논문 2006-43CI-5-6

대역확산과 심리음향 모델을 이용한 고음질 오디오 워터마킹

(High Quality Audio Watermarking using Spread Spectrum and Psychoacoustic Model)

노 진 수*, 이 강 현**

(Jin Soo NOH and Kang Hyeon RHEE)

요 약

본 논문에서는 심리음향 모델과 MDCT/IMDCT(Modified DCT/Inverse Modified DCT)를 이용하여 고음질 오디오 워터마킹 알고리즘을 제안하였다. 일반적으로 디지털 오디오 워터마크는 디지털 오디오 신호를 주파수 영역으로 변환 한 다음 주파수 영역에 워터마크를 삽입하지만 삽입된 워터마크에 의해 디지털 오디오 음질이 영향을 받게 된다. 제안된 알고리즘에서는 디지털 오디오 데이터를 PN(Pseudo Noise) 코드를 사용하여 확산시킨 다음 심리음향 모델을 참조하여 MDCT 과정을 통하여 오디오 워터마크를 삽입시킨다. MDCT 과정에서 고음질의 오디오를 얻기 위해 필터뱅크 출력의 첨예도에 따라 256, 1,024 또는 2,048 포인트의 윈도우가 선택되어진다. 본 논문에서 워터마크 계수 a가 2.5 이하일 때, 워터마크의 검출률이 SDMI(Secure Digital Music Initiative)의 제안 조건을 50% 이상 상회 하며, SNR은 4종류의 공격(압축, 절단, FFT, 에코)에 대해 50~68dB 값을 가짐을 확인하였다.

Abstract

In this paper, we proposed the high quality audio watermarking algorithm using MDCT/IMDCT (Modified DCT/Inverse Modified DCT) with psychoacoustic model. Generally, a digital audio watermark is embedding the frequency domain after frequency transform of the digital audio data but the digital audio quality is affected by watermarking. In our scheme, the digital audio data is spread with PN((Pseudo Noise) code and then audio watermark is embedded in MDCT processing that refers psychoacoustic model. In MDCT processing, according to the shape of filter bank output, the block switching selects a window sequence that has 256, 1,024 or 2,048 points interval for high quality audio. The author confirm that when watermark weight a is 2.5 below, the detection ratio of watermark is a satisfied to SDMI's(Secure Digital Music Initiative) recommendation 50% above and SNR is 50~68dB range with mainly 4 kind of attacks(Compression, Cropping, FFT and Echo).

Keywords : Audio Watermarking, Psychoacoustic Model, MDCT, IMDCT

I. Introduction

In the present digital world, the value of information is very great. Information enables users to claim their copyright or ownership by copying and modifying data as information can be accessed in various manners due to the development of internet. As another protection technologies being issued, watermarking expresses intellectual property right of a digital work by inserting copyrighter's own information into it. Digital watermarking is a high technology to protect intellectual property right of digital works on web. Watermarking can be applied both to video and audio data. For audio watermarking, watermark is embedded

* 학생회원, ** 평생회원, 조선대학교 전자공학과
(Dept. of Electronic Engineering, Chosun University)
접수일자: 2006년7월24일, 수정완료일: 2006년8월27일

into frequency domain after frequency transform. During such watermarking, audio data is influenced by audio quality. In audio aspect, the quality of audio is very important, as value of commodity is reduced if it is damaged by watermarking when watermark is embedded into audio data [1-3].

Several digital audio watermarking techniques have been proposed over the last ten years. They can be categorized into five major techniques: least-significant-bit (LSB) coding, phase coding, echo coding, patchwork and spread spectrum [4-7, 11-14]. And recently, an active research both of the considered psychoacoustic model and time-to-frequency transform such as DCT and DWT are published 3. In [8],the watermark extractor must know a quantization scale that is used at the watermark embedder for audio watermarking thus, computation quantity is increasing. Filter bank in [9] divides into 32 constant-bandwidth subband. In [10], DCT also has a constant block size of transform for audio watermarking. In case of audio signal is suddenly change, the transform processing between each other block can raise a noise phenomenon. That is not desirable for high quality audio.

This paper deals with high quality audio watermarking using psychoacoustic model and MDCT/IMDCT to solve such problems. In our scheme, the digital audio data is spread with PN code and than audio watermark is embedded in MDCT processing that refers psychoacoustic model. In MDCT processing, according to the shape of filter bank output, the block switching selects a window that has 256, or 1,024 or 2,048 points for high quality audio. In IMDCT processing, psychoacoustic model is referred again and that results multiplying by PN code is produce the watermarked audio.

The organization of this paper is as follows: Sec. II illustrates the theoretical background of psychoacoustic model and MDCT/IMDCT for high quality audio watermarking. Sec. III addresses the proposed audio watermarking algorithm. Sec. IV shows the experimental results and discussion of the proposed algorithm. Finally, the conclusion is drawn in Sec. V.

II. Theoretical background

Among audio watermarking methods, Low-bit coding and Phase coding, Spread spectrum, and Echo hiding are the most basic[4-7, 11-14]. Fig. 1 shows general watermark embedding and extraction process.

In general, sampling, quantization and digitalization of audio data are performed on the basis of sampling theory. Especially, audio data sampled at 44.1KHz, quantized at 16bit and digitalized by PCM for high quality audio. Audible frequency band can be all covered to be digitalized by such means with very high quality audio maintained. In addition, digital data can be easily copied in a complete form. Such full repeatability is a great merit for digitalization, while essentially requires copyright protection of digital works[16,17]. Boney used a method of embedding copyright information into audio data in an imperceptible type. The Boney's method of embedding digital watermark by the auditory masking is not proper for high quality audio data and has a problem in the concealment of watermark as the amplitude element of audio is directly embedded [11,13,15]

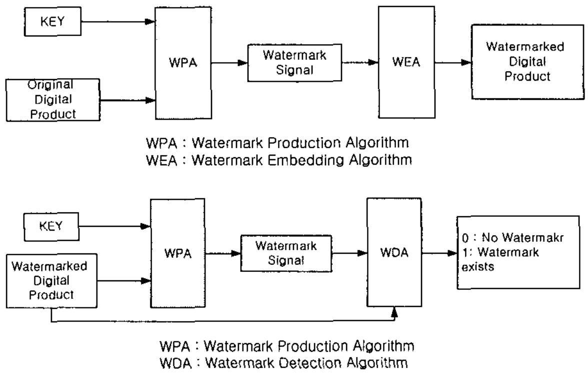


그림 1. 워터마킹 삽입(위), 추출(아래)알고리즘
Fig. 1. Watermark embedding (top) and extraction (bottom) Algorithm.

1. PN-sequence

PN-sequence is a basic element of watermarking, as it is not influenced by external interference. Auto-correlation is a feature of PN-sequence. Here, multiplying n bit of PN-sequence by even times and dividing it by N becomes 1, multiplying it by odd times and dividing it by N becomes 1/N, and multiplying it

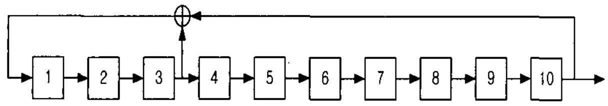


그림 2. PN 시퀀스
Fig. 2. PN-sequence.

by one bit delayed PN-sequence value becomes -1/N. Eq. (1) shows such relation, and Fig. 1 shows PN-sequence generation module has 10 registers and feedback with the outputs of the registers no. 3 and 10.

$$R(k)=\begin{cases}1 & k=0\\ \dfrac{-1}{N} & 0<|k|<N\end{cases} \tag{1}$$

2. Psychoacoustic model and MDCT/IMDCT

Psychoacoustic model is input audio data that is multiplied by PN-sequence to provide necessary an information for MDCT/IMDCT. Audible band ranges from 20Hz to 20KHz. In this range, the feature would be different by frequency, and the sound is inaudible until the signal level reaches more than a certain level. This is called the masking threshold value of the auditory frequency and is presented as the following Eq. (2) and Fig. 3.

$$ATH(f)=3.46\,(f/1000)^{-0.8}-6.5e^{-0.6(f/1000-33)^2}+10^{-3}(f/1000)^4 \tag{2}$$

ATH(f) indicates dB within ATH, and if f is frequency value. The frequency level of lower and higher band is more high than middle band (1KHz~5KHz). It means that important information of an audio is contained in the middle band. The lowest

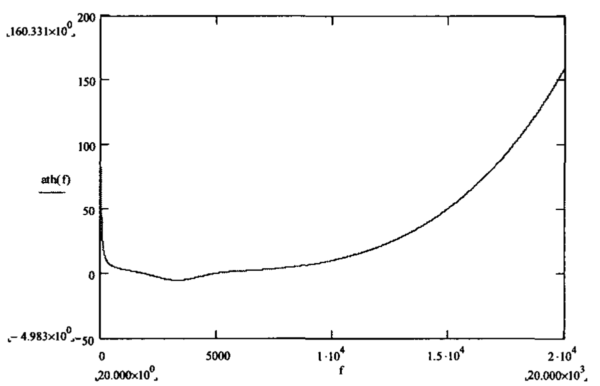


그림 3. 절대 가청 주파수
Fig. 3. Absolute threshold frequency.

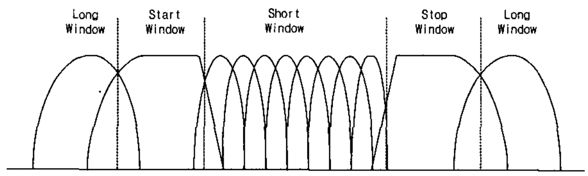


그림 4. MDCT 분해 윈도우
Fig. 4. MDCT analysis window.

point in the above curve has the same sound pressure with 1 LSB of sine level. Definitely, human ear cannot maintain fully high frequency of a momentary sound for a certain period.

In this module, Masking Threshold of input signal is calculated. For masking threshold, hanging window is weighted, and audio data input for 512 or 2,048 sample audio segments is analyzed for 50% overlay of the continuous blocks to provide window information for MDCT/IMDCT. MDCT shall perform frequency transform with the window information. Fig. 4 shows MDCT analysis window.

MDCT removes aliasing by subband coding with TDAC(Time Domain Aliasing Cancellation). MDCT /IMDCT used in this paper are shown in Eq. (3) and (4) each. MDCT converts the time domain data into the frequency domain, and reversibly IMDCT converts the frequency domain data into the time domain data. MDCT/IMDCT is block operation and therefore window is used for smooth connection between blocks. N, the length of synthesis window for transform is decided by window_sequence expressed in 2 bits. Table 1 shows N, the length of synthesis window by window_ sequence.

$$X(i,k)=2\cdot\sum_{n=0}^{N-1}x(i,n)\cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right)$$
$$for\ \ 0\le n \prec N/2$$
$$\begin{cases}x_{(i,K)} = Spectral\ coefficient\\ n = Sample\ index\\ k = \cos ine\ coefficient\ index\\ i = Block\ index\\ N = Window\ length\\ n_0=\left(\dfrac{N}{2}+1\right)\dfrac{1}{2}\end{cases} \tag{3}$$

표 1. 윈도우 시퀀스
Table 1. Window sequence.

N	2048, ONLY_LONG_SEQUENCE(0X0)
	2048, LONG_START_SEQUENCE(0X1)
	256, EIGHT_SHORT_SEQUENCE(0X2)
	2048, LONG_STOP_SEQUENCE(0X3)

$$X(i,n)=\frac{2}{N}\cdot\sum_{n=0}^{\frac{N}{2}-1}x(i,k)\cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right)$$
$$for\ \ 0\le n\prec N/2$$
$$\begin{cases} x_{(i,K)}=Windowed\ input \\ n=Sample\ index \\ k=Spectral\ coefficient\ index \\ i=Window\ index \\ N=Window\ length \\ n_0=\left(\frac{N}{2}+1\right)\frac{1}{2} \end{cases} \tag{4}$$

In Table 1, they are LONG_BLOCK using 2,048 samples except EIGHT_SHORT_SEQUENCE. Therefore, 256 samples are processed in block with SHOT_BLOCK only in case of window_sequence=0x3, and others are processed with LONG_ BLOCK.

As MDCT/IMDCT suggested in this study uses derived window and sine window, shape as well as length of window shall be decided. It is decided WINDOW_SHAPE of the current block and WINDOW_SHAPE_PREVIOUS_LOCK of the past window shape. And, by making the whole length of window into 1/2, the window shape of WINDOW_ SHAPE is used in the first half and that of WINDOW_ SHAPE_PREVIOUS_ BLOCK is used in the latter half.

In case of WINDOW_ SHAPE/WINDOW_SHAPE_ PREVIOUS_BLOCK=0x1, then KBD window is used. Like the above, by directly embedding watermark into the frequency transform domain, decoding it, and then multiplying PN-sequence, watermarked audio data without audio deterioration can be obtained [17-20]

III. Proposed high quality audio watermarking

1. Embedding watermark

In this scheme, to guarantee encapsulation of embedded watermark, spread spectrum was used, and to guarantee high quality audio data, psychoacoustic model and MDCT/IMDCT adopted in MPEG audio are used.

The whole algorithm is configured as follows;

<Algorithm configuration for the audio watermark>

A: Audio data
W: Watermark
W′: Watermark embedded in audio data
*A*PN*: Multiplication of audio data and
PN: PN-Sequence
PSY: Psychoacoustic model
MDCT: Modified DCT
IMDCT: Inverse Modified DCT

{*Procedure 1*} Multiplication A and PN.
{*Procedure 2*} Sending the A*PN value to MDCT and PSY.
{*Procedure 3*} Computation of window value in PSY.
{*Procedure 4*} Performing frequency transform of the A*PN value with the result value of PSY in MDCT, and then embedding W.
{*Procedure 5*} Transform into the time domain by IMDCT.
{*Procedure 6*} Multiplication Procedure 5′ result and PN for synchronization.

Pseudo code is as follows;

<Pseudo code for the audio watermark proposed>

```
WHILE N DO
 A*PN
PSY(A*PN)
WHILE N DO
MDCT(A*PN)
THEN WATERMARK INSERTION
 IF W′= TRUE
         IMDCT(W′)
```

ELSE WATERMARK INSERTION
END IF
*W'*PN*
END

And, the proposed flow chart of audio watermarking is shown in Fig. 5.

DSSS(Direct Sequence Spread Spectrum), one of spread spectrum types, is used to guarantee encapsulation of digital watermark by embedding watermark into audio data. As DSSS needs private key of the PN-Sequence is set as unique key of digital watermark. Multiplication of audio data and two signals in the proposed flow chart of Fig. 5 is authorized into MDCT. Spectrum signal that used for protection from interference and channel noise. Because PN-sequence inhibits interference of noise and has excellent auto-correlation feature.

Synchronization by inserting watermark into the frequency domain, converting the time domain by IMDCT, and multiplying PN-sequence restores original audio data to be decoded. Audio signal is spread by multiplication of PN-sequence and audio data is input value of psychoacoustic model and MDCT for the embedding audio watermark.

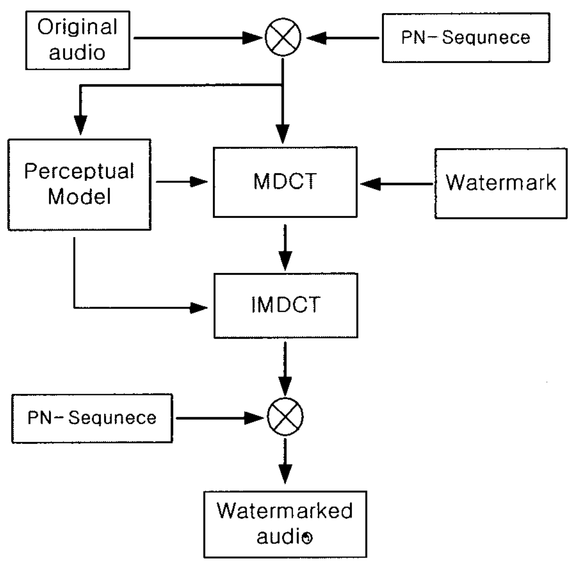


그림 5. 제안된 오디오 워터마킹 블록도
Fig. 5. Proposed block diagram of audio watermarking.

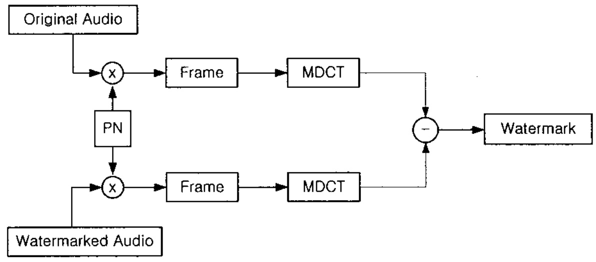


그림 6. 워터마크 추출 블록도
Fig. 6. Block diagram of watermark extraction.

2. Watermark extraction

Amount of watermark insertion is Z obtained with audio sampling frequency and number of samples N as shown in Eq. (5).

$$Z = \frac{f_s}{N} \tag{5}$$

Therefore, compared to audio deterioration, bits for each frame can be embedded readily and amount of embedding is proportional to number of embedding bits per frame. For the process of audio watermark extraction, watermarked audio data and PN-sequence of the same module as that used in watermark embedding are multiplied and synchronized, and then watermark is extracted in the frequency domain by performing again MDCT. This procedure is shown in Fig. 6.

If difference PN-sequence from the watermark insertion stage is used, the audio quality itself may be deteriorated, as well as right watermark cannot be detected. When watermark is inserted into audio data with high quality audio, to prevent reduction of the audio quality by watermark is very important. Hence, the insertion of watermark into high quality audio data must be considered.

Ⅳ. Experimental Results and Discussion

The proposed audio watermarking algorism for high quality audio was implemented with C++ for experiment of this paper and, the experiment is executed on Pentium-IV with 3GHz and 1G memory. In this scheme, stereo audio data at 44.1KHz and 16bit

표 2. 원 신호와 워터마크된 신호 사이의 오디오 특성 비교 결과
Table 2. Compared results of audio characteristics between original and watermarked signal.

Measured type	Results of original audio.(A)	Results of watermarked audio. (B)	(A)-(B)
# of Channel	1	1	-
Min. sample value	-28315	-28315	-
Max. sample value	28384	28384	-
Peak amplitude	1.31dB	1.24dB	0.07dB
Min. RMS power	-28.4dB	-27.4dB	-1.0dB
Max.RMS power	18.58dB	17.58dB	1.0dB
Average RMS power	20.28dB	19.28dB	1.0dB
Total RMS power	20.55dB	19.57dB	0.98dB
Using RMS window	50ms	50ms	-

표 3. 여러 가지 타입의 공격에 대한 [11]과 제안된 알고리즘의 성능 비교(Digital audio : 44.1KHz, 16bit, Stereo)
Table 3. Comparison between [11] and proposed scheme on several attacks type(Digital audio : 44.1KHz, 16bit, Stereo).

Watermark Weight	a=1.5		a=2.5		a=4	
Comparison	Conventional	Proposed	Conventional	Proposed	Conventional	Proposed
(a) No attack	17.02	17.02	17.02	17.02	17.02	17.02
(b) MP3 compression	17.02	16.88	17.02	17.02	17.02	17.02
(c) Cropping	16.06	16.83	16.06	16.85	16.06	16.85
(d) FFT	7.04	7.32	9.79	10.05	11.39	11.63
(e)Echo	7.21	7.33	10.19	10.25	11.95	12.54
(f) Average SNR(b)~(e)	17	68	13	53	10	45

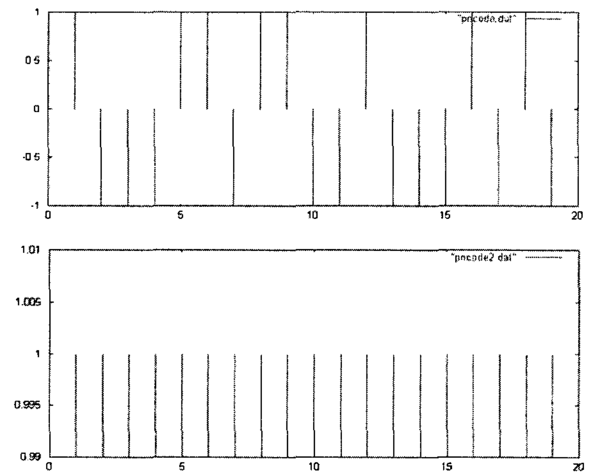


그림 7. PN 시퀀스(위)와 자기 상관(아래)
Fig. 7. PN-sequence(top) and its auto-correlation (bottom).

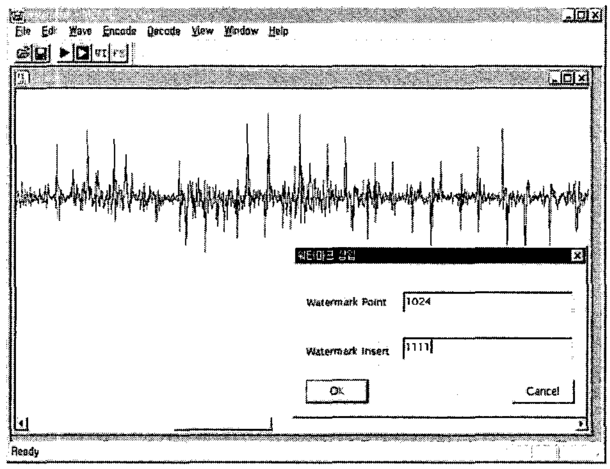


그림 8. 오디오 워터마킹 응용 프로그램
Fig. 8. Audio watermarking application program.

quantized classic audio data and castanets audio data were used. In case of audio data for experiment is stereo sound, the author was embedded watermark data only into single channel. Original audio is multiplied by PN-sequence before MDCT processing and psychoacoustic Modeling and then watermark data is embedded. Fig. 7 shows PN-sequence ant its auto-correlation waveform.

Fig. 8 shows the embedding watermark on the application program of proposed algorithm in this

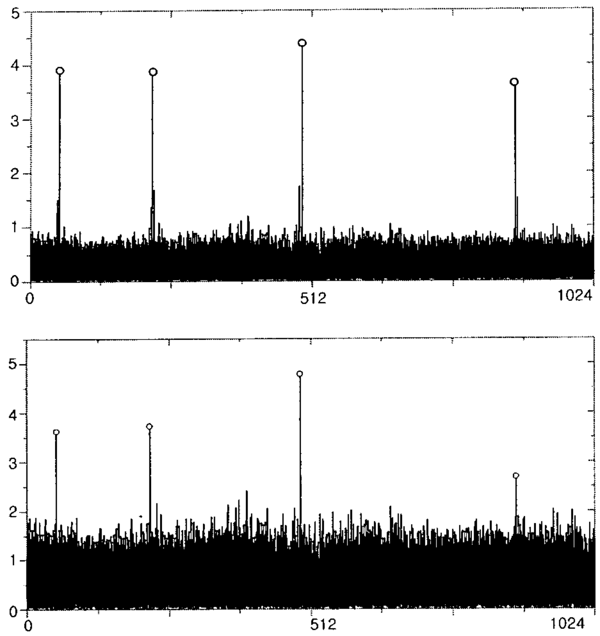


그림 9. 추출된 워터마크(위)와 MPEG-1 Layer 3 압축 및 복호된 후 추출된 워커마크(아래)

Fig. 9. Extracted watermark(top) and Watermark extraction after MPEG-1 Layer 3 compression and decoding(bottom).

paper. Watermark is embedded into block point of audio data that is, each 256, 1,024 and 2,048 MDCT windows as shown in Fig. 4.

The proposed watermarking algorism for evaluation, the watermarked audio signal in this paper has the attacks of MP3 compression, Cropping, Echo, FFT and Gaussian Noise. Fig. 9 shows watermark extraction after embedding watermark into the frequency domain and decoding it with by IMDCT and the watermarked audio data after MPEG-1 Layer III compression and decoding. When the point of watermarking is on the sequence of 30, 250, 500 and 890, the given watermark weight a is 3.9, 3.9, 4.3 and 3.7 each other. The attenuation ratio of extracted watermark is smaller at the center than beside window.

Fig. 10 shows original audio, watermarked audio signal and there difference each other. Table 2 shows the analyzed and compared the results of original data and watermarked data.

In Table 2, the difference power of average and total RMS are 1dB below. In Table 3, when a is 2.5 below, the detection ratio of watermark is a satisfied to

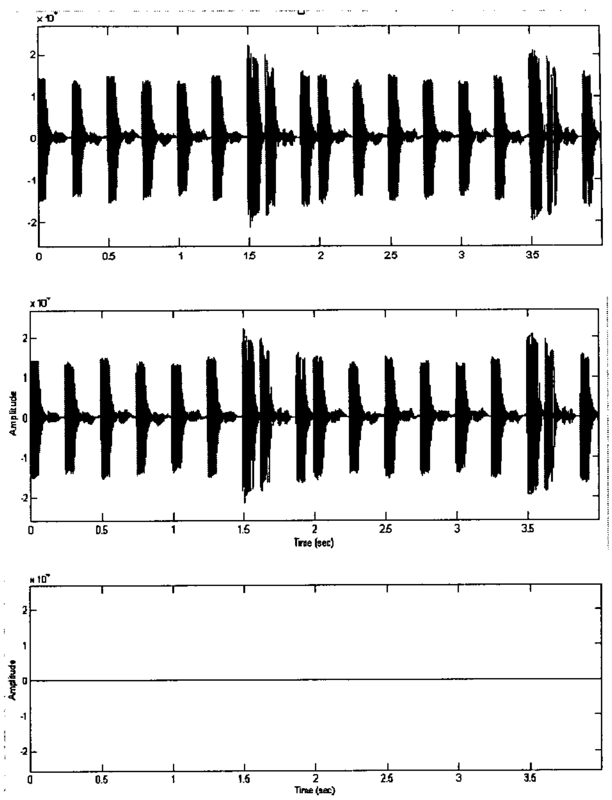


그림 10. 원 오디오 신호(위), 워터마킹된 오디오 신호(중간) 두 신호간의 차이(아래)

Fig. 10. Original audio signal(top), Watermarked audio signal(middle) and Difference of original and watermarked audio signal(bottom).

SDMI's recommendation above 50%, and SNR is obtained 50~68dB range with mainly 4 kind of attacks.

V. Conclusion

This paper proposed watermarking for high quality audio. Our scheme used a method of directly embedding watermark into the frequency domain which is processed on MDCT/IMDCT with psychoacoustic model. For the experiment, the configuration of audio input data is 44.1KHz sampling rate, 64Kbps~256Kbps of bit rate(stereo and mono). According to the window sequence of MDCT/IMDCT, it able to adjust the intervals of embedded watermark thus amount of watermark is flexibly variable. And watermark data can be embedded into a certain point of input audio data with parameter that is window sequence and watermark weight. The proposed algorithm used MDCT/IMDCT with psychoacoustic model so it can be widely applied to MPEG audio watermark in multimedia area.

References

[1] KIM Jeong-tae, and RHEE Kang-hyeon, "A Study on the design of MDCT/IMDCT for MPEG Audio," Summer Academic Meeting, The Institute of Electronics Engineers of Korea, vol. 22, pp. 530-533,.

[2] "gISO/IEC MPEG-2 Advanced Audio Coding4382(N-1)"-Presented at the 101st Convention,1996 November 8-11 Los Angeles, California, AUDIO ENGINEERING SOCIETY PREPRINT,.

[3] Brandenburg K, & Stoll G, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," Journal of the Audio Engineering Society, No. 10, pp. 780-792, Oct. 1994.

[4] Cox, I. J., Miller M. L., Bloom J. A., "Digital Watermarking," Academic Press, 2002.

[5] Schmucker, A. M., Wolthusen S. D., "Techniques and Applications of Digital Watermarking and Content Protection," Artech House, 2003.

[6] Bender, W., Gruhl D., Morimoto N., Lu A., "Techniques for data hiding," IBM Systems Journal, Vol. 35, No. 3 & 4, pp.313-336, 1996.

[7] Xing, H., Scordilis, M. S., Iliev, A., A Novel High Capacity Digital Audio Watermarking System, in Proc. of IEEE ICASSP 2004, Montreal, May, 2004.

[8] Libin Cai and Jiying Zhao, "AUDIO QUALITY MEASUREMENT BY USING DIGITAL WATERMARKING," CCECE-CCGEl, Niagam Falls, Maylmai, pp.1159-1162, 2004.

[9] Xiaomei Quan and Hongbin Zhang, "Perceptual Criterion Based Fragile Audio Watermarking Using Adaptive Wavelet Packets," Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), 2004.

[10] Dong Hoon WOO, "Digital audio Watermarking Inserted in the Specific Coefficients of the Frequency Domain," Master Degree Thesis, Graduate school of Ulsan University, Korea, Dec. 2002.

[11] Ingemar J. Cox, Joe Kilian, Tom Leighton and Talal Shannon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Trans. on Image Processing, Vol. 6, No. 12, pp.1673-1687, 1997.

[12] Gruhl, A. Lu, and W. Bender, "Echo Hiding," in Proc. Information Hiding Workshop (University of Cambridge, U.K), pp.295-315, 1996.

[13] V. Basia and I. Pitas, "Robust audio watermarking in the time-domain,"Proc. Europe. Signal Processing Conf., Sept. 1998.

[14] Jack Lacy, Schuyler R. Quackenbush, Amy R. Reibman and James H. Snyder, "Intellectual property protection systems and digital watermarking," OPTICS EXPRESS. Vol. 3, No. 12, 7 Dec. 1998.

[15] L. Boney, A.H. Tewfik, & K.N. Hamdy, "Digital watermark for audio signals," Proc. IEEE Int. Conf. on Multimedia Computing and Systems, June 1996.

[16] F. Hartung & M. Kutter, "Multimedia watermarking Technique," Proceedings of IEEE, Vol. 87, No. 7, July 1999.

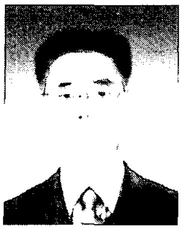
[17] R. Wolfgang, C. Podichuk and E. Delp, "Perceptual watermarks for images and video," Proceedings of the IEEE, May 1999.

[18] W. Bender, D. Gruhl and N Morimoto, Techniques for data hiding, In Proc. of SPIE, Volume 2420, p.40, Feb. 1995.

[19] R. Dixon, Spread Spectrum System with Commercial Applications, Wiley, New York, 1994.

[20] R. G. van Schyndel, A. Z. Tirkel and C. F. Osborne, "A Digital Watermark," Proc. ICIP94, vol. II, p.86, 1994.

저 자 소 개



노 진 수(학생회원)
2002년 조선대학교 전자공학과 학사졸업.
2004년 조선대학교 전자공학과 석사졸업.
2006년 조선대학교 전자공학과 박사과정.
<주관심분야 : UWB, 생체인식, 양자컴퓨팅>

이 강 현(평생회원)-교신저자
1979년, 1981년 조선대학교 전자공학과 공학사 및 석사
1991년 아주대학교 대학원 공학박사
1977년~현재 조선대학교 교수
1991년, 1994년 미 스탠포드대 CRC 협동연구원.
1996년 호주 시드니대 SEDAL 객원교수
2000년~현재 한국 멀티미디어기술사협회 이사
2002년 영국 런던대 객원 교수
2002년 대한전자공학회 멀티미디어연구회전문위원장
2003년 한국 인터넷 방송/TV 학회 부회장
2003년~현재 대한전자공학회 상임이사
2005년~현재 조선대학교 RIS 사업단장
<주관심분야 : 멀티미디어 시스템설계, Ubiquitous convergence>