AUDIO WATERMARKING BASED ON SINUSOIDAL AMPLITUDE MODULATION

Akira Nishimura

Tokyo University of Information Sciences,
Department of Media and Cultural Studies, Faculty of Informatics,
1200–1, Yatoh-cho, Wakaba-ku, Chiba-city, Chiba 265-8501, Japan

ABSTRACT

A new watermarking system based on amplitude modulation is proposed. Sinusoidal amplitude modulations at relatively low modulation frequencies applied to the neighboring subband signals in opposite phase are used as the carrier of embedded information. The embedded information is encoded in the form of the relative phase differences between the amplitude modulations applied to several groups of subband signals. Extraction of the amplitude modulations from the watermarked signal is performed by calculating the logarithmic ratio of the amplitude envelopes extracted from the neighboring subband signals. Deterioration of the sound quality resulting from the watermarking is barely detectable because modulation detection interference in the human auditory system and masking in the frequency-domain disturb the perceptual detectability of the additional amplitude modulations.

1. INTRODUCTION

Recent studies on audio watermarking have resulted in significant progress in inaudibility and reliability. Audio watermarking techniques have achieved robustness against, for example, MPEG compression, additive noise, low pass filtering, pitch change and time scale modification [1, 2]. The inaudibility and reliability of most techniques depend on the acoustic characteristics of the host signal. However, few studies have confirmed the reliability of the watermarking techniques using a number of musical samples or actual sample sounds. In addition, robustness against emission in closed spaces, i.e., reverberation and reflection disturbance, has rarely been considered. Robustness against reverberations and reflections are important for audio watermarking techniques for live performances.

In the present paper, a new watermarking system based on subband coding using amplitude modulation is proposed. The system is robust for perceptual audio codings such as MP3 or RealAudio, reflections and reverberations, additive Gaussian or colored noise and spectral modifications. A key is required for embedding and decoding, and no host signal is required for decoding. Robustness testing has been conducted for 100 pieces of music of various genres.

2. AUDIO WATERMARKING SYSTEM BASED ON AMPLITUDE MODULATION

2.1. Embedding method

At the beginning of the embedding process, a host signal $H(t)$ of the length of a data frame period is separated into $2n$ subband signals $h_m(t)$ by a filterbank with equal bandwidth:

$$H(t) = \sum_{m=1}^{2n} h_m(t). \tag{1}$$

Sinusoidal amplitude modulations (SAMs) at relatively low modulation frequency (f-Hz) are applied to the neighboring subband signals $h_{2m}(t)$ and $h_{2m+1}(t)$ in opposite phase. An embedding key defined by a known pseudo-random number generator arbitrary classifies n subband pairs into k subband groups. It also defines random initial phase angles $r(m)$ of SAM for each subband pair. The output of an amplitude modulated subband pair $x_m^i(t)$ which belongs to the i-th subband group is given by

$$\begin{aligned} x_m^i(t) \quad &= h_{2m}(t)(1 + A(m)\sin(2\pi ft + r(m) + p(i))) + \\ &\quad h_{2m+1}(t)(1 - A(m)\sin(2\pi ft + r(m) + p(i))), \end{aligned} \tag{2}$$

where $A(m)$ is the depth of SAM of the m-th subband pair. Embedded information is encoded by Phase Shift Keying (PSK), defined as the differences between phase angles of SAM of the first subband group and that of the i-th subband group, $p(1)$ and $p(i)$ $(i = 2, ..., k)$. 4-PSK encodes 2-bit information $(D_i = 0, 1, 2, 3)$ to every $\pi/2$ phase angle of $p(i)$.

$$p(i) = \begin{cases} 0 & i = 1; \\ \dfrac{\pi D_i}{2} & i = 2, ..., k. \end{cases} \tag{3}$$

As a result, $2(k-1)$ bits information is embedded per data frame period. Multiplex watermarking can be applied using different modulation frequencies simultaneously. Finally, a watermarked signal $X(t)$ is obtained by summing up all amplitude modulated signals $x_m(t)$.

$$X(t) = \sum_{m=1}^{n} x_m(t). \tag{4}$$

The embedding key determines the initial phases of the SAMs applied to all pairs of the neighboring subbands and also determines which neighboring subbands belong to the same subband group. Combined with the small SAM depth applied to each subband, the embedded watermark offers high concealment.

Synchronization of the data frames is achieved by inverting the relative phase of the SAMs between successive frames for the first subband group.

2.2. Extraction method

The amplitude envelopes of the subbands $E_m(t)(m = 1, ..., 2n)$ of the watermarked frame signal can be derived from the amplitude spectrum of the half-overlapped running FFT, where t is the unit of time defined by a half of the FFT length. Extraction of the embedded modulation waveform $G_m(t)$ is performed by calculating the logarithmic ratio of the amplitude envelopes extracted from the neighboring subband signals.

$$G_m(t) = \log \frac{E_{2m}(t)}{E_{2m+1}(t)}. \tag{5}$$

Synchronized addition of the AM waveforms for the i-th subband group is conducted after compensation of the initial phase differences $r(m)$ in order to emphasizes the modulation waveform (Fig.1). Consequently, the modulation depth $A(m)$ in the embedding process can be kept small. Initial phase differences between the first and the i-th subband, that is embedded information, are obtained by comparing phase angles of the FFT spectra calculated from the extracted AM waveforms.

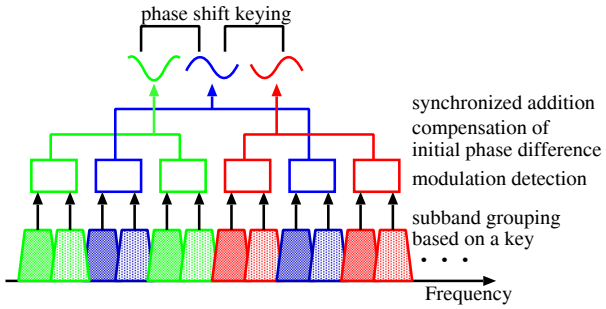

Fig. 1. Overview of watermark extraction.

2.3. Finding a starting point of the data frame

Before decoding PSK data, a starting point of the embedded data frame must be detected. A rectangular temporal window $W(t)$ of the data frame length T is iteratively applied to the modulation waveform $G^1(t)$ extracted from the first subband group. A starting point of the windowing is denoted u in Eq. 6. Then, $F(u)$ is derived by subtracting the synchronized addition of the modulation waveforms in the odd order windows from the synchronized addition of the modulation waveforms in the even order windows.

$$\begin{aligned} H_u &= W(t)G^1(t) \\ &= \{G^1(u), G^1(u+1), ..., G^1(u+T-1)\} \end{aligned} \tag{6}$$

$$F(u) = \sum_{v=0} H_{u+2vT} - \sum_{v=0} H_{u+(2v+1)T} \tag{7}$$

The Fourier amplitude of $F(u)$ which corresponds to the modulation frequency f, referred as $\mathrm{AMP}_f(F(u))$, exhibits a maximum when the positions of the window completely overlap with the positions of the frame. Consequently, the starting point y of the data frame is given by

$$y = \underset{u}{\mathrm{argmax}}\, \mathrm{AMP}_f F(u). \tag{8}$$

2.4. Method used to decide the depth of the SAMs

The depth of SAM $A(m)$ for the m-th subband pair is determined relative to the weighted sum of the magnitude of the amplitude fluctuations around the modulation frequency (MF) observed in the amplitude envelope of the subband signal $h_{2m} + h_{2m+1}$. In the present study, the weighting function of the amplitude fluctuation, that is, the characteristics of the modulation filter, is –10 dB at MF/2 and 2MF, and 0 dB at MF. The intensity of the watermarking is defined as the modulation depth relative to the amplitude at the output of the modulation filter. $A(m)$ is re-calculated for every two to four modulation periods.

This decision process takes into account modulation masking in the human auditory system [3]. Since the perceptual detectability of amplitude modulation is disturbed by the inherent amplitude fluctuations contained in the host signal, and the sidebands induced by the SAM are apt to be masked by the musical signal, deterioration of the sound quality induced by the watermarking is considered to be barely detectable.

3. COMPUTER SIMULATION OF WATERMARKING AND EXTRACTION

A computer simulation was conducted to confirm that the watermarking was successful for 100 pieces of music

in a music database containing various types of music (RWC-MDB-2001-G)[4]. Amplitude modulated watermarks of 3 Hz or 12 Hz are embedded in the initial 60-second left-channel signal of each piece of music. The details of the simulated conditions and parameters are shown in Table 1.

Modifications to the watermarked music included MP3 encoding and decoding, RealAudio8, reverberation, additive Gaussian noise and amplitude quantization conversion. Reverberation was applied by convolving a random Gaussian sequence with exponential decay. The level of additive Gaussian noise was determined relative to the level of the sinusoid of maximum amplitude in 16-bit quantization.

The bit detection rate, defined as the ratio of the number of correct bits to the total number of embedded bits, is shown in Table 2. The results revealed that the watermarking was robust for the perceptual audio codings, such as MP3 or RealAudio. More than half of the pieces of music retained 100% correct watermarks after MP3 encoding and decoding at greater than 48 kbps for an embedding intensity of –10 dB. The same was true for reverberations, except for the 12-Hz modulation frequency. Watermarks of low SAM frequency are robust for reverberations because valleys of SAM are difficult to fill by reverberation.

Table 1. Embedding conditions and parameters.

parameters	values
sampling freq.	44100 Hz
embedding region	$<$ 11025 Hz
subband pairs (n)	32
subband groups (k)	3
frame period	5 s (3-Hz), 2 s (12-Hz)
bit rate	0.8 bps (3-Hz), 2 bps (12-Hz)

4. PERCEPTUAL DETECTABILITY

Perceptual detection thresholds of the intensity of the watermarking were obtained by the transformed up-down method with the AXB discrimination task (70.7% threshold) for three 5-seconds musical signals, which were relatively discriminable among the other pieces of music in the music genre database. Watermarks of 3-Hz and 12-Hz amplitude modulations were simultaneously applied to the host signal. The other embedding conditions and parameters are the same as the computer simulation and are shown in Table 1. Three trained listeners participated in the experiment. All stimuli were presented diotically through headphones (STAX Lambda Nova Classic). The average detection thresholds obtained from at least three measurements are shown in Table 3.

Computer simulation of watermark extraction for these three musical signals revealed that bit correction rates were more than 95% at an embedding intensity of –20 dB after the modifications tested in the previous section, except for 32 kbps perceptual codecs and reverberation at 12-Hz modulation embedding.

These results confirmed that the minimal required magnitude for extraction is lower than the perceptual detection thresholds of the watermark. In other words, the effective embedded watermarks were barely perceptible.

Table 3. Detection thresholds of the watermark in five seconds of music from RWC-MDB-G-2001. (dB)

Track no.	subj. 1	subj. 2	subj. 3
No. 45	-21.7	-11.9	-14.7
No. 69	-10.5	-11.1	-16.4
No. 87	-18.6	-18.1	-15.7

5. DISCUSSION

Since the watermarking method is still relatively weak with respect to malicious attacks such as pitch change and time scale modification, there is room for improvement. Theoretically, pitch change is more serious than time scale modification, because pitch change shifts frequency boundaries between neighboring subbands and causes negation of amplitude modulation in the subbands. The amount of pitch change can be estimated by the decrease of the intensity of extracted modulation in high frequency region. Efficient step-by-step searching by expansion and compression along both the frequency and time axes to extract salient amplitude modulations should be introduced in the future.

Although the robustness for spectral modifications, additive colored noise and lowpass filtering are not practically confirmed in the present paper, these modifications will not seriously damage the present watermarking method. Spectral modification may change relative levels between neighboring subbands, however, the level change only affects a direct current component in the extracted SAM waveform after calculation of the logarithmic ratio between amplitude envelops of neighboring subbands. Since embedded amplitude modulations are equally applied to the wide frequency range of the host signal, watermarks survive after lowpass filtering or additive colored noise as long as the effective frequency regions for sound quality are not disturbed.

From non-official listening tests on several pieces of watermarked music, perceptual quality degradation of musical signals that contain sounds of several musical instruments at the same time were difficult to detect. Perceptual quality degradation was not detectable even when the intensity of watermarking was greater than 0 dB. The present decision scheme of the modulation depth is based on amplitude fluctuations within the channel.

Table 2. Bit detection rate after various modifications to the watermarked music. Numbers in parentheses indicate the number of pieces of music having a detection rate of 100% out of 100 pieces of music.

Mod. freq. and Mod. ratio	3 Hz, –10 dB	3 Hz, –20 dB	12 Hz, –10 dB	12 Hz, –20 dB
bit rate		MP3 encoding and decoding		
32 kbps	94.9% (29)	60.3% (0)	90.8% (1)	70.5% (0)
48 kbps	99.2% (86)	83.3% (20)	99.0% (47)	90.2% (9)
64 kbps	99.8% (96)	93.1% (54)	99.9% (92)	96.2% (36)
bit rate		RealAudio 8 encoding and decoding		
32 kbps	96.0% (6)	90.6% (2)	98.0% (13)	89.7% (0)
44 kbps	95.9% (15)	93.4% (5)	98.5% (17)	94.7% (0)
reverberation time		reverberation		
1 sec.	99.1% (76)	89.0% (22)	71.7%(0)	64.0% (0)
0.3 sec.	99.9% (97)	96.5% (52)	97.2%(18)	86.8% (10)
noise level (overall)		addition of white noise		
–50 dB	98.9% (88)	79.3% (2)	95.9% (41)	87.8% (1)
–70 dB	99.5% (86)	93.8% (49)	99.8% (84)	97.3% (33)
resolution		sampling bit conversion		
8 bit	97.0% (59)	91.0% (5)	96.3% (51)	94.5% (1)

However, it is well known that amplitude fluctuations of remote channels can interfere with modulation detection, which is called modulation detection interference (MDI)[5]. Therefore, MDI makes perceptual detection of embedded amplitude modulation difficult, and wide variation in the detection thresholds observed among various kinds of music may be partly due to MDI. Considering the results of the perceptual detectability test, detailed modeling of the MDI in the human auditory system [6] will enable more information embedding without perceptual deterioration of the sound quality.

The present watermarking method does not include a scheme for error correction to embedded data. The present paper reports the first proposal of a new method for data hiding and embedding. Practical implementation should adopt effective block or convolutional coding in order to achieve high reliability.

6. SUMMARY

A new watermarking system based on amplitude modulation is proposed. The system is robust against perceptual audio codings, reverberations and additive Gaussian noise. Detecting the existence of the watermarking is difficult without knowing the key used in the encoding process. No host signal is required for the decoding process. Perceptual detection of the quality degradation of the watermarked music is quite difficult if the watermark is embedded with an appropriate intensity.

Acknowledgments: This work is partly supported by a grant from the Okawa Foundation for Information and Telecommunications.

7. REFERENCES

[1] Ryuki Tachibana, Shuichi Shimizu, Seiji Kobayashi, and Taiga Nakamura, "Au audio watermarking method robust against time- and frequency-fluctuation," in Proc. of Security and Watermarking of Multimedia Contents III, SPIE, 2001, vol. 4314, pp. 104—115.

[2] Wei Li and Xiangyang Xue, "Audio watermarking basd on music content analysis: Robust against time scale modification," in Digital Watermarking: Second International Workshop, IWDW 2003, LNCS 2939, 2004, pp. 289—300, Springer-Verlag.

[3] T. Houtgast, "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am., vol. 85, pp. 1676—1680, 1989.

[4] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, "Rwc music database: Music genre database and musical instrument sound database," in Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003), 2003, pp. 229—230.

[5] William A. Yost and Stanley Sheft, "Modulation detection interference: Across-frequency processing and auditory grouping," Hearing Research, vol. 79, pp. 48—58, 1994.

[6] Torsten Dau, Birger Kollmeier, and Armin Kohlrausch, "Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration," J. Acoust. Soc. Am., vol. 102, no. 5, pp. 2906–2919, 1997.