Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.
Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-21

A Reversible and Imperceptible Acoustic Watermarking Using Partially-applied Huffman Lossless Compression

Xuping Huang

Abstract— **In this paper, a reversible and robust acoustic watermarking based on lossless compression is proposed. Stereo speech data is represented by 16 bit for each sampling point and then divided into frames. Then Huffman lossless compression algorithm is applied to insignificant 4 bits in each sampling point partiality to reserve hiding capacity. An average of 0.7035 and the best 0.686 compression ratio depending on different frame lengths are achieved, which promises about 1.188 bits for payload hiding in each sampling point. Since Huffman algorithm is applied partially to each sampling point, stego data is comprehensive after embedding and complexity towards attack is promising. Result of Perceptual Evaluation of Speech Quality based on ITU-T recommendation P.862 and signal-noise ratio (SNR) show the proposed method achieved imperceptibility.**

Keywords: acoustic watermarking, lossless compression, reversibility, imperceptibility

I. Introduction

Reversible watermarking has been proposed to protect the integrity of data with probative importance or personal data with privacy vulnerability as an alternative authentication method recently. It can be applicable to many kinds of scenarios, for example: military images, medical operation video and investigation recordings, which are with legal issues involved.

A. *Reversible Watermarking*

In these scenarios, the reversibility of original data is essential due to the probative importance of the data. If algorithm is loss, original data cannot be reconstructed from stego data. Any data loss is not allowed, even though the loss might be slight for few samples. Since for applications where availability of original data is important, which may be used as the evidence materials in the court, according to the law, the data is not reliable and cannot be used any more even though the loss of samples did not affect to comprehensiveness of the content of itself.

Examples include military materials, medical recordings and so on. Besides, in some cases, it is difficult to reproduce the content again, which is with great value as the documentary in some certain environment and should be protected without any distortion or data loss.

In military and medical cases, the recorded data subjected to certain situation and cannot be re-provided but are important for strategy and medical treatment. The importance of reversibility means that the algorithms to hide payload, including information for verification and positional data, extract payload, and reconstruct original data should be lossless.

B. *Conventional hiding method and proposed method*

For hiding algorithm, two main algorithms are used to hide audio information, mainly into frequency domain and time domain: (1) histogram-based watermarking [1][2], payload is embedded into expanded components in frequency domain; (2) linear predictive coding based watermarking [3][4]: payload is embedded into expanded residual between predicted value and the real value; and (3) compression-based watermarking [5], in which the payload is hidden in the capacity reserved by the compression algorithm. The challenge of information hiding by compression is that the compression rate of an audio file is quite limited with the existing algorithms: MP3, Huffman, MonkeyAudio, TTA, LPAC, TAK, WavPack, etc. Some algorithms use a proprietary format (MonkeyAudio), making them unsuitable for public usage. Furthermore, stego data are not audible without proper decoding because of the proprietary format in many information hiding methods that combine compression and package loss makes it irreversible.

Thus, in this paper, being different from the conventional compression method, we represent each sampling data into 16 bit and apply lossless compression algorithm Huffman to less significant bits to reserve capacity to hide payload. Data is divided into frames before processes. For the target application, authenticate use is one of the applications, since integrity of an audio signal is determined by detecting changes in its feature value. Reversible information hiding for authentication use on image and video has been proposed [5-9]. In these works, fingerprint information [6], time sequence information [7], hash values [5, 8], and binary numbers [9] are widely used to verify the integrity of original data. In this paper, a

Xuping Huang
School of Applied Information Technology
The Kyoto College of Graduate Studies for Informatics (KCGI)
7 Tanaka-Monzencho, Sakyo-ku, Kyoto, Japan, 606-8225

Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.
Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-21

content-based RSA digital signature digest value of each frame is calculated and then is embedded into the frame as payload.

As experimental evaluation, objective testing based on Perceptual Evaluation of Speech Quality (PESQ), ITU-T recommendations showed the implementation of the proposed method is imperceptive that the distortion is tolerated and acceptable.

In this paper, the details of proposed algorithm are described in Section 2. The experimental results and evaluation on audio quality are described in Section 3, and the conclusion and future work are described in Section 4.

II. Proposed audio watermarking based-on Huffman Compression

Our proposed reversible method for acoustic data is achieved by applying Huffman algorithm partially into less significant bits of original data. The illustration of frame division, compression, and embedding processes are shown in Fig.1. The S'$_{fk}$= *Huffman(cover[index])*, *index*∈[1,4] means the data after compression has applied to data from the least significant bit (lsb) to the 4th bit. *a* means the digital signature of the previous frame, which is to be embedded in the space reserved by compression.

Since the acoustic data should be usable as probative data, the challenge is to generate audible stego data without additional conversion after partial compression has been applied to the sampling points and payload embedded. Furthermore, the acoustic distortion should be kept imperceptible in accordance with the compression ratio, the embedding algorithm, and the capacity reserved by compression.

The stream of cover data is first divided into small fixed-sized frames. The tampering detection rate depends on the frame length. The frame length *N* can be 65,536, 32,768, 16,384, …, 2,048, 1024, etc. The shorter the frame is, the more precise the detection of modification positions. Frame f_k, k∈ [*1,N*] is represented by sampling points in time domain represented by 32kHz-16bit-stereo, which means each sampling is represented by 16 bits.

The compression rate applied to acoustic data is lower than that applied to images if the algorithm is applied to the file because of data correlation. We thus partially apply the compression algorithm to the [*1*st, *index*th] bit of each sampling point after frame division and sample computing for *index*∈[2,4]. The 1st bit means the least significant bit at each sampling point.

A. Data structure after embedding

Data after embedding consists of three types of data by taking the *k*th frame as an example: (1) data in significant bits: S$_{fk\_msb}$ = *cover_data[index]*, *index*∈[5,16]; (2) data after partially applied compression: object data for compression is S$_{fk\_lsb}$ = *cover_data[index]*, *index*∈[1,4], which is collected in a data pool and then compressed, and S$_{fk\_lsb}$ = *Huffman(cover[LSB, 4th])*; and (3) payload: signal a_k is the digital signature for the *k*th frame and a_k *=(hash*$_k$ || *private_key (d,n))*, which is the signal for verification to be embedded in the capacity reserved by the compression algorithm. Here, sign () is a modulo calculation with private key pair *(d, n)*.

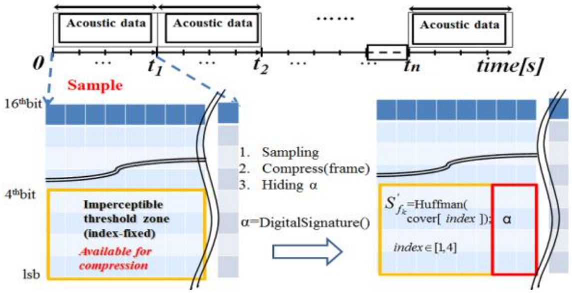

Figure 1. Illustration of proposed method with partially adaptive compression

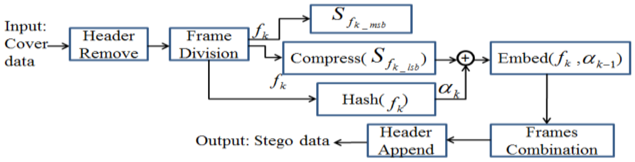

Figure 2. Embedding process flow

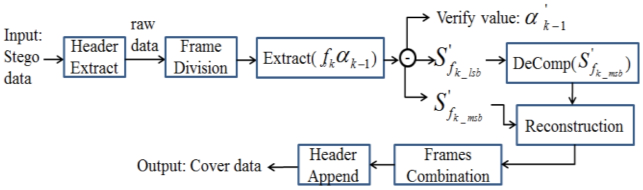

Figure 3. Extraction process flow

B. Implementation

(1) Process flow for embedding

There are four steps for embedding phase:

Step 1 Remove header information from input audio signal, divide signal into fixed-size frames, and pulse code modulation sampling.

Step 2 Calculate verification value using hash function and digital signature algorithm

Step 3 Select cover data from *LSB* to 4th bits for each sampling point in data pool for each frame.

Step 4 Compress data pool generated by step (3) using Huffman algorithm and write verification value a_k in reserved capacity after compression to generate stego data

The embedding process flow is shown in Fig.2.

Since the header information of cover WAVE data is not the target for compression or embedding capacity, the header information is extracted before processing. After compression and embedding, the header information is appended to frames after frame combination to enable audible stego data to be generated.

The capacity reserved by compression is estimated using: *t/b*compR*f*time*, where *(a)* the compression ratio is: *compR*, which is calculated and plotted in Fig.4 and Fig.5. *(b)* threshold index for compression is *t*; *(c)* sampling frequency is *f*; *(d)* the number of sampling bits is *b*, and *(e)* the time length of the sampled frame is *time*.

Since Huffman algorithm needs to append a table to indicate the tree, with which decompression is achievable, it occurs that the output data is larger than input data if the compression ratio is too inefficient to reserve any space in a frame. In this case, the frame is skipped without any embedding.

(2) Process flow for extraction and reconstruction of cover data

The input of the extraction phase is stego data, and the output is reconstructed cover data. In this phase, the extraction algorithm is used to extract the embedded verification data a'_k, and the partially compressed data S'_{fk_lsb} and S'_{fk_msb}. S'_{fk_lsb} is then decompressed and combined with S'_{fk_msb} to reconstruct the cover data.

The extraction process flow is shown in Fig. 3.

(3) Process Flow for Verification

The input data for verification are the *(a):* reconstructed cover data, and *(b):* a'_k. Since payload is content-based, verification can be finished only with stego data, which makes this method a cover-blind verifiable method.

The verification phase is based on the digital signatures and includes the signing and verification procedures:

Step 1 Since the compression, embedding and extraction are lossless, in theory, the cover data is reversible; the reconstructed data is *DeComp*(S'_{fk_lsb})+S'_{fk_msb}. Function *Decomp()* means decoding of compression. If there has been without modification, the result should match the original cover data, thereby proving that this method is a reversible solution.

Step 2 After applying decompression algorithm to data S'_{fk_lsb} and computing the hash function of the reconstructed data, the result *Hash (DeComp*(S'_{fk_lsb})+ S'_{fk_msb}*)* is used to compare with the digital signature value extracted from the frame and thereby verify the content integrity.

Step 3 From the extracted a'_k, we can extracted the hash value h_k by using the decoding calculation $a'^{e}_{k}(\mathrm{mod}d)$ with the public key pair *(e, n)*. The last step for verifying the integrity of the stego data is to compare the result from step 2 with $a'^{e}_{k}(\mathrm{mod}d)$.

To verify the integrity of the target acoustic data with signature, the signing and verification transformations must satisfy if and only if *Hash* (*DeComp*(S'_{fk_lsb})+ S'_{fk_msb}) = $a'^{e}_{k}(\mathrm{mod}d)$.

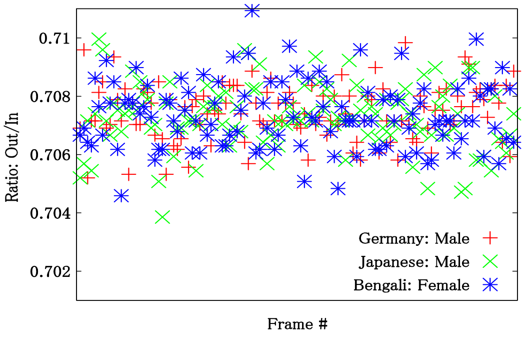

Figure 4. Compression ratio with different language and gender when frame length is 32768

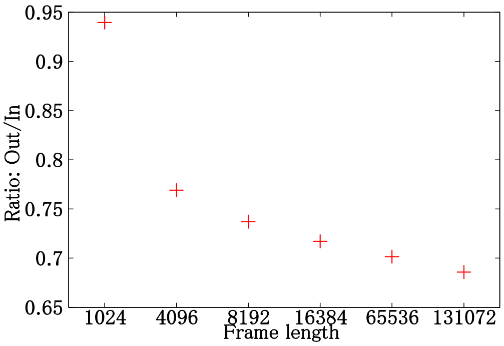

Figure 5. Compression ratio with different frame lengths

III. Evaluation

A. *Dataset for test*

Since the target application can be used to authenticate the integrity of speech data, we record speech data with cooperators who speak in Germany (male), Japanese (male) and Bengali (female) to analyze whether language and gender effect on compression ratio. Samples are recorded by stereo microphone audio-technica AT9941 by 32kHz-16bit-stereo.

B. *Compression Ratio*

Compression ratio is calculated by compare size of output data (after compression) and input data (original data) by out/in to estimate the capacity for payload. The data is represented by 32 kHz-16bit-stereo. The compression ratios are plotted in Fig.4 and Fig.5. According to Fig.4, gender and linguistic difference seldom effect on compression ratio and the average ratio is around 0.703 when frame length is 32768 (a middle length is selected to check the average value), which means 4bits*(1-0.703) = 1.188 bits are available for hiding capacity in each sampling point. According to Fig. 5, it is obvious that longer frame size may

Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.
Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-21

have better compression ratio, while when frame length 1024, the compression ratio is the worst, since a table with about 100 bits takes up a high rate in each frame.

C. *Audio quality*

Fig. 6 plots the amplitude of original data and stego data, and Fig.7 plots the difference between stego data and original data in both of DCT domain and time domain with the sample of Germany male speaker in 958464 samples (32kHz-16bit-stereo, frame size=32768).

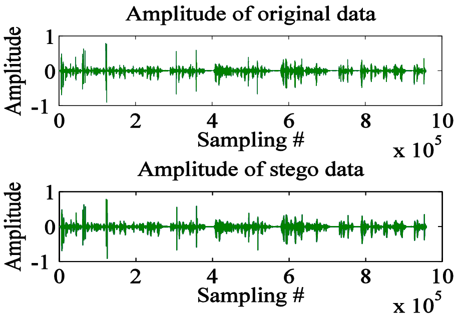

Figure 6. Ampliutude of original data and stego data (Germany Male)

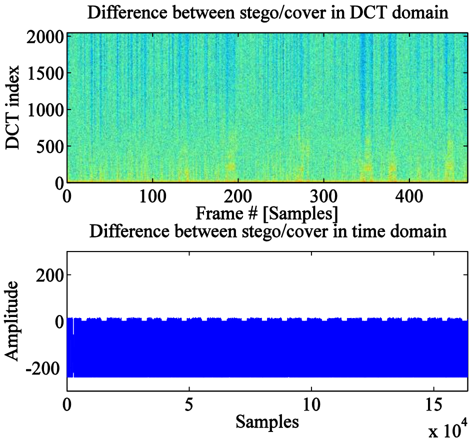

Figure 7. Difference between stego data and original data (Germany Male) in DCT domain (log of absolute DCT value) and in time domain

The difference is time domain is between [-240, 15] in both of channels in this sample. The difference is small out of 32768 as maximum amplitude. According to Figures 6 and 7, it is obvious that the difference between stego data and original data is small to be distinguished by human hearing system. Bengali (Female) has the similar distribution of difference data.

To evaluate audio quality objectively, we calculated Perceptual Evaluation of Speech Quality based on ITU-T recommendation P.862 with raw mean opinion scores (MOS) in the range -0.5 to 4.5 (best). Signal-noise ratio (SNR) is used to calculate the difference in time domain. AFsp packages version 9.0 is used for calculation. MOS and SNR listed in Table 1.

TABLE I. MEAN OPINION SCORES AND SNR

Language	*MOS*	*SNR(dB)*	*Instruction*
Bengali	2.918	21.558	Female, 32kHz-16bit-stereo
Germany	2.690	20.455	Male, 32kHz-16bit-stereo
Japanese	2.924	23.705	Male, 32kHz-16bit-stereo

IV. Conclusion

In this paper, we propose a reversible speech watermarking with partially applied Huffman compression algorithm as an alternative method. About 1.188 bits in each sampling point is available for hiding that high capacity for embedding is achievable by the proposed method. Since the compression ratio is each frame is complexity is achievable. By analyzing amplitude of stego data and original data and their difference with audio quality evaluation by MOS, we got the result that the stego data generated by the proposed method is imperceptible.

As the future work, we are going to evaluate audio quality with more samples objectively and subjectively, and to calculate the robustness towards Stirmark Benchmark.

References

[1] S. Zmudzinski, and M. Steinebach, “Perception-Based Audio Authentication Watermarking in the Time-Frequency Domain, Information Hiding Workshop LNCS vol.5806, pp.146–160, (2009)

[2] X.P. Huang, I.Echizen., and A, Nishimura, “A Reversible Acoustic Steganography Scheme to Authenticate Use, H.-J. Kim, Y. Shi, and M. Barni (Eds.): digital watermarking”, Lecture Notes in Computer Science LNCS 6526 (Springer-Verlag Berlin Heidelberg, vol (2011), pp: 305-316, (2011)

[3] D. Q. Yan, and R. D. Wang, “Reversible Data Hiding for Audio Based on Prediction Error Expansion”, Proc. of Intelligent Information Hiding and Multimedia Signal Processing, pp. 249–252, (2008)

[4] A. Nishimura, “Reversible Audio Data Hiding Using Linear Prediction and Error Expansion”, Proc. of Intelligent Information Hiding and Multimedia Signal Processing, pp: 318–321, (2011)

[5] M. Kaur, R. Kaur, “Reversible watermarking of medical images: authentication and recovery-A survey”, Journal of Information and Operations Management, vol (3)-1, pp.241-244, (2012)

[6] E. Gomez, P. Cano, L.D. Gomes, “Batlle, and E., Bonnet, M.: Mixed watermarking fingerprinting approach for integrity verification of audio recordings”, International Telecommunications Symposium ITS 2002, Brazil (2002)

[7] I. Echizen, T. Yamada, S. Tezuka, S. Singh, and H. Yoshiura, “Improved video verification method using digital watermarking”, Proc. Intelligent Information Hiding and Multimedia Signal Processing, pp.445-448, (2006)

[8]R. Caldelli, F. Filippini, R. Becarelli, “Reversible watermarking techniques: an overview and a classification”, EURASIP Journal on Information Security, vol.2010, pp: 1-19, (2010)

[9] X. Zeng, Z. Y. Chen, M. Chen, Z. Xiong, “Reversible video watermarking using motion estimation and prediction error expansion”, Journal of Information science and engineering, vol. (27), pp: 465-479, (2011)

Proc. of the Second Intl. Conf. on Advances In Computing, Communication and Information Technology- CCIT 2014.
Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-051-4 doi: 10.15224/ 978-1-63248-051-4-21

About Author (s):

She received B.S. and B.A. degrees from the Department of Software Science, Dalian JiaoTong University, China in 2007 and an M.S degree from the Department of Information Science, Graduate School of Iwate Prefectural University, Japan in 2009. From 2009 to 2014, she was a Ph.D. candidate at the Graduate University for Advanced Studies (SOKENDAI) and a research assistant at the National Institute of Informatics, Japan. She is now the Assistant Professor of The Kyoto College of Graduate School for Informatics. Her research interests include information hiding and audio signal processing. She is a member of IEICE and IPSJ.