Vol.2011-CSEC-55 No.8
2011/12/5

A novel content-based tampering detective steganography for acoustic data

Xuping Huang[†1] Akira Nishimura[†2] Isao Echizen[†3]

We propose and implement a content-based tampering detective steganography scheme using acoustic data with probative value in this paper. The purpose is to verify and identify malicious modification. Content-based hash function SHA-1 is adapted to detect tampering. After transforming data from time-domain to frequency domain using integer Discrete Cosine Transform (int-DCT), the amplitude of the highest frequency domain is expanded to reserve embedding capacity which is necessary for hiding hash value and index table for hiding. Lossless embedding and extraction algorithm ensure this scheme a reversible alternative scheme to meet the requirements of acoustic media with probative value. Hash digest is applied to data units after the original data is divided to frames to detect tampering in frame unit and to ensure the reversibility of the rest data even tampering occurs partially. The numerical simulation experiments on detection precise and acoustic degradation indicate that the proposed scheme satisfied highly variability and reversibility, while the acoustic degradation of stego data is imperceptible on the basis of the ITU-R BS.1387 (PEAQ) standard.

†1 The Graduate University for Advanced Studies
†2 Tokyo University of Information Sciences
†3 National Institute of Informatics

1. Introduction

In this paper, we discuss the issues and problems associated with illegal tampering with acoustic evidence using the conventional reversible information hiding scheme. Maintaining the integrity of digital content and verifying whether illegal tampering has occurred are serious issues, especially for acoustic data that must be kept as material evidence for probative purposes. Such data may include police-investigation tapes, last wills and testament tapes, telephone recordings, phone banking records, emergency calls, and air-traffic communications. There are digital signature to authenticate use, however, digital signature is usually appended on the header of the file and while the header was removed, the digital signature is not valid at all. We apply information hiding in this scheme to embed hash value of each divided frames to the high frequency domain of the frame itself, where the distortion is difficult to be realized. There are three requirements for these scenarios: (1) the scheme must be able to determine whether malicious tampering has occurred; (2) the scheme must be reversible since the original is required for probative purposes; and (3) the distortion of the stego data must be controlled to be imperceptible. Many verifiable or reversible watermarking schemes have been proposed in the field of image processing. These methods take advantage of the Human Visual System and cannot be applied to acoustic data, which will cause auditory distortion.

This paper is organized as follows. The approaches taken here and those in the previous studies are introduced in Section 1. Section 2 details the proposed method, including the hash digest, embedding, and extraction processes, and verification of tampering. Section 3 introduces the implementation of acoustic steganography, and Section 4 summarizes the results from an objective experiment. Section 5 concludes this paper.

1.1 Conventional Works and Their Problems

Alternative reversible hiding methods have already been studied[1),2),3),4),5),6)]. Technologies have also been used to verify the integrity of the transmitted speech signal[7)], enhance the security of speaker identification system[8)], and many others. Methods have also been proposed to detect audible modifications or to identify uses[9),10),11)].

Popular reversible hiding policies for integrity verification were classified into two

main classifications by Mehmet et al.[6] i.e., (a) During decoding a spread spectrum signal corresponding to the information payload (embed data) is superimposed onto the host signal (cover data), and during decoding the payload is subtracted from the stego data, and (b) the features of the original data are embedded as the watermark payload. The first classification offers bit robustness while the second offers a higher capacity. In both ways, the watermarked signal is visible (images), audible (acoustic data) and perceptible in the payload.

Algorithms for reversible watermarking can mainly be categorized into three classifications: (1) Data compression based methods[12]: the original portions of the cover data will be replaced by a payload using compression and data alteration, and during decoding, the feature information is extracted and decompressed. (2) Methods based on modifying the difference expansion[13]: these schemes represent the features of the original data using small values, then the value is expanded to embed the payload in the LSB. (3) Histogram bin shifting methods[14]: the embedding target is replaced with the histogram of a block. These methods are used to enhance the robustness of reversible watermarks.

The approach of imperceptible payload hiding has been a difficult issue to conceive conventionally. A distortion-free scheme for embedding has been proposed[15]; however, it cannot be used for integrity verification. According to our investigations, none of the past studies have met the requirements to combine reversible and verifiable approaches for the integrity verification of acoustic data, and there has been a lack of applications to maintain the integrity of acoustic data or verify the authenticity. In the proposed scheme, flexible amplitude expansion and countermeasures against overflow/underflow make it possible to achieve reversibility and verifiability, which are two important issues for protecting the integrity of sensitive acoustic data, when the usual criteria do not apply. In our previous work[16],[17], a reversible and verifiable steganography was proposed, but the detection is functionally to a whole file and frequency is divided to half of low-frequency domain and the other half part of high-frequency domain. In the previous work[17], detection for malicious tampering is accurate to the frame unit, and processing effectives have been improved. The amplitude of data in high frequency domain (half of the window) is expanded by a bit to reserve space for capacity hiding. However, the positions reserved by amplitude expansion is larger than bit amounts which are necessary for verification data and index table. In this work, We apply amplitude expansion only for necessary hiding space for integrity verification, since the purpose of this work is not to create large amount of hiding capacity, but better acoustic quality for integrity verification of the data in frame unit for high detection precise.

1.2 Function Properties

The target application of this scheme is to verify whether malicious modifications have occurred that may lead to unjust accusations. Integrity and reliability can be verified without needing the original data. Probative use requires the digital content to be faithful to the original and be reliable. This means three issues should be considered.

- **Verifiability** The scheme identifies any malicious modifications to the acoustic data and finds the specific domain where tampering has occurred that may lead to unjust accusations to ensure the digital acoustic data are reliable. Integrity and reliability can be verified without original data.
- **Reversibility** The original acoustic data are required by courts and judges for particular purposes. Therefore, the original sound (cover data) recording should be recoverable even after the feature data have been extracted from the stego data. Reversibility requires both the embedding and extracting algorithms to be lossless.
- **Imperceptibility** When the feature data are camouflaged in the cover sound, distortion needs to be kept below a perceptible level to keep the stego data audible without needing additional data processing.

This paper explains the general principles behind lossless embedding and reversible and imperceptible schemes to verify whether tampering has occurred. We extract the hash data as the authentication information and embed the hash and payload into the high-frequency spectrum domain after computing the integer discrete cosine transform (intDCT) in the segmented host audio data. Prior to embedding, an amplitude expansion is applied to the DCT coefficients to achieve a totally reversible steganography scheme.

2. Proposed method

The feature data (hash) are calculated from the original data and embedded in a

redundant space in the high frequency domain of the cover data after the intDCT and amplitude expansion. There are three processing phases involved with this process. Figures 2, and 3 are block diagrams of the proposed model. All the processes are repeatedly applied to seamlessly segmented audio signals of length N. N can be set to 256, 512, 1024, 2048, etc. Figure 1 shows the preparation processing necessary to obtain the feature value from each frame of the original data.

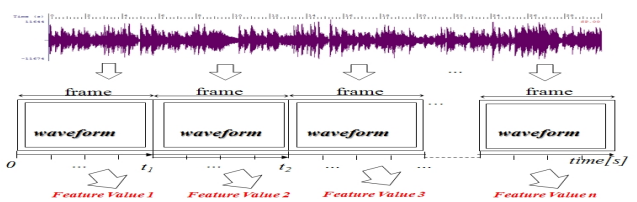


図 1 Frame division and feature value extraction

2.1 Embedding phase

The data to be embedded consists of the hash data of this frame, the index data for embedding. The embedding phase is conducted in the frequency domain calculated with the intDCT of the host signal. The space necessary for hiding is reserved from the highest frequency DCT coefficients after data expansion by taking advantage of the auditory feature where the data changes in the high frequency domain are difficult to be distinguished. The frequency spectrum is converted into waveform data by using an inverse intDCT to obtain the stego signal. If an amplitude overflow or underflow of the stego signal occurs, the amplitude estimation process prior to the inverse intDCT is repeated until the overflow and underflow no longer occurs. The details are given in subsection3.1. Figure 2 outlines the embedding phase.

2.2 Extraction phase

The input stego data are converted by computing the intDCT in the extraction phase. The output data are the re-extracted feature value, the re-constructed original host signal, and the extra payload data. The details are given in Subsection 3.2.

2.3 Verification phase

The hash value is detected in the verification phase from the reconstructed original

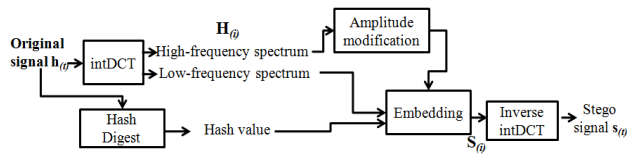


図 2 Structure of embedding phase

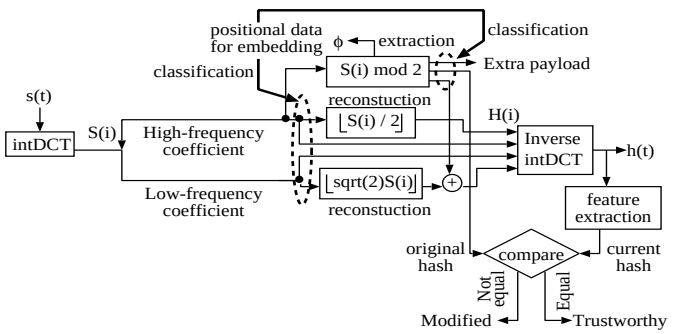


図 3 Structure of extraction and verification phases

data. The current hash value is compared to the original hash value. If they differ, the stego data are modified. Figure 3 outlines the extraction and verification phases.

3. Implementation

3.1 Process flow for embedding phase

3.1.1 intDCT

The intDCT Type IV algorithm proposed by Haibin et al.[?] is used for time-frequency data conversion. There are a total of $2.5N$ rounding operations, where N is the DCT size (the length of a block).

Let $h(t)$ (t=0,1,. . . , $N-1$) be a real-valued input sequence (host signal waveform). We assume that $N = 2^p$, where p is a positive integer. The length N of the type-IV DCT of $h(t)$ is defined as

Vol.2011-CSEC-55 No.8
2011/12/5

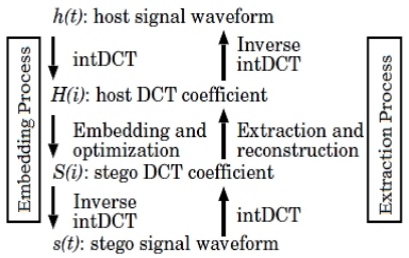


図 4 Signal representations. t is discrete time; $0 < t \leq N$. i is discrete frequency; $0 < i \leq N$.

$$H(i) = \sum_{t=0}^{N-1} h(t) \cos \frac{\pi(2t+1)(2i+1)}{4N}, i = 0, 1, \ldots, N-1 \quad (1)$$

Let C_N^{IV} be the corresponding transform matrix, i.e.

$$C_N^{IV} = \left(\cos \frac{\pi(2t+1)(2i+1)}{4N} \right), \quad i, t = 0, 1, \ldots, N-1 \quad (2)$$

3.1.2 Amplitude Expansion for Embedding

First, the DCT coefficients $H(i)$ are divided into M ($M = 16$ in the current implementation) frequency regions. The power levels are then calculated for each frequency region $p(m)(1 \leq m \leq M)$. Index data for hiding $\vec{\varphi}$ with a length of M bits correspond to the location of the frequency region, i.e., $\vec{\varphi}(m)$ indicates the frequency region of the DCT coefficients $H(i)(i = (m-1)N/M + 1, ..., mN/M)$. Each bit value of $\vec{\varphi}$ classifies the frequency regions to be used for embedding. The last 176 bits indicated from the highest frequency regions for hiding data. $\vec{\varphi}$ is initialized as:

$$\vec{\varphi}(m) = \begin{cases} 1 & \text{if } m \leq M/2; \text{no manipulation} \\ 0 & \text{available for embedding} \end{cases} \quad (3)$$

The high-frequency DCT coefficients, which are indicated by $\vec{\varphi}(m) = 1$, are expanded by doubling to reserve the embedding space. DCT coefficients of the expanded frequency regions are replaced by embedding data. Figure 5 contains the details on the embedding and amplitude expansion.

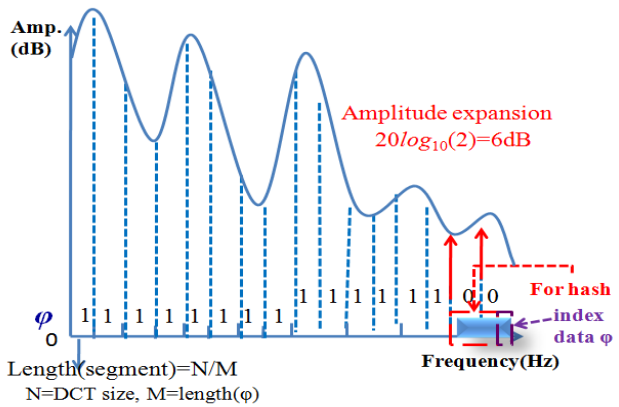


図 5 Amplitude expansion: highest frequency region is reserved to embed index data $\vec{\varphi}$, with length equals $M = 16$. The embedding region is divided into two parts: hash data (length=160 bits), and index table for hiding $\vec{\varphi}$(length $= M$);

3.1.3 Overflow countermeasure

We have to be careful when applying inverse intDCT to the expanded DCT coefficients $S(i)$ of intense signal segments because the amplitude of the stego waveform $s(t)$ can possibly overflow or underflow, which means the amplitude of the waveform exceeds the upper bound (32767 for 16-bit signed audio and 65535 for unsigned) or the lower bound (-32768 for 16-bit signed audio and 0 for unsigned). Yan et al.[2] used a location-map technique to prevent overflow/underflow problems.

The overflow countermeasure in the proposed method is to estimate the amplitude result after expansion. Overflow is due to amplitude expansion, if overflow or underflow is supposed to occur in $s(t)$, the index of the embedding (expansion) start index is shifted to the right by a length of one segment N/M until the overflow no longer occurs. No hiding in a frame where the reserved length is not enough for length of φ and the hash value.

3.1.4 Hash Data and Payload Data

The payload embed region is reserved due to the amplitude expansion at a value of $20log_{10}(2) \approx 6$dB. We used the SHA-1 standard to extract the hash code of the original

data at a length of 160 bits. The hash function plays an important role in authentication schemes to verify the originality and authenticity of recordings in typical practical applications.

The highest frequency region is reserved to embed the index data ($|\vec{\varphi}|$; M bits), the hash data (160 bits). The frequency regions that are indicated by $\vec{\varphi}(m) = 0$ are available to store the payload data.

3.2 Extraction and reconstruction of host waveform

Our method focuses on the high-frequency spectrum domain. As seen in Figure 3, the input datum is the stego signal, and the output data are the re-constructed original signal and the detected feature value. After intDCT is applied to the stego signal $s(t)$, we obtain the intDCT coefficients $S(i)$.

The embedded data are extracted by applying modulo 2 to $S(i)$. The extracted and de-scrambled $|\vec{\varphi}|$ data in the highest frequency region indicate which frequency region has been modified by embedding hidden data into it. The hash data are also extracted and de-scrambled from the same highest frequency region. The DCT coefficients of the high frequency region with hidden data where $|\vec{\varphi}(m)| = 0$ are divided by two to recover the original host spectrum $H(i)$. The reconstructed $h(t)$ is obtained by performing an inverse intDCT.

3.2.1 Verification

We detect the feature value of the reconstructed original data, and compare it with the re-extracted feature value in the verification phase. The extracted feature value (original hash) is compared with the feature value obtained from the reconstructed host signal. If they are not equal, the stego data are untrustworthy, i.e., they may have been modified by a third party. The main disadvantage is that the signals are fragile and authentication fails when there are any attempts to modify the stego data, including regular modifications, such as sampling size conversions, compression, and re-sampling.

4. Experimental Evaluation

We evaluated our method on the RWC Music Database[18]. We did an experiment with an L-channel waveform with a 44.1-kHz sampling and 16-bit quantization. The samples were cut to the initial 30 seconds of playback time.

4.1 Verifiability

We marked 160 bits of the feature data extracted from the stego signal as the original *hash*, and we calculated the hash value of the reconstructed host data, *hash'*. We modified the stego data using Hex Editor, and after the extraction process, the system determined $hash \neq hash'$, which means that the modified data were verified as not being trustworthy. Detection for malicious tampering is accurate to the frame unit. The precision for verification depends on the DCT size N. The precision rate is listed in Table 1 for a signal with 1323008 samples with different DCT sizes. As shown in Table 1, the smaller the DCT size is, meaning the detection is more accurate for a smaller time block unit, the higher the precision rate is. With a 2048 DCT size, the verification can be accurate to a 0.046 seconds sample block.

表 1 Detection precision rate

DCT size	Number of Frames	Precision Time block
512	2584	0.012 sec
1024	1292	0.023 sec
2048	646	0.046 sec

However, the risk that distortion is perceptible is higher when the hiding process is a small DCT size because the amplitude expansion in a high-frequency domain greatly affects the sound quality. An experimental test on the ODG value with different DCT sizes using tracks with the worst ODG values with size of 2048 and half-frequency expansion is shown in Figure 6. When feature value is appended with a larger frame size, the distortion can be imperceptibly controlled, though the detection precise is lower.

4.2 Reversibility

We compared the reconstructed host data with the original data. The difference between them was 0, which meant that our method was reversible. One hundred tracks were used and the original cover data are reversed.

4.3 Imperceptibility

We evaluated the distortion in the acoustic quality on the basis of the ITU-R BS.1387 (PEAQ) standard[19].

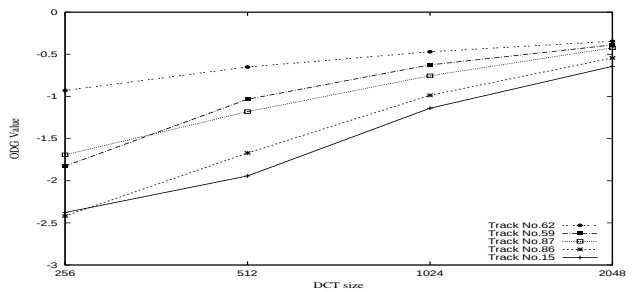


図 6 ODG with different DCT sizes

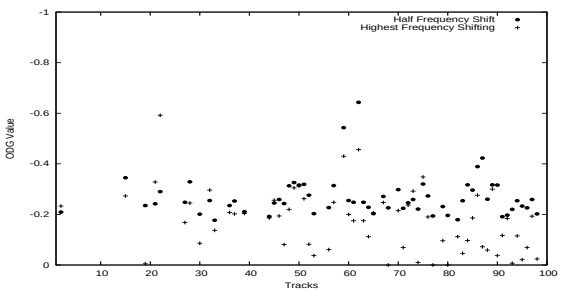


図 7 Comparison of ODG of 61 Tracks between half frequency shift and least shift, N=2048

The impairment scale, an objective difference grade (ODG), ranged from 0 to -4 and could be interpreted as: *0* imperceptible, *-1* perceptible but not annoying, *-2* slightly annoying; *-3* annoying, and *-4* very annoying.

The acoustic data were re-sampled at 48 kHz before the evaluation. The experiments[⋆1] were for the total payloads including the hash and index data for hiding is 103360 when frame size is 2048. We used 61 tracks [⋆2]with top worst ODG value in our previous work[17] for the evaluation, and the ODG results are presented in Figure 7; the average ODG was -0.174, while the value applied with shifting high-frequency domain (half of the window) is -0.267.

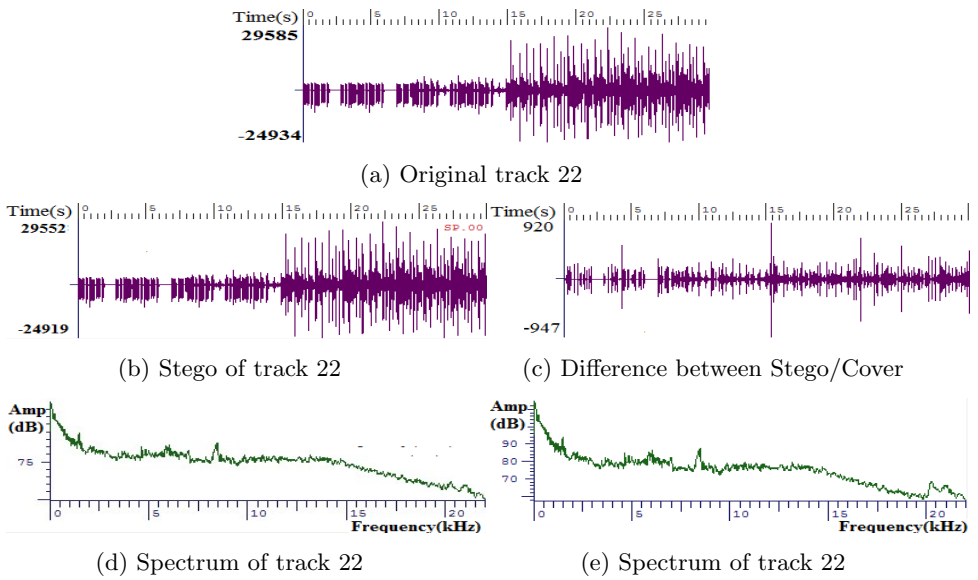
The ODG results are between 0 and -0.5, which seem to indicate good audio-quality after embedding. In some tracks with bad ODG, the he annoying loss of quality was caused by discontinuities at the boundaries between the waveform segments in the DCT operation. Figure 7 presents the different results obtained when using half-frequency shift in work[17] and when using the proposed steganography technique. Decreasing the number of embedding frequency regions, i.e., reducing the extra payload data, may reduce the degradation in sound quality. Among the 61 tracks, track 22 had the worst ODG -0.592; although, the distortion is imperceptible according to the results from a subjective experiment. Figure 8 presents the original data track 22, stego data after embedding with hash information, and the difference between the stego data and cover data in a time domain. The Y-axis in Figure 8(c) is zoomed in on to detect any small changes. According to Figure 8(d) and Figure 8(e), we can find the highest spectrum in track 22 has been changed due to the hidden information.

We calculated the SegSNR results obtained from using the proposed method when using track 22, with with a 30 seconds playback time and a DCT size of $N = 2048$. The segSNR value was 46.108 dB.

5. Conclusion

We proposed a reversible approach that uses steganography to verify whether the acoustic data was free of tampering. A hash function was used to extract the feature value of divided frame of the original cover data, and this feature value was used as the payload for verifying whether or not tampering had occurred. The amplitude expansions were concentrated on in the highest frequency spectrum domain for necessary capacity to make the hidden data less perceptible. This scheme is a reversible way of guaranteeing the integrity and reliability of data without the need to apply additional conversion to the object data. Detection for information addition and deletion is listed to be the future work.

⋆1 We used RWC-MDB-G2001 No.10 with a 30 sec playback time
⋆2 Samples at 44.1 kHz, and 16 bits, in mono with 30 sec of playback time (1323008 samples) and N=2048



(a) Original track 22

(b) Stego of track 22

(c) Difference between Stego/Cover

(d) Spectrum of track 22

(e) Spectrum of track 22

図 8 Imperceptible experimental results

参 考 文 献

1) Vander,V.M., Leest,A.V., and Bruekers, F.,"Reversible Audio Watermarking", Audio Engineering Society, pp.5818, 2003
2) Yan,D.Q., and Wang, R.D., "Reversible Data Hiding for Audio Based on Prediction Error Expansion", Intelligent Information Hiding and Multimedia Signal Processing, pp. 249–252, China, 2008
3) Gomez,E., Cano,P., Gomes,L.D., Batlle,E., and Bonnet,M., "Mixed watermarking fingerprinting approach for integrity verification of audio recordings", International Telecommunications Symposium ITS, Brazil, 2002
4) Gulbis, M., Muller, E., and Steinebach, M.:Audio integrity protection and falsification estimation by embedding multiple watermarks,International Conference on Intelligent Information,2006
5) Mih,M.K., and Venkatesan,R., "A perceptual audio hashing algorithm: A tool for robust audio identification and information hiding", Moskowitz, I.S. (ed.) IH 2001. LNCS, vol.2137, pp.51,Springer, Heidelberg, 2001
6) Celik,M.U., Sharma,G., Tekalp,A.M., and Saber,E., "Reversible Data Hiding", Proc. of International Conference on Image Processing, Vol.2, pp.157–160, USA, 2002
7) Chen,O.T.C., and Liu,C.H., "Content-dependent watermarking scheme in compressed speech with identifying manner and location of attacks", Audio, Speech, and Language Processing, IEEE Transactions on Vol.15(5), pp:1605–1616, 2007
8) Faundez,Z.M., Hagmuller,M., and Kubin,G., "Speaker verification security improvement by means of speech watermarking", Speech Communication, Vol.48(12), pp:1608–1619, 2006
9) Radhakrishnan,R., andMemon, N.D., "Audio content authentication based on psychoacoustic model" E.J. Delp, and P.W. Wong (eds)., Proc. SPIE, Security and Watermarking of Multimedia Contents IV, vol. 4675, pp. 110–117, 2002
10) Zmudzinski,S., and Steinebach,M., "Perception-Based Audio Authentication Watermarking in the Time-Frequency Domain", Information Hiding Workshop LNCS vol.5806, pp.146–160, 2009
11) Kalker,T., Haitsma,J.A., and Oostveen,J.C., "Robust audio hashing for content identification", Content based multimedia indexing (CBMI), pp. 2091–2094, Italy, 2001
12) Celik, M.U., Sharma, G., Tekalp, A.M., and Saber, E.: Localized lossless authentication watermark (LAW), International Society for Optical Engineering, Vol.5020, pp.689–698, USA, 2003
13) Tian, J., "Reversible data embedding using a difference expansion", IEEE Transactions on Circuits Systems and Video Technology, Vol.13, no.8, pp.890-896, 2003
14) Chang, C.C., Tai, W. L., and Lin, M.H.: A reversible data hiding scheme with modified side match vector quantization, Proceedings of the International Conference on Advanced Information Networking and Applications, Vol.1, pp.947–952 Taiwan, 2005
15) Goljan.M., Fridrich,J., and Du,R., "Distortion-free Data Embedding, Proceedings of 4th Information Hiding Workshop", Vol.2137, pp.27–41. Pittsburgh, 2001
16) Huang, X.P., Echizen, I., and Nishimura, A., "A New Approach of Reversible Acoustic Steganography for Tampering Detection", IIHMSP 2010 (10.1109/IIHMSP.2010.137), pp.538–542, Germany, 2010
17) Huang, X.P., Echizen, I., and Nishimura, A.,"A Reversible Acoustic Steganography Scheme to Authenticate Use", Kim,H.J., Shi.Y., and Barni.M. (Eds.): IWDW 2010 LNCS 6526(Springer-Verlag Berlin Heidelberg 2011), pp.305-316, Oct.2010
18) Goto,M., Hashiguchi,H.,Nishimura, T., and Oka,R., "RWC Music Database: Music Genre Database and Musical Instrument Sound Database", Proceedings of the

Vol.2011-CSEC-55 No.8
2011/12/5

4th International Conference on Music Information Retrieval (ISMIR), pp. 229–230, 2003

19) Kabal, P.,"An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality", TSP Lab Technical Report, Dept. Electrical, Computer Engineering, pp.1–89, 2002

20) Kobayasi,T., Itahashi,S. ,Hayamizu,S. and Takezawa,T., "ASJ continuous speech corpus for research", The Journal of the Acoustical Society of Japan, vol. 48(12), pp.888-893, 1992