \quad \text{if } A_j > +|T| \\
&WO_j = 0, \quad \text{if } A_j < -|T|
\end{aligned} \tag{11}
$$

where $A_j = \frac{1}{k} \sum_{i=1}^{k} Ws_m$, $m = n(i-1) + j$.

According to the field of application, the watermark system adjusts threshold T to control the false alarm probability. By adjusting α and β, it is possible to respond to the requirements of sound quality and robustness. For robust watermarking, it is advantageous to make the difference between α and β as large as possible. However, as the difference between α and β is made large, it results in significant variations in signal, thereby resulting in distortions of sound quality occurring in the boundary between frames. Considering both sound quality and robustness, α and β must be determined.

It is important to find a relatively stable tone set in the proposed method. There are frequency signals having large energy during an instantaneous time among audio signals, especially pop audio signals. In terms of a general definition of tone, such signals cannot be considered a tone. If the frame size is big, such signals can be filtered out more from the tone set. Accordingly, for robust watermarking, determining a proper FFT frame size is required.

Moreover, the proposed method should compensate for the noise on the border of frames that is produced due to the modification of tonal maskers. The main purpose of audio watermarking is not to reconstruct the original signal; it is to include information while not creating a noticeable difference in sound quality from the original piece. To maintain the sound quality and, at the same time, reduce the loss of the modified tonal maskers, it is better to use a trapezoid window [8] rather than to use a Hann window or Hamming window [9]. Furthermore, the length of the overlap between frames must also be kept in mind. This must be sufficient in order to preserve the sound quality, but it is better for it to be as little as possible for the watermark extraction.

Considering the practical implementation, psychoacoustic modeling and selection of a tone, shown in Figs. 8 and 9, can be integrated into one process. That is, it is possible to discover a tone while applying (2) and (3) only to the frequency range selected by the key. Therefore, in the case of a platform where FFT is made in real time, it is possible to extract a watermark in real time.

This proposed watermarking system is implemented in a mono channel with 2048 sample frames and a 64 overlap length at 44.1 kHz / 16-bits per sample. The frequency range of tones for watermarking is also limited to between 0.86 and 5.2 kHz, which is relatively stable against various types of signal processing. This system embeds a 72-bit binary watermark sequence nine times in a 30 second audio file. The values of α and β are set to 0.05 and 0.4, respectively, and the threshold during extraction T is set to 0.0.

IV. Experimental Results

The proposed system was evaluated in regard to the imperceptibility of its watermarking. The test was conducted

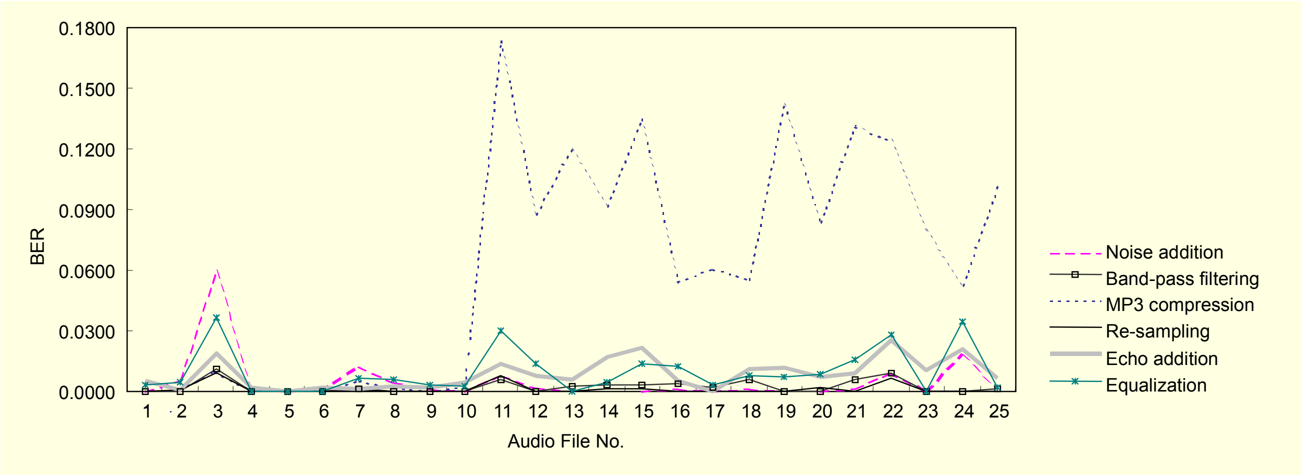

Fig. 11. BER of the proposed system.

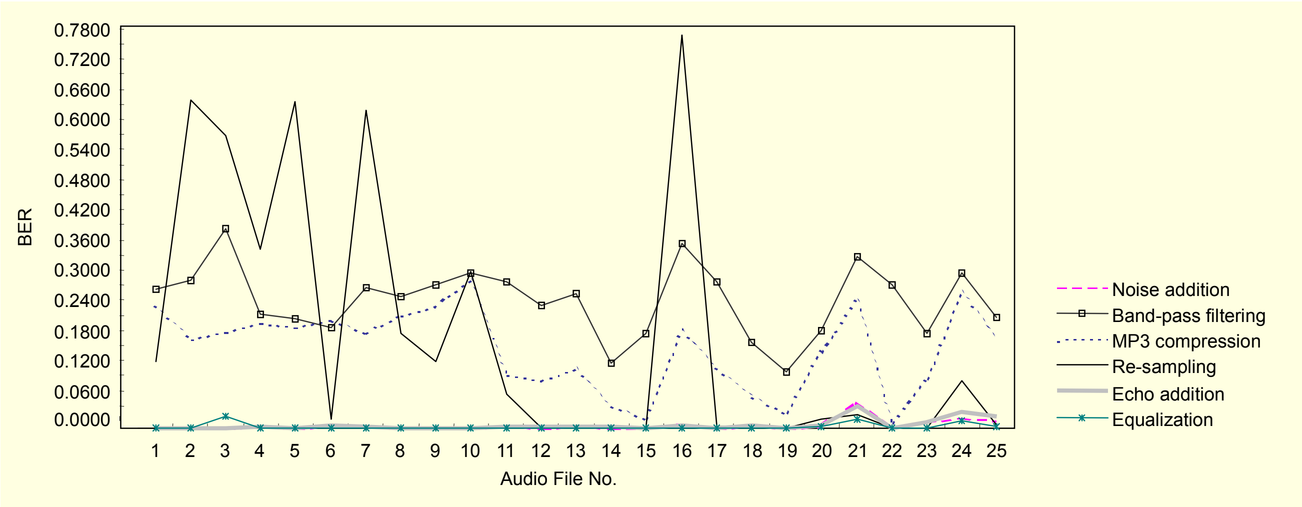

Fig. 12. BER of the SS-LP system.

using ITU-R's "double-blind triple stimulus with hidden reference" [10]. The results from 45 general listeners' evaluations showed an average subjective difference grade (SDG) of -0.285 with classical music and an average SDG of -0.340 with pop or light rock music. The resulting sound quality is believed to be satisfactory to a general listening audience.

The robustness of the proposed system was evaluated in comparison with the spread spectrum watermarking system using linear prediction filtering (SS-LP) [4]. In the comparison, the SS-LP system embedded a pseudo random noise (PN) sequence corresponding to 1 or 0 bits in each frame, and during extraction 10th order LP filtering was applied to separate audio from noise signals. After repeatedly calculating correlated values of two PN sequences, the embedded 72-bit watermark sequence was extracted in the statistical process.

To conduct an impartial comparison of robustness, the parameters of the SS-LP system were adjusted to make the SS-LP's watermarked audio quality to be similar to that of the proposed system. In the evaluation, there were 168 answers to compare the sound quality of the audio produced by these two systems. Seventeen of the answers said the sound quality was similar, 95 answers said that the sound quality of the proposed system was superior, and 56 said that the SS-LP system produced higher sound quality.

With these results, a robustness evaluation was conducted. We watermarked 25 audio pieces on each system and compared the extracted watermark's bit error rate (BER) after applying various types of signal processing.

Figure 11 shows the BERs from the proposed method, while Fig. 12 displays those of the SS-LP. The experimental data from 1 to 10 in Figs. 11 and 12 refer to the results from classical music, while data from 11 to 25 refer to the results from pop music, which includes various genres such as acappella, male or female singers, rock, and jazz. As shown in

Fig. 11, the proposed method is more robust with classical music than with popular musical forms. The results of pop audio show a relatively high BER. With classical audio pieces in general, the distinction between the tone and other frequency signals in the spectral domain is markedly maintained, while with pop audio pieces, other frequency signals have a high sound pressure level and are distributed widely in low- and mid-frequency ranges, showing many cases where the distinction between them and the tone is obscured.

As shown in Fig. 12, the SP-LP system is weak in re-sampling, band-pass filtering, and MP3-encoding in general. These three signal processing types each reduce noise signals in an audio piece, resulting in higher loss of the SS signals embedded in the time domain. Particularly in re-sampling, the embedded SS signals are changed notably because the signals are processed by interpolation and decimation. In counterpoint to the SS-LP system, the proposed system shows robustness in these types of processing because the modified frequency coefficients have little loss after signal processing attacks that reduce noise.

The proposed method in general, showed an average BER of 0.012 and a maximum BER of 0.18, while the SS-LP achieved an average BER of 0.093 and a maximum BER of 0.8. The proposed method also showed a relatively even level of robustness against each signal processing compared to that of the SS-LP method.

In addition to the above comparative test, different aspects of the proposed technique were analyzed. For a frame-synchronization error, which is a significant problem of the spread spectrum watermarking technique, the robustness of the proposed technique was evaluated. With the proposed technique it is possible to extract watermark sequences of 91% on average for the frame-synchronization error of 100 samples.

However, for time scaling, it is impossible to extract watermarks with the proposed technique. Time scaling (pitch-invariant time scaling ±4%) is generally used when a very small increase/reduction in audio length is required. In order to keep pitch-invariant, many samples are irregularly copied or deleted. For 44.1 kHz audio, about 500 samples are copied in two or three positions during a one-second period. Like most of the watermarking techniques vulnerable to such attacks, the proposed technique needs a preprocessing process, requiring the development of a frame synchronization algorithm.

V. Conclusion

In this paper, we proposed a new scheme for digital audio watermarking and evaluated its performance. The proposed watermarking scheme utilizes a modified tonal masker as an audio watermarking carrier. We explained that the human auditory system has difficulty in distinguishing the modification of tonal maskers through the existing studies. Also, a variation rate analysis has shown that the signals of the tonal maskers underwent only a very small change when subjected to a variety of signal processing types. This means that the modification of the tonal masker is a certainly viable method to be used in audio watermarking that is both robust and imperceptible.

An evaluation of the proposed method's "imperceptibility" concluded that the system produced an average SDG of -0.285 with classical music and an average SDG of -0.340 with pop music. The robustness of the proposed system was tested in comparison with that of the SS-LP watermarking system. The tonal masker scheme proved itself to be robust against common signal processing types such as noise addition, band-pass filtering, compression, re-sampling, echo addition, and equalization.

To improve the robustness, a new method that can achieve a more stabilized tone set is needed. Further research will focus on solving this problem.

References

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for Data Hiding," *IBM Systems J.*, vol. 35, 1996, pp. 313-336.

[2] M.D. Swanson, B. Zhu, A.H. Tewfik, and L. Boney, "Robust Audio Watermarking Using Perceptual Masking," *Signal Processing*, vol. 66, 1998, pp. 337-355.

[3] D. Kirovski and H. Malvar, "Robust Spread-Spectrum Audio Watermarking," *Proc. ICASSP 2001*, vol. 3, 2001, pp. 1345-1348.

[4] J. Seok, J. Hong, and J. Kim, "A Novel Audio Watermarking Algorithm for Copyright Protection of Digital Audio," *ETRI J.*, vol. 24, no. 3, June 2002, pp. 181-189.

[5] J.D. Gordy and L. Bruton, "Performance Evaluation of Digital Audio Watermarking Algorithms," *Proc. IEEE MWSCAS*, 2000, pp.456-459.

[6] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," *Proc. IEEE*, vol. 88, no. 4, Apr. 2000, pp. 451-513.

[7] ISO/IEC IS11172-3, *Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s—Part 3: Audio,* 1993.

[8] E. Zwicker, *Psychoacoustics-Facts and Models*, Springer-Verlag, 1982, pp. 208-213.

[9] H. Oh, *Robust Audio Watermarking Using Imperceptible Echo Embedding Technique*, doctorial dissertation, Yonsei Univ., Korea, 2001, pp. 122-131.

[10] ITU-R Rec. BS.1116, *Methods for The Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems*, 1994.

Hee Suk Lee received the BS degree in computer sciences from Changwon National University, Changwon, Korea, the MS degree in computer engineering from Kyungpook National University, Daegu, Korea, and in 2003 the PhD degree in computer engineering from Changwon National University, where she is currently teaching in the Department of Computer Engineering. Her research interests include digital signal processing in the field of digital content protection and management.

Woo Sun Lee received the BS, MS and PhD degrees in the Department of Electronic Engineering of Korea University, Seoul, Korea in 1981, 1983, and 1988. He is currently a Professor in the Department of Computer Engineering at Changwon National University. His research interests are in digital signal processing and microprocessor applications.