Study on information hiding based on cochlear-delay characteristics for sound signals

Toshizo Kosugi (0910023)

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

February 8, 2011

Keywords: cochlear delay caracteristics, parallel architecture, cascade architecture, composite architecture, speech steganography.

1 Introduction

The multimedia contents (text, sound, image, and video) are now widely available from the internet due to the popularity of PCs and the proliferation of high-speed large-capacity networks. In particular, the speech communication and music contents are recently the rapidly growing. Since demands to protect the multimedia content have greatly increased in recent years, multimedia information hiding (MIH) is currently a very important technology in this research field. MIH techniques have aimed to preserve the values of multimedia information (such as text, digital-audio, images, and video) from their unauthorized copying by hiding imperceptible marks such as copyright notice. MIH techniques are, in general, composed of content protection of multimedia information such as watermarking and steganography. Since it is possible to use MIH techniques together with cryptographic techniques, they are applicable for secure content authentication such as fingerprint.

In general, multimedia information hiding methods must satisfy four requirements to provide: (a) **inaudibility** (inaudible to humans with no sound distortion caused by the embedded data), (b) **confidentiality** (secure and undetectable concealment of embedded data), (c) **robustness** (not affected when subjected to techniques such as data compression), and (d) **embedding limitation** (to increase embedding limitations). The first requirement (**inaudibility**) is the most important in the method of information hiding for sounds because this must not affect the sound quality of the original signal. The second requirement (**confidentiality**) is important to conceal embedded data to protect copyright, and it is important that users do not know whether the embedded data contains information hiding or not. The third requirement (**robustness**) is important to ensure that the information hiding methods are tamper-proof to resist any manipulations by illegal users. The last requirement (**embedding limitations**) is important to increase the flexibility of information hiding technique. Although several methods have been proposed, these methods have suffered from serious drawbacks in any of the four requirements. As the method solving the problems with regard to requirements (a) and (c), a method of information hiding based on cochlear delay has been proposed by Unoki & Hamada (a base architecture). This research proposes a composite architecture by reasonably incorporating parallel and cascade architectures to further improve embedding limitations.

2 Composite architecture

A cochlear-delay filter is designed 1st-order IIR all-pass filter to model cochlear delay characteristics. An IIR all-pass filter is usually used to control delays in which amplitude spectra are passed equally without any loss. Based on the expression of 1-bit, it is also possible to control 2-th cochlear delays using the base architecture. Imabeppu *et al.* improved the previous approach to improve embedding limitations with the method by using a parallel architecture. Based on the expression of N-bits, it is also possible to control $M(=2^N)$-th cochlear delays using the parallel architecture. However, M-th cochlear delays must not be beyond the cochlear delay, which was scaled by 1/10. We have improved our previous approach to improve embedding limitations with the method by using a cascade architecture. Based on the expression of L-bits, it is also possible to control $R(=2^L)$-th cochlear delays using the cascade architecture. However, inaudibility is affected increasing the number of R-th cochlear delays. In addition, the value of parameter b must be from 0 to 1. Thus, we propose a composite architecture by reasonably incorporating parallel and cascade architectures. Based on the expression of $N \cdot L$-bits, it is also possible to control $U(=2^{N \cdot L})$-th cochlear delays using the composite architecture. We designed the composite architecture for the cochlear delay filter $H_{\mathrm{Cmp}}(z)$ as follows:

$$H_{\mathrm{Cmp}}(z) \quad := \quad \prod_{\ell=1}^{L} H_{\ell,m}(z) = \prod_{\ell=1}^{L} \frac{-b_{\ell,m} + z^{-1}}{1 - b_{\ell,m} z^{-1}} \tag{1}$$

where $\ell = 1, 2, \cdots, L$ and $m = 0, 1, \cdots, M-1$. Here, the group delay, $\tau_{\mathrm{Cmp}}(\omega)$, can be obtained as:

$$\tau_{\mathrm{Cmp}}(\omega) \quad = \quad \sum_{\ell=1}^{L} \tau_{\ell,m}(\omega) \tag{2}$$

$$\tau_{\ell,m}(\omega) \quad = \quad -\frac{d\arg(H_{\ell,m}(e^{j\omega}))}{d\omega} \tag{3}$$

For example, the group delays in the composite architecture with $N = 2$ and $L = 2$ are represented as 16-types of $\tau_{Cmp}(\omega)$ in Eq. 3. Therefore, the composite architecture can embed 4-bits per frame into the original signal.

Embedded information data were detected as follows: (1) We assume that both original signal and embedded information data signal are available with this method. (2) The original signal and the embedded information data signal are decomposed to become overlapping segments using the same window function used in embedding the data. (3) The phase difference is calculated in each segment. (4) The summed phase differences of phase difference to the respective phase spectrum of the filters are calculated to estimate the group delay characteristics of used for embedding the data. (5) The embedded data are detected using the U-th cochlear filter.

3 Objective evaluations

We evaluated the improved methods (the parallel ($N = 1, 2, 3,$ and 4), the cascaded ($L = 1, 2, 3,$ and 4), and the composite architectures ($L = 2$ and $N = 2$)) by carrying out three objective experiments in audio signal. The results of objective evaluations revealed that embedding limitations with the parallel ($L = 1$ and $N = 2$) and cascade architecture ($L = 2$ and $N = 1$) were 1024 (= 512 fps $\times 2$) bps. The results of robustness tests revealed that embedding limitations with the composite architecture ($L = 2$ and $N = 2$) was 256 (= 64 fps $\times 4$) bps. We evaluated the improved methods by carrying out robustness tests in speech signal. In addition, we evaluated the improved methods by carrying out StirMark benchmark for audio tests. These results revealed that the improved methods are robust against (i) **Noise**, (ii) **Amplitude**, (iii) **Bit**, and (v) **Filtering**. However, the improved methods which embedded a watermark in phase domain is not robust to the attack of (iv) **Data**, (vi) **Phase**, and (vii) **Echo**.

We evaluated the improved methods by carrying out three objective experiments in speech signal. The results of objective evaluations revealed that embedding limitations with the parallel ($L = 1$ and $N = 2$) were 512 (= 256 fps $\times 2$) bps. We evaluated the improved methods by carrying out robustness tests in audio signal. The results of robustness tests revealed that embedding limitations with the parallel architecture ($L = 1$ and $N = 3$) was 384 (= 128 fps $\times 3$) bps.

We proposed speech steganography by using our method. We evaluated the speech steganography by carrying out embedding 32×32 bitmap image. As a result, the bit-detection rate is 97.5% (bit error 26 bit). We evaluated the speech steganography by carrying out robustness test of PCM codec. As a result, the bit-detection rate is 90.9% (bit error 93 bit). Speech steganography is not only inaudibility, but robustness also very important. The proposed method can be used to speech steganography.

4 Conclusions

We investigated how the proposed approach could be implemented to produce an efficient architecture to further improve embedding limitations with our proposed approach. We carried out objective evaluations and robustness tests on the proposed architectures including base, parallel, cascade, and composite architectures. As results, the problems with regard to requirements (**inaudibility** and **robustness**) were solved by the proposed method. Moreover, embedding limitations of the parallel, cascade, and composite architectures increased from that of the base architecture. It suggests that my proposed method can be used as useful multimedia information hiding technique rather than the base architecture proposed by Unoki & Hamada.