MP3-Resistant Audio-Adaptive Steganography

Paul Bao, *Senior Member IEEE* and Xiaohu Ma
Nanyang Technological University

Abstract- In this paper, we propose a novel audio steganography scheme for embedding high-capacity covert data in a music carrier, where the carrier is first transformed to a 2D arrangement (image) and represented by a wavelet domain singular value decomposition (SVD), and a quantization-index-modulation process is then applied on the SVD for the covert data embedding. The proposed scheme, due to its indirect embedding on the singular values in the visual domain, possesses a high and flexible embedding capacity and an excellent MP3 robustness while retaining good inaudibility. The MP3 robustness and the imperceptibility are further enhanced by adaptively modeling the quantization parameters based on the statistics within subbands.

1. Introduction

Image steganography has been extensively explored [1,2] and schemes based on bit modification of pixels and coefficients insensitive under the various psychovisual and statistical models such as spread spectrum, generalized Gaussian in wavelet coefficients, qualified significant wavelet trees, variable-size model, Quantization Index Modulation have been presented. All the schemes are characterized with high embedding capacity (large payload), excellent imperceptibility and oblivious and accurate extraction, thanks to the low sensitivity of the human visual system (HVS) to luminance.

However compared to image steganography, the relatively less success in audio steganography with high MP3 resistance is partly due to the superb sensitivity of the human auditory system compared to that of human visual system so that less data could be embedded into audio signals inaudibly and robustly. Gang *et al.* [3] presented a audio steganographic scheme specifically designed for high MP3 resistance where the covert data embedded in the signal properties which would be least affected by the compression process, namely amplitude, DFT phase domain and noisy component. The method can achieve an embedding capability at 20~60 bits/second but with a poor imperceptibility. Gopalan *et al.* [4] presented a steganography method specifically aimed at embedding covert *voice* messages in wireless communication using a spread spectrum based utterance embedding strategy. Cvejic *et al.* [5] proposed a method aimed at increasing the embedding capacity of LSB-based audio steganography by adopting the minimum error replacement and modifying error diffusions in the embedding process.

In this paper, we present an audio steganographic scheme by extending the wavelet domain SVD image watermarking [6] to audio signals. The experiment shows that the proposed audio steganography scheme can achieve a very high embedding capacity and an excellent resistance against MP3 compression compared to the existing audio steganography schemes while retaining a comparable imperceptibility.

2. Δ-Transform

The relatively less success in audio steganography, let alone music steganography, in comparison to its counterpart in image steganography is mainly attributed to the fact that the degree of the *sensitivity* and the range of dynamics of HAS, on which any audio steganographic strategy would be based, is much *higher* than that of HVS and thus the amount of the signal samples which could be inaudibly altered are much lesser, leading to seemingly uncompromised objectives: imperceptibility and high-capacity. This motivated us to the idea if the host *audio* signal could be *reversibly* Δ-transformed into a 2D *image* so that if the visual perception is well preserved with a certain watermarking embedding scheme on image, the audio perception on the audio inversely transformed from the watermarked image would be well preserved, audio steganography would become an easier image steganographic issue. In this paper, we design a reversible mathematical transform between the signal wavelet coefficients and a 2D arrangement of the coefficients (image), where an increased correlation between the PSNR measures on images and the two widely used

perceptual audio quality models on audio, namely, PAQM (*perceptual audio quality measure*) and NMR (*noise-to-mask ratio*) is preserved. With this transform, dubbed, Δ*-transform,* the image watermarking based on the wavelet domain singular values modification could be applied for the audio signal to form an audio steganography scheme.

The Δ-transform is a simple procedure to convert a 1D sequence of coefficients (quantized to range [0,255]) $C = \{c_1, c_2, \cdots, c_k, \cdots\}$ to a 2D image by down-sampling sequence C into sequence $P = \{p_1, p_2, ..., p_m, ...\} = \{c_a, c_{a+\Delta}, c_{a+2\Delta}, \cdots, c_{a+(m-1)\Delta}, \cdots\}$ (Δ*-transform*) and arrange P in the raster scan order into a 2D image I_A of size $n \times n$. This can be formulated mathematically as follows

$$I_A = (I_{ij})_{n\times n} \qquad I_{ij} = p_m = c_k,$$
$$k = a + (m-1) * \Delta, m = (i-1) * n + j, 1 \le i, j \le n \quad (1)$$

where parameter Δ represents the interval of between the selected samples and thus controls the ratio of the coefficients which could be potentially affected by watermarking and parameter a marks the beginning of the downsampling, which is randomly selected for the enhanced security. Both parameters serve as watermarking secret keys. In our experiment, Δ is set to 13, so that for a 10 seconds audio signal A sampled at 44.1 kHz can be Δ-transformed into an image I_A with 128×128 pixels. Then one bit of covert data can be embedded into a block of size $w \times w$ pixels so that the covert data (image, audio, text) of $\lfloor 2^7 / w \rfloor \times \lfloor 2^7 / w \rfloor$ bits can be embedded. The embedding capacity can be controlled by adjusting parameters Δ and w. For example, for $w = 8$ and $w = 4$, 2^8 bits and 2^{10} bits can be embedded, resulting in capacities at 25.6bps and at 102.4bps, respectively.

3. Audio Steganography Based on Wavelet Domain SVD

The proposed audio steganography scheme can be viewed as the extension of the image-adaptive watermarking scheme based on wavelet domain SVD. The embedding process is comprised of five steps.

1) Apply discrete wavelet transform (DWT) on the input audio signal A at level one to obtain the baseband coefficients C_A^L and high-pass subbands coefficients C_A^H;
2) Form image I_A from coefficients C_A^L by a simple Δ*-transform* procedure;
3) Apply the image-adaptive SVD watermarking on I_A to obtain the watermarked image I_A^W;
4) Apply the inverse Δ*-transform* on I_A^W to obtain the watermarked baseband coefficients $\tilde{C}_A^L$;
5) Perform the inverse discrete wavelet transform (IDWT) on baseband coefficients $\tilde{C}_A^L$ and high-pass subbands coefficients C_A^H to produce the watermarked audio signal A_W.

4. Experiments

Experiments on some benchmark audio signals have been conducted for measuring the imperceptivity and the robustness to MP3 compression. We applied the scheme on 10 samples of wideband audio signals with each signal being a raw WAV file of 10 seconds in length, sampled at 44.1 kHz and quantized to 16 bits per sample (CD quality). The benchmarks were selected in an attempt to experiment how well the proposed steganography scheme can comprehend a wide range of signals possessive of different spectral characteristics.

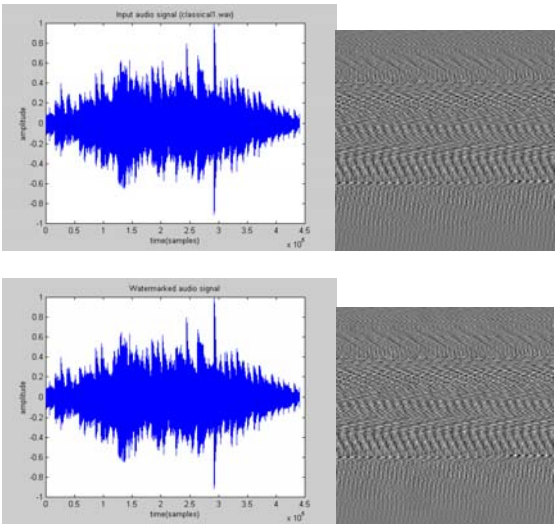

Figure 1. Clockwise: Input original audio signal A; Image I_A; Watermarked I_A^W with PSNR=42.3333 between I_A and I_A^W; Watermarked audio signal A_W with SNR=32.8782 between A and A_W.

Figure 1 shows the watermarking embedding and extraction, respectively, for audio file *classical1*. Table 1 lists the signal-to-noise ratio (SNR) of the stego signals and the embedding capacities corresponding to the parameter w set to various values $w = 4,5,6,7,8$, respectively.

From Table 1, it can be seen that the proposed scheme outperforms better in SNR measures than Echo Coding, Phase Coding and Frequency Mask methods. Table 2 shows the detection errors for MP3 compressions at various rates. Note that embedded data can be extracted error-free from the MP3-compressed audio signals at any rates above 48kpbs, an excellent resistance to MP3 compression.

Table 1: SNR and Hiding Capacity of Watermarked Audio

	25.6 bps	32.4 bps	44.1 bps	62.5 bps	102.4 bps
Classical1	32.8782	31.3507	30.5106	29.1643	26.8774
Classical2	31.2507	30.0946	29.3014	27.9348	25.3278
Country1	38.1446	36.9574	35.0462	34.0922	32.1360
Country2	37.3702	35.9334	34.0524	33.0994	30.4366
Folk1	39.3753	38.4527	36.7213	35.2549	33.2998
Folk2	39.5238	39.5238	38.5005	36.9099	32.9009
Pop1	41.2147	40.0362	39.4157	37.8552	35.4519
Pop2	39.0620	38.9678	35.7043	34.5645	32.2893
Blues1	37.7858	37.3779	35.7415	34.3379	32.0426
Blues2	35.9672	34.6455	33.4757	31.9881	30.1286

Table 2: Robustness to MP3 Compression

	25.6 bps	**32.4 bps**	**44.1 bps**	**62.5 bps**	**102.4 bps**
Classical1	32.8782	31.3507	30.5106	29.1643	26.8774
Classical2	31.2507	30.0946	29.3014	27.9348	25.3278
Country1	38.1446	36.9574	35.0462	34.0922	32.1360
Country2	37.3702	35.9334	34.0524	33.0994	30.4366
Folk1	39.3753	38.4527	36.7213	35.2549	33.2998
Folk2	39.5238	39.5238	38.5005	36.9099	32.9009
Pop1	41.2147	40.0362	39.4157	37.8552	35.4519
Pop2	39.0620	38.9678	35.7043	34.5645	32.2893
Blues1	37.7858	37.3779	35.7415	34.3379	32.0426
Blues2	35.9672	34.6455	33.4757	31.9881	30.1286

On the other hand, the scheme is vulnerable to many intentional signal manipulations thus unsuitable for applications such as copyright protection, digital rights management and access controls. But since the steganographic music contents are perceptually indiscernible to the observers, they will be unlikely targeted by the intentional manipulations.

5. Conclusion

The audio steganography presented in this paper embeds a covert message (audio signal, image or text) into a host audio signal by an indirect audio-adaptive wavelet domain SVD-based watermarking scheme. Experiments show that the audio steganography scheme leads to an embedding capacity up to 102.4bps while still retaining a good perceptual quality (high SNR values) and excellent resistance to MP3 compression in comparison with the existing audio steganography schemes. Meanwhile, the embedded data extraction is oblivious and secure that the embedding model is solely determined by the quantization parameters and/or transform parameters Δ and a, the malicious detection of the watermark would not be possible without knowledge of them.

References

[1] N. Provos, P. Honeyman, Hide and seek: an introduction to steganography**,** *Security & Privacy, IEEE Magazine,* Volume: 1 Issue: 3, pp. 32 -44N, May-June 2003.

[2] F. Johnson and S. Katzenbeisser, A Survey of steganographic Techniques, in S. Katzenbeisser and F. Petitcolas (Eds.): Information Hiding, pp. 43-78, Artech House, Norwood, MA, 2000.

[3] Litao Gang, A.N. Akansu, M. Ramkumar, MP3 resistant oblivious steganography**,** *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,* Vol. 3, pp. 1365-1368, 7-11 May 2001.

[4] K. Gopalan, Audio steganography using bit modification, *Proceedings of International Conference on Multimedia and Expo,* Vol. 1, pp. 629-632, 6-9 July 2003.

[5] N. Cvejic, T. Seppiinen, Increasing the capacity of LSB-based audio steganography**,** *Processings of IEEE Workshop on Multimedia Signal*, pp. 336 –338, 2002.

[6] P. Bao and X. Ma, Image Adaptive Watermarking Using Wavelet Domain Singular Value Decomposition Authentication, *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 15, No. 4, April, 2005.