ROBUST AND RELIABLE AUDIO WATERMARKING BASED ON DYNAMIC PHASE CODING AND ERROR CONTROL CODING

Nhut Minh Ngo, Brian Michael Kurkoski, and Masashi Unoki

School of Information Science, Japan Advanced Institute of Science and Technology

1–1 Asahidai, Nomi, Ishikawa, 923–1292, Japan

ABSTRACT

This paper proposes an audio watermarking method based on dynamic phase coding and error control coding. The technique of quantization index modulation is employed for embedding watermarks into the phase spectrum of audio signals. Since most of the audio information is distributed in moderately low frequencies, to increase robustness, this frequency region is chosen for embedding watermarks. Phase modification causes sound distortion in a manner that is proportional to the magnitude. Therefore, the amount of phase modification is adjusted according to the magnitude to balance inaudibility and robustness. Error control coding is incorporated to further increase reliability of watermark detection. The experimental results show that the watermarks could be kept inaudible in watermarked signals and robust against various attacks. Error control coding is effective in increasing watermark detection accuracy remarkably.

***Index Terms*—** audio watermarking, quantization index modulation, inaudibility, robustness, error control coding

1. INTRODUCTION

Recent developments in multimedia communication technologies have made life much easier but at the same time put the security of digital multimedia data at risk [1]. In that context, digital watermarking has attracted many researchers' interest in finding a solution not only for protecting copyright and ownership of multimedia data [2] but also for issues such as copy control, tampering detection, and covert communication [1]. In general, audio watermarking methods should satisfy four requirements: *inaudibility*, *blindness*, *robustness*, and *high capacity*. The solution is very hard because there is a trade-off among these requirements. It is straightforward that perceptually insensitive features should be exploited for embedding watermarks. But this is a challenge for robustness, since processing can distort the watermark without degrading the sound quality. Selecting suitable audio features for watermarking that satisfy both inaudibility and robustness is an important task for the design of watermarking algorithms.

Audio watermarking methods typically embed a watermark directly into audio samples in the time domain or audio features in a transformed domain. Some methods replace least significant bits (LSB) with watermark bits or insert a watermark which are perceptually shaped according to the human auditory system (HAS) [3]. Other methods take the advantages of simultaneous masking characteristics of HAS [4] or relative insensitivity of phase change [5] to embed inaudible watermarks. Phase has been exploited for inaudible audio watermarking since controlled phase alteration results in inaudible change in sound to HAS [6]. Several audio watermarking methods have been proposed based on quantization index modulation (QIM) [7–10] and showed that QIM is a promising technique for robust watermarking.

We previously proposed an audio watermarking method based on phase coding that applies QIM to the phase of low frequency components [11]. The experimental results show that the watermark is robust but the sound quality decreases when the bit rate increases. In this method, to embed watermarks, the phase of frequency components is statically modified, regardless of how resistant each frequency component is. However, strong frequency components could be less modified to reduce sound distortion while the resistance of watermarks is still ensured.

In this paper, we extend the previously proposed method to obtain a reasonable trade-off between inaudibility and robustness. We replace the static modification with a dynamic phase coding scheme for watermarking, in which the amount of phase modification is adjusted according the frequency component's magnitude. Large-magnitude frequency components are more sensitive to the modification of the phase of that component, with respect to sound distortion. Accordingly, to ensure low sound distortion and high robustness, larger-magnitude frequency components will have small phase modification, whereas smaller-magnitude frequency components will have somewhat higher phase modification.

In some applications, such as fingerprinting and authentication, very high capacity is not required but watermarks

This work was supported by a Grant-in-Aid for Scientific Research (B) (No. 23300070) and an A3 foresight program made available by the Japan Society for the Promotion of Science.

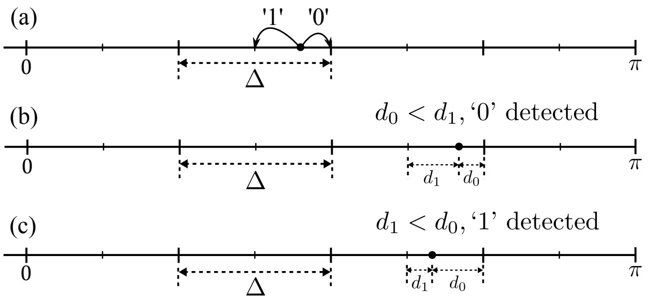

Fig. 1. Illustration of watermarking based on QIM: (a) embedding, (b) detection of '0', and (c) detection of '1'

need to be perfectly extracted. We further reduce watermark detection error rates by incorporating error control coding (ECC) into the watermarking system. The experimental results show that the embedded watermarks are inaudible and robust against various attacks. The incorporation of ECC is effective in a manner that watermarks could be extracted without any detection error at a bit rate of 102 bps.

2. PROPOSED METHOD

2.1. Quantization index modulation

QIM has been considered as a class of provably good methods for digital watermarking [7]. The procedure of embedding and detecting watermarks is quite simple. Figure 1 shows an illustration of embedding and detection processes. To embed a bit m, '0' or '1', into a scalar variable x, we quantize x to the nearest point that is an even or odd multiple of $\frac{\Delta}{2}$, respectively as (1). The obtained variable, y, is sent to receivers and might be affected by channel noise, hence becomes $\hat{y}$. To decode the embedded bit from $\hat{y}$, we calculate the distances between $\hat{y}$ and the nearest even multiple of $\frac{\Delta}{2}$, d_0 and the nearest odd multiple of $\frac{\Delta}{2}$, d_1 and then compare d_0 and d_1 to decode the bit as (2) and (3).

$$y = Q(x,m) = \begin{cases} \Delta \left\lfloor \dfrac{x}{\Delta} + \dfrac{1}{2} \right\rfloor & \text{if } m = \text{'0'} \\ \Delta \left\lfloor \dfrac{x}{\Delta} \right\rfloor + \dfrac{\Delta}{2} & \text{if } m = \text{'1'} \end{cases} \quad (1)$$

where $\lfloor . \rfloor$ is the floor function and Δ is the QIM step size.

$$d_j = \hat{y} - Q(\hat{y}, j), \quad j = \{\text{'0'}, \text{'1'}\} \quad (2)$$

$$\hat{m} = \arg\min_j d_j \quad (3)$$

2.2. Principle of watermark embedding

We apply QIM to the phase spectrum of audio signals to construct an inaudible, robust, and reliable audio watermarking system with the following considerations. (i) Phase alteration is relatively inaudible [6], hence slightly modifying the phase keeps watermarks inaudibly embedded. (ii) Most of the audio information is distributed in moderately low frequencies [12], thus this frequency region is more robust against attacks and should be chosen for embedding. (iii) Since modifying the phase of a frequency component causes sound distortion in a manner that is proportional to the magnitude of that component, the amount of phase modification should be adjusted to the magnitude. (iv) To increase the reliability, non-meaningful frequency components, i.e., very low magnitude components, are not used for embedding at all. (v) To further reduce detection error, ECC is employed to correct a number of errors.

Considerations (i), (ii), and (iv) were investigated and verified in the previous method [11]. In this paper, we investigate whether (iii) can help increase robustness and inaudibility simultaneously and (v) can further lower bit error rate by using ECC to encode watermarks before embedding process.

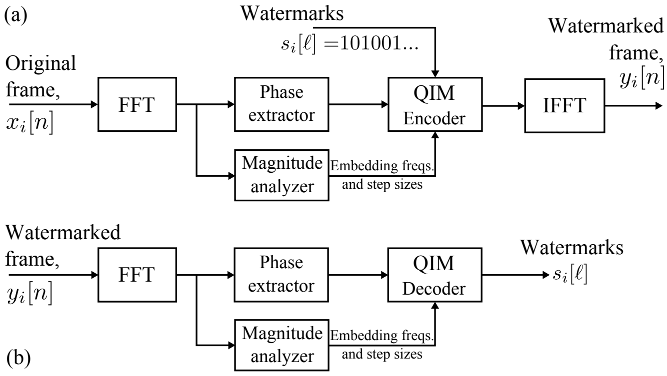

Fig. 2. Proposed scheme of audio watermarking: (a) embedding process and (b) detection process

2.3. Watermark embedding

The embedding process starts with frame segmentation of the original signal, $x[n]$ into frames $x_i[n]$ with a fixed frame size. Watermark bits $s_i[\ell]$ are embedded into audio frame $x_i[n]$. Figure 2(a) depicts a block diagram of the four steps that embed the watermark into an audio frame as follows.

Step 1. Original frame $x_i[n]$ is transformed into the Fourier spectrum $X_i(\omega)$ by fast Fourier transform (FFT). Magnitude spectrum $|X_i(\omega)|$ and phase spectrum $\angle X_i(\omega)$ are calculated.

Step 2. We select the frequency components up to K kHz that are meaningful, i.e., their magnitude is greater than a threshold ϵ. The watermark bits are embedded into only these selected components to increase reliability.

For each embedding component, the amount of phase modification that is quantified by a QIM step size is determined based on its magnitude. Firstly, the magnitudes are normalized to 1 and linearly divided into L ranges in which each range has a corresponding QIM step size. The higher range has a smaller QIM step size.

Step 3. The bits $s_i[\ell]$ are encoded into the phase of the selected components by (1) and a quantized phase spectrum

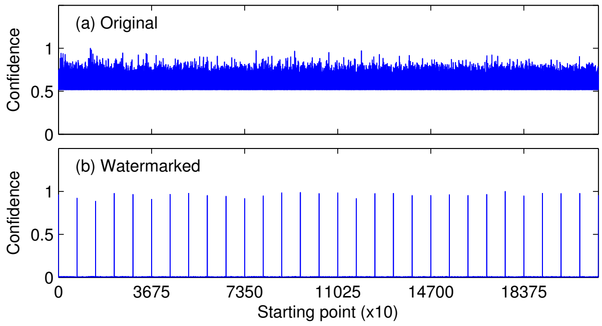

Fig. 3. Frame synchronization in the case frame length = 7350 points: (a) original signal and (b) watermarked signal

$\angle Y_i(\omega)$ is obtained. Although each bit can be embedded in only one component, it is embedded in several components to increase robustness. The bit rate is adjusted by changing the number of components for each bit.

Step 4. The magnitude spectrum, $|X_i(\omega)|$ and the quantized phase spectrum, $\angle Y_i(\omega)$, are combined into Fourier spectrum $Y_i(\omega)$ which is then transformed into time domain signal $y_i[n]$ by inverse Fourier transform (IFFT).

Finally, all the processed frames are combined together to yield a watermarked signal $y[n]$.

2.4. Watermark detection

The detection process also starts with frame segmentation of the watermarked signal, $y[n]$ into frames $y_i[n]$ with the same frame size as in the embedding process. Figure 2(b) shows a block diagram of the process that detects watermark bits from a watermarked frame involving three steps as follows.

Step 1. Watermarked frame $y_i[n]$ is firstly transformed into $Y_i(\omega)$ by FFT. Phase spectrum $\angle Y_i(\omega)$ is calculated.

Step 2. The embedding frequency components and corresponding QIM step sizes are determined as in Step 2 in the embedding process.

Step 3. The embedding components are decoded by (3) to extract all the bits. The output bits, $s_i[\ell]$, are determined by majority decision, e.g., if the number of '0', N_0, are greater the number of '1', N_1, the output is '0'.

These steps are repeated until we reach the final frame.

2.5. Frame synchronization

The detection process works with an assumption that the frame positions are available. In practice, the frame positions might be unavailable, so we have to identify the starting point before detecting watermarks. It is noteworthy that a bit is detected from a watermarked frame by majority decision. The values, N_0 and N_1, are compared to decide the output bit, $s_i[\ell]$. We can see that N_0 much greater than N_1 implies that the probability $P(s_i[\ell] = \text{'0'})$ is much higher. In other words, the confidence that $s_i[\ell]$ is correctly detected is higher. In general, we define the detection confidence of a bit by: $\delta_i[\ell] = \max\left(N_0/N_1, N_1/N_0\right)$.

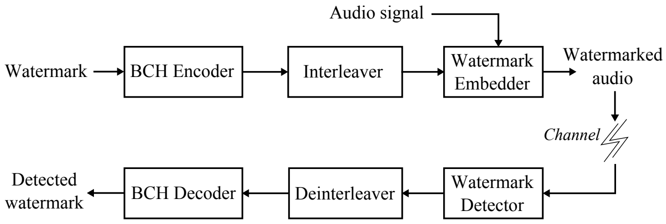

Fig. 4. Proposed framework of audio watermarking in incorporation with error control coding

We can search for a correct frame position over a frame length of signal. It is obvious that if we select the correct frame i, $\sum_\ell \delta_i[\ell]$ is maximized. Figure 3 depicts an illustration of frame synchronization. We calculate detection confidence over 32 frames in two cases: (a) without watermark and (b) with watermark. The detection confidence is normalized to 1. There is no obvious peak in Case (a) while very high peaks occur at the correct frame-starting points in Case (b). The search procedure is performed at the beginning of the detection process. Once the starting point is determined, all the frame positions can be synchronized.

3. INCORPORATION OF ECC TO THE SYSTEM

Figure 4 shows a diagram of the proposed framework of audio watermarking with ECC. The watermark is firstly encoded by a BCH encoder after which certain codewords are obtained. We choose BCH codes because they are binary error-correcting codes with excellent properties [13]. In order to improve the performance of ECC against burst errors, the BCH codewords are interleaved to distribute the errors into different codewords. The interleaved codewords are then embedded into an audio signal. At the receiver side, the watermark detector is firstly used to extract the interleaved codes. Then the extracted codewords are deinterleaved and deinterleaved codewords are finally decoded by BCH decoder. Watermark embedding and watermark detection processes are presented in the previous sections. The next two subsections give descriptions of BCH codes and the interleaving technique.

3.1. BCH codes

BCH codes [14] form a class of parameterized error-correcting codes which have been applied to many applications, such as satellite communications, DVD players, and two-dimensional bar codes. The principal advantages of BCH codes is that they are binary codes with excellent minimum distance properties, and can be decoded via an elegant algebraic method which allows very simple electronic hardware to perform the task. BCH codes are also highly flexible, allowing control over block length and acceptable error thresholds.

A BCH code is represented by (n_0, k_0, t_0) in which n_0 is the code length, k_0 is the data length, and t_0 is the number of

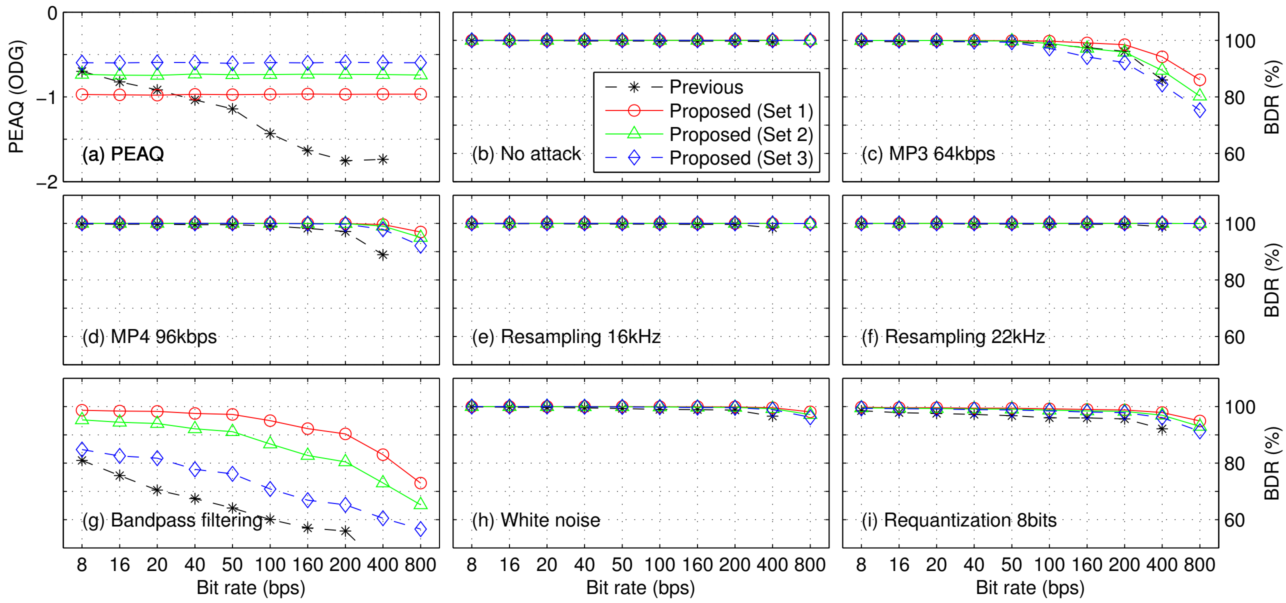

Fig. 5. Sound quality (PEAQ) and bit detection rate (BDR) with respect to bit rate in comparison with the previous method: (a) PEAQ, (b) BDR without attack, and (c)–(i) BDR against attacks

correctable bits [13]. If the number of errors occurring during transmission is less than or equal to t_0, all the errors can be corrected. Otherwise, BCH code fails to correct the errors.

BCH codes add additional parity bits to the data bits, hence when applied to audio watermarking, BCH codes reduce the bit rate of watermark. The higher value of t_0 creates a stronger BCH code, i.e., it can correct more errors. However, k_0 becomes smaller which reduces the actual bit rate of watermark. A suitable BCH code can help improve reliability. In practice, the strategy for section of BCH code's parameters is firstly fixing the criteria for sound quality and bit error rate, then selecting the parameters (n_0, k_0, t_0) so that those criteria could be met and the bit rate is maximized.

3.2. Interleaving

In audio watermarking systems, errors typically occur in bursts rather than independently. Interleaving multiple codewords can be used to improve performance of error correcting codes. If the number of errors due to a burst are greater than the error-correcting code's capacity, the error-correcting code cannot recover the original codeword successfully. Interleaving shuffles the source symbols across several codewords in order to create a more uniform distribution of errors, reducing the effect of burst errors.

4. EVALUATION

Experiments were carried out to evaluate inaudibility and robustness of the proposed method with 102 RWC music tracks [15] that have a sampling frequency of 44.1 kHz and 16-bit quantization. The frame size is set to 500 ms. The FFT size is equal to the frame size and the rectangle window was used. The watermarks were randomly generated. The parameters, K, ϵ, and L, were determined by experimental analysis and set to 1.6 kHz, 10^{-4}, and 5 respectively. The QIM step sizes are chosen as integer divisions of π to reduce wrapping errors. We investigated the proposed method with three sets of five QIM step sizes: Set 1 ($\frac{\pi}{2}$, $\frac{\pi}{4}$, $\frac{\pi}{6}$, $\frac{\pi}{8}$, $\frac{\pi}{10}$), Set 2 ($\frac{\pi}{3}$, $\frac{\pi}{5}$, $\frac{\pi}{7}$, $\frac{\pi}{9}$, $\frac{\pi}{11}$), and Set 3 ($\frac{\pi}{4}$, $\frac{\pi}{6}$, $\frac{\pi}{8}$, $\frac{\pi}{10}$, $\frac{\pi}{12}$).

Inaudibility was tested by perceptual evaluation of audio quality (PEAQ) [16] which rates sound quality by the objective difference grade (ODG) from -4 (very annoying) to 0 (imperceptible). Detection accuracy was measured by bit detection rate (BDR), the ratio between the numbers of correct bits and total bits. Robustness was investigated with the following processing: MP3 64 kbps, MP4 96 kbps, adding white Gaussian noise 36 dB, requantization 8 bits, resampling 22 kHz and 16 kHz, and bandpass filtering with passband [0.1, 6] kHz and stopband attenuation -12 dB/octave.

Figure 5 shows the results of PEAQ and BDR of the proposed method with three sets of QIM step size and in comparison with the previous method [11]. The bit rate was varied from 8 to 800 bps. All the PEAQs are greater than -1 ODG (not annoying) and the sound quality of watermarked signals remains unchanged as the bit rate increases. The sound quality becomes better when the QIM step size decreases from the values in Set 1 to those in Set 3. PEAQs of the previous method decrease when the bit rate increases. The sound quality of the proposed method is almost better than that of the previous method, especially at higher bit rates.

In the cases of no attack and resampling, the BDRs are greater than 99.9% for all the bit rates and do not change much among three sets of QIM step size. In the cases of MP3, MP4, adding white noise, and requantization, the BDRs are greater than 98% for the bit rates less than or equal to 200

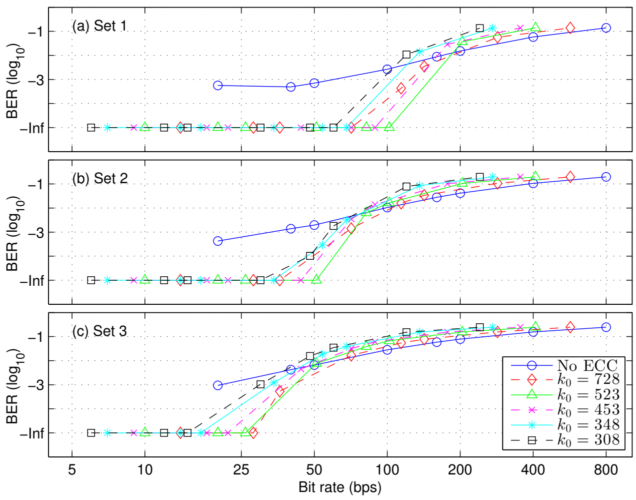

Fig. 6. Bit error rate (BER) after BCH decoding in the case of MP3 attack with respect to embedding bit rate

bps with Set 1 and slightly decrease with Set 2 and Set 3. The bandpass filtering seems to be the strongest attack which makes the BDRs around 90% for the bit rates less than or equal to 200 bps with Set 1 and decrease much more with Set 2 and Set 3. The BDRs of the proposed method with all the sets of QIM step size are almost greater than the previous method except for MP3 with Set 3.

These results suggest that the proposed method is more effective and has better performance compared with the previous method. Moreover, the proposed method provides good sound quality in the watermarked signals and high robustness against most types of processing.

We evaluated the effectiveness of incorporation of ECC into the watermarking system in the case of MP3. We chose to investigate MP3 attack because it is popularly used in practice and is the strongest attack except for bandpass filtering. If ECC can correct errors after MP3 compression, it can also correct errors from the other attacks. The five codes with the length of 1023 and different values of k_0 have been used.

Figure 6 shows the bit-error rate (BER) after BCH decoding, with respect to embedding bit rate with three sets of QIM step size. The results show that the system can extract watermarks without any detection error at a bit rate of 102, 51, and 28 bps with Set 1, Set 2, and Set 3, respectively. Compared with the case that ECC is not used to encode watermarks, the incorporation of ECC is remarkably effective in correcting all the errors at relatively high bit rates.

5. CONCLUSION

We proposed an audio watermarking method based on dynamic phase coding and ECC. Watermarks are embedded into the phase of moderately low frequency components. The QIM step size for each component is adjusted according to the magnitude to balance inaudibility and robustness. BCH coding is applied in encoding watermarks before embedding process to increase reliability. The experimental results verify that the watermarked signals have high sound quality and the embedded watermarks are robust against various attacks. The incorporation of ECC is effective for audio watermarking to carry more reliable watermark in practice.

References

[1] C. I. Podilchuk and E. J. Delp, "Digital watermarking: Algorithms and applications," *IEEE Signal Proc. Magazine*, vol. 18, no. 4, pp. 33–46, 2001.

[2] N. Cvejic and T. Seppänen, *Digital Audio Watermarking Techniques and Technologies*. IGI Global, 2007.

[3] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Trans. Multimedia*, vol. 3, no. 2, pp. 232–241, 2001.

[4] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Elservier Signal Processing*, vol. 66, no. 3, pp. 337–355, 1988.

[5] A. Takahashi, R. Nishimura, and Y. Suzuki, "Multiple watermarks for stereo audio signals using phase-modulation techniques," *IEEE Trans. Signal Proc.*, vol. 53, no. 2, pp. 806–815, 2005.

[6] H. L. F. Helmholtz, *On the Sensations of Tone, 2nd ed.* Dover Publications, 1954.

[7] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Info. Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[8] M. Narimannejad and S. M. Ahadi, "Watermarking of speech signal through phase quantization of sinusoidal model," in *Proc. of Iranian Conf. on Elec. Engineering 2011*, pp. 1–4.

[9] N. Khademi, M. A. Akhaee, S. M. Ahadi, M. Moradi, and A. Kashi, "Audio watermarking based on quantization index modulation in the frequency domain," in *Proc. of Int. Conf. on Signal Proc. and Comm.*, pp. 1127–1130, 2007.

[10] L. Gang, A. N. Akansu, and M. Ramkumar, "MP3 resistant oblivious steganography," in *ICASSP 2001*, pp. 1365–1368.

[11] N. M. Ngo and M. Unoki, "Robust and reliable audio watermarking based on phase coding," in *ICASSP 2015*, pp. 345–349.

[12] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. California Tech. Pub, 1997.

[13] S. Lin and D. J. Costello, *Error Control Coding, 2nd ed.* Prentice Hall, 2004.

[14] R. C. Bosea and D. K. Ray-Chaudhuria, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, no. 1, pp. 68–79, 1960.

[15] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Music genre database and musical instrument sound database," in *Proc. ISMIR 2003*, pp. 229–230, 2003.

[16] P. Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," tech. rep., Dept, Elect. Comp. Eng., 2002.