19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

SUBJECTIVE AND OBJECTIVE QUALITY EVALUATION FOR AUDIO WATERMARKING BASED ON SINUSOIDAL AMPLITUDE MODULATION

PACS: 43.10.Pr, 43.60.Ek

Nishimura, Akira[1]
[1]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; akira@rsch.tuis.ac.jp

ABSTRACT
Audio watermarking based on sinusoidal amplitude modulation, which was previously proposed by the author, has been confirmed to be robust against perceptual audio codecs, reverberations and reflections, additive noises, and lowpass filtering by computer simulation. In order to evaluate the watermarking performance, two types of listening test as well as objective measurement of audio quality degradation based on ITU-R BS.1387 perceptual evaluation of audio quality (PEAQ) were conducted. The transformed up-down method was used to determine perceptual thresholds of the watermarking intensity for several musical sounds. A subjective evaluation test of audio quality degradation, which followed ITU-R BS.1116, was conducted for several watermarked or MP3 musical pieces. The resultant subjective difference grade (SDG) values of MP3 encoded music were significantly correlated with the objective difference grade (ODG) values obtained from the PEAQ measurement, whereas the SDG values of watermarked music were not. Improvements to the embedding algorithm of watermarking were effective in improving the robustness of the watermark by computer simulation using 100 pieces of music in the music genre database.

INTRODUCTION
The inaudibility and reliability of most audio watermarking techniques depend on the acoustic characteristics of the host signal. However, few studies have confirmed the reliability and inaudibility of a technique using a number of musical samples or actual sample sounds. In addition, robustness against emission in closed spaces, i.e., reverberation and reflection disturbances, has rarely been considered. Robustness against reverberations and reflections is important for audio watermarking techniques for live performances.

The author has developed a new audio watermarking technique based on sinusoidal amplitude modulation. The newly developed technique is robust against perceptual audio codings, additive Gaussian noise, and spectral modifications [1]. It is also robust against reflections and reverberations, because the technique applies relatively slow amplitude modulation in a long embedding frame of several seconds. The results of computer simulation with respect to robustness against these modifications, which were examined in a previous study [2], are shown in part in Table 2.

The present study evaluates the audio watermarking based on sinusoidal amplitude modulation in terms of subjective and objective quality evaluation for watermarked pieces of music. Computer simulation using 100 pieces of music revealed that modifications to the embedding algorithm and its parameter values improved the robustness of the watermark.

AUDIO WATERMARKING BASED ON AMPLITUDE MODULATION

Embedding process
At the beginning of the embedding process, a host signal *H(t)*, which is the length of one data frame period, is split into *2n* subband signals $h_m(t)$ using an equal-bandwidth filterbank:

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

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

$$x^i_m(t) = h_{2m}(t)\Big(1 + A(m)\sin\big(2\pi f t + r(m) + p(i)\big)\Big) + h_{2m+1}(t)\Big(1 - A(m)\sin\big(2\pi f t + r(m) + p(i)\big)\Big), \quad (2)$$

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

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

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

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

Synchronization of the data frames is achieved by inverting the relative phase of the SAMs between successive frames for the first subband group. Finding a starting point of the data frame before extraction.

Before decoding the phase shift keying data, the starting point of the embedded data frame must be detected. A rectangular temporal window of the data frame length *T* is iteratively applied to the modulation waveform $G_1(\tau)$ extracted from the first subband group. The starting point of the windowing is denoted by *u* in Eq. 5. 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.

$$R_u = \{G^1(u), G^1(u+1), ..., G^1(u+T-1)\}. \quad (5)$$

$$F(u) = \sum_{v=0} R_{u+2vT} - \sum_{v=0} R_{u+(2v+1)T}. \quad (6)$$

The Fourier amplitude of *F(u),* which corresponds to the modulation frequency *f*, is denoted by $AMP_f F(u)$ and exhibits a maximum when the position of the window completely overlaps the position of the frame. Consequently, the starting point of the data frame *y* is given by

$$y = \underset{u}{\mathrm{argmax}}\, \mathrm{AMP}_f\big(F(u)\big). \quad (7)$$

Extraction process

The amplitude envelopes of the subbands $E_m(\tau)(m = 1, 2, ..., 2n)$ of the watermarked frame signal can be derived from the amplitude spectrum of the half-overlapped running FFT, where o is the period of time defined as half of the FFT length. The embedded modulation waveform $G_m(\tau)$ is extracted by calculating the logarithmic ratio of the amplitude envelopes extracted from adjacent subband signals.

$$G_m(\tau) = \log \frac{E_{2m}(\tau)}{E_{2m+1}(\tau)}. \quad (8)$$

After compensation of the initial phase difference *r(m)* for each $G_m(\tau)$, synchronized addition of

$G_m(\tau)$, which belongs to the i-th subband group, produces an accumulated AM waveform $G_i(\tau)$, which emphasizes the AM waveform. Consequently, the modulation depth $A(m)$ in the embedding process can be kept small. Initial phase differences between the first and i-th subband groups, i.e., embedded information, are obtained by comparing the phase angles of the FFT spectra calculated from $G_1(\tau)$ and $G_i(\tau)$.

The intensity of a watermark in the embedding process is defined as the depth of the SAM $A(m)$. $A(m)$ is determined relative to the inherent fluctuation power obtained by the detection operation before embedding, which is division between the amplitude envelopes of the adjacent subband. In the previous study and the present listening tests, $A(m)$ was re-calculated for every two to four modulation periods [2]. In the present simulation, $A(m)$ is reconsidered and modified to be re-calculated for each data frame period.

SUBJECTIVE EVALUATION FOR WATERMARKED MUSIC

Two listening tests were conducted in order to assess the inaudibility and quality degradation of watermarked music. The parameter values of watermarking for the musical signals used in the listening tests are shown in Table 1.

Table I.- Watermarking conditions.

sampling freq.	44,100 Hz
embedding region	below 11,025 Hz
no. of subband pairs	128
no. of subband groups	5
mod. frequencies	2, 3, 5 Hz
max. AM depth per mod. freq.	0.316
length of data frame	5 s
total bit-rate of embedded data	4.8 bps

Measurement of thresholds of watermarking intensity

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 five-second musical signals, which were relatively discriminable among the other pieces of music in the music genre database RWC-MDB-G2001[3]. Four 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 four measurements are shown in Fig. 1.

The results revealed that the detection threshold of watermarking was approximately -10 dB for the most detectable musical sound for sensitive listeners. The following section deals with the amount of perceptual quality degradation induced by supra-threshold watermarking.

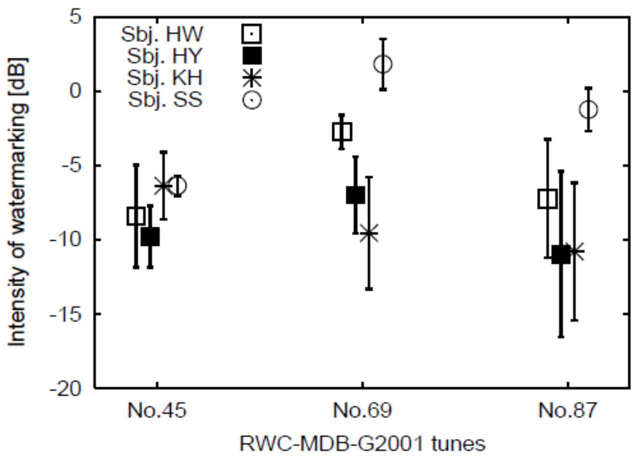

Figure 1. Detection threshold of watermarking intensity. Error bars show 1 standard error.

Assessment of subjective quality degradation

Assessment of subjective quality degradation for the audio signal must be treated carefully in order to avoid artifacts related to the experimental procedure. Listening tests based on the ITU-R recommendation BS.1116-1 were conducted for the method of the subjective assessment of small impairments in audio systems in order to assess the subjective quality degradation for watermarked musical signals.

A sixty-second period of degraded music and the reference music were randomly assigned to buttons labeled 'A' and 'B' on a computer screen. A button labeled 'X' was also displayed on a screen and was assigned to the reference. The subject can listen to each piece of music arbitrary by pressing one of the three buttons during the course of a trial. The reproduced sound was not interrupted by pressing the buttons until the sound ended because the three sounds were synchronized in the time scale. The listener was able to listen repetitively and was asked to assess the impairments on 'A' or 'B' compared to 'X' using a five-grade impairment scale with a resolution of one decimal place. All sounds were reproduced via headphones (STAX Lambda Nova Classic), connected to the audio card (M-AUDIO Delta 1010) through the amplifier (DENON AVC-1890).

The degradation of the musical signals was produced by watermarking and MP3 encoding. The embedding intensities of the watermarks were -5 dB and 0 dB for the three pieces of music used in the experiment to obtain the thresholds of watermarking intensity. In addition, the embedding intensities of the watermarks were 0 dB and +10 dB for No. 99, the quality degradation of which was found to be difficult to perceive in the preliminary experiment. The bit-rates of MP3 encoding were 128 kbps and 96 kbps. The MP3 encoded sounds were introduced to indicate typical sound quality degradation. Three trained listeners participated in the experiment.

Trials for 16 experimental conditions, consisting of the combinations of four pieces of music and four types of degradation, were repeated five or six times. The subjective difference grades (SDGs), the difference between the grades given to the hidden reference and the degraded object, were corrected and analyzed statistically. One- tailed t-tests revealed that each listener detected degraded versions of music significantly ($p < 0.06$) in more than 13 experimental conditions. The SDGs of non-significant detection were discarded in the following analysis.

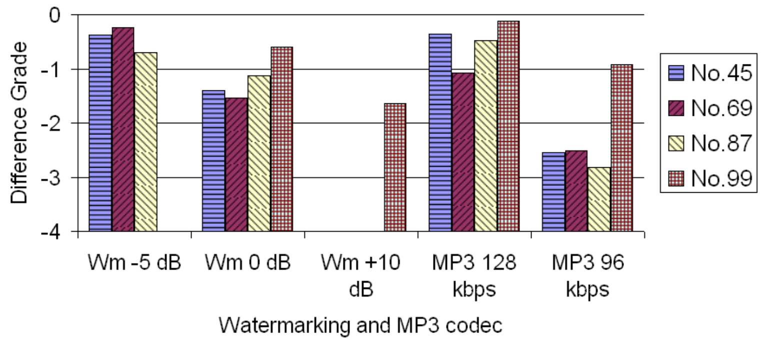

Figure 2. Difference grade obtained by ITU-R BS.1116-1.

The mean SDGs obtained from the three listeners on each degraded condition are shown in Fig. 2. The results reveal that the sound quality of -5 dB watermarked music is comparable to 128 kbps MP3 encoded music. The SDGs of 0 dB watermarked music exhibit an audio quality of MP3 encoding of between 128 kbps and 96 kbps.

OBJECTIVE EVALUATION OF QUALITY DEGRADATION

A method for the objective measurement of perceived audio quality (ITU-R Recommendation BS.1387), which is called PEAQ, uses a number of psycho-acoustical measures that are combined to provide a measure of the quality difference between two instances of a signal (a reference and a test signal). The implementation of the basic version of PEAQ by Kabel [4] was used to assess the objective quality degradation of the watermarked and MP3 encoded signals used in the previous section. The PEAQ measurement outputs the Objective Difference Grade (ODG), which corresponds to the SDG obtained from the procedure of BS.1116-1.

Figure 3 shows the ODGs obtained from 16 types of degradation and their SDGs for the three listeners. The filled squares represent MP3 encoded signals, and the filled circles represent watermarked signals. The SDG values of MP3 encoded music correlated significantly with the ODG values (R=0.515, p=0.007), whereas the SDG values of the watermarked music were not (R=0.192, p=0.22). The results of the correlation analysis indicate that the PEAQ measurement is effective for predicting the SDG of MP3 encoding to some extent, as PEAQ should be expected. However, the PEAQ measurement may not be effective for predicting the SDG of the watermarking based on AM.

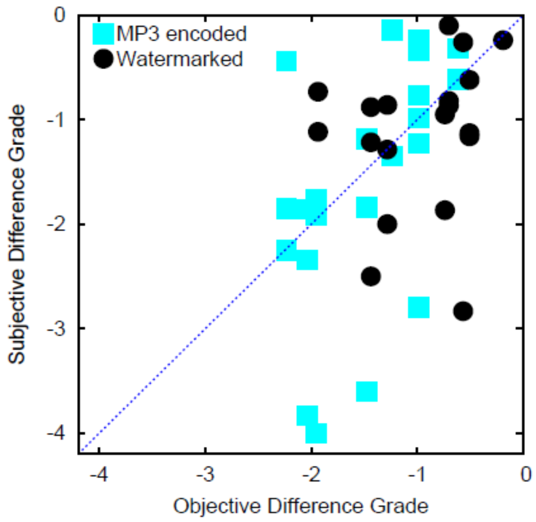

Figure 3. Objective and subjective difference grade of watermarked and MP3 encoded music.

PARAMETER TUNING AND COMPUTER SIMULATION

The effect of parameter tuning on robustness and the perceptual quality of watermarked sound was examined. Since the measurement of PEAQ may not be suitable to assess the perceptual audio quality degradation of an AM based watermark, the average effective AM depth over the embedding frequency region was calculated in order to compare the audio quality before and after improvement of the embedding algorithm.

The effective AM depth in decibels is a logarithmic measure of the root mean square of the instantaneous modulation waveform *AM (t)* relative to a reference value over a piece of music. The reference value is the depth of a 100% sinusoidal AM.

Two improvements were introduced to the embedding algorithm. The first is the timing of watermark intensity decision to be performed every data frame interval. The second is the number of subband pairs to be reduced from 128 to 64. The former is effective in improving the precision of measuring the inherent amplitude fluctuation at the modulation frequency in the subband signal, since a longer time frame leads to higher frequency resolution. As a result, the embedding intensity can be determined appropriately. The latter provides robustness against pitch shifting because a wider bandwidth absorbs frequency divergence among the embedding subbands and reduces robustness against reverberation.

Computer simulation of watermark detection using 100 pieces of music in the music genre database [3] was conducted. Modifications to the watermarked musical signals were RealAudio encoding and decoding, which includes sampling frequency conversion, reverberation, additive Gaussian noise, time stretch, and pitch shift. The embedding intensity in the previous study was -5 dB and that in the present simulation is 0 dB. This is due to smaller estimation of the inherent fluctuation by the improvement in frequency resolution at the stage of modulation frequency analysis in the embedding process.

In Table 2, the obtained results are compared with those of the previous study [2] in terms of the bit detection rate and the number of musical pieces that achieve over 85% bit detection. Improvements to both performance indexes, except for the time stretch, were observed. Since the average effective AM depth was 0.62 dB smaller than that obtained using the previous method, the present method will cause no impairment to the perceptual audio quality, as compared with the previous method.

Table 2. Mean bit detection rate and number of pieces that achieve over 85% detection after severe modifications to watermarked music. Comparison between the results of the previous study [4] and those of the present simulations.

Modification	Mean detection rate %		Number of pieces that achieve over 85% detection	
	previous	present	previous	present
RealAudio 21 kbps/ch	89.7	93.2	82	99
Reverberation 1 s	89.2	91.1	87	91
White noise SNR 20 dB	90.9	92.7	81	91
Time stretch 2%	94.9	94.2	91	91
Pitch shift 0.6%	54.4	90.9	0	93

DISCUSSION

Despite the improvement in robustness against pitch shift, the new method is not able to maintain a detection rate above 1% pitch shifting. The detection rate drops to 65% by a pitch stretch of 1%. This characteristic may make the watermarking system vulnerable to attacks made by those attempting to deactivate the copyright protection system. Logarithmic frequency spacing for the filter bank, instead of linear frequency spacing, is a way to improve robustness against pitch shift.

The most useful application to the AM-based watermarking system is data hiding in sound radiated in the air. The watermark is so robust against reverberation that modifications to frequency response and additive noise that acoustic communication in a hidden channel can be achieved. The data payload can be increased more than tenfold with respect to the tolerance of audio quality degradation.

SUMMARY

In order to evaluate the performance of watermarking based on amplitude modulation, two types of listening test as well as objective measurement of audio quality degradation based on PEAQ were conducted. The transformed up-down method was used to determine perceptual thresholds of the watermarking intensity on several perceptible musical sounds. The subjective evaluation test of audio quality degradation, which followed ITU-R BS.1116, was conducted for several watermarked or MP3 musical pieces. The resultant SDG values of MP3 encoded music were significantly correlated with the ODG values obtained from the PEAQ measurement, whereas the SDG values of watermarked music were not. Improvements to the embedding algorithm of watermarking were effective for improving the robustness of the watermark by computer simulation for 100 pieces of music in a music genre database.

References:

[1] A. Nishimura. Audio watermarking based on sinusoidal amplitude modulation. In Proceedings of ICASSP 2006, IV, pages 797-800. IEEE, 2006.
[2] A. Nishimura. Audio watermarking based on subband amplitude modulation. In Proceedings of the 2006 Symposium on Cryptography and Information Security, number 3F4-2. IEICE, 2006.
[3] M. Goto, H. Hashiguchi, T. Nishimura, and R. 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), pages 229-230, 2003.
[4] P. Kabal. An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality, TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, 2002.