TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 07/21 pp55-61
Volume 14, Number 1, February 2009

Performance Analysis of Data Hiding in MPEG-4 AAC Audio*

XU Shuzheng (徐淑正)**, ZHANG Peng (张 鹏), WANG Pengjun (王鹏军), YANG Huazhong (杨华中)

**Institute of Circuits and Systems, Department of Electronic Engineering,
Tsinghua University, Beijing 100084, China**

Abstract: A high capacity data hiding technique was developed for compressed digital audio. As perceptual audio coding has become the accepted technology for storage and transmission of audio signals, compressed audio information hiding enables robust, imperceptible transmission of data within audio signals, thus allowing valuable information to be attached to the content, such as the song title, lyrics, composer's name, and artist or property rights related data. This paper describes simultaneous low bitrate encoding and information hiding for highly compressed audio signals. The information hiding is implemented in the quantization process of the audio content which improves robustness, signal quality, and security. The imperceptibility of the embedded data is ensured based on the masking property of the human auditory system (HAS). The robustness and security are evaluated by various attacking algorithms. Tests with an extended MPEG4 advanced audio coding (AAC) encoder confirm that the method is robust to the regular and singular groups method (RS) and sample pair analysis (SPA) attacks as well as other statistical steganalysis method attacks.

Key words: advanced audio coding (AAC); steganography; steganalysis; human auditory system (HAS); data hiding; information hiding

Introduction

Data hiding, or steganography, is a challenging area of research which has attracted much attention for centuries[1]. As an active area of research, new techniques are constantly emerging. In contrast to the active research focusing on image or video schemes, there is little research on data hiding in audio schemes, especially in high compressed audio like the MPEG-4 advanced audio coding (AAC). Human ears can detect even a small amount of embedded noise especially when the signal is weak.

Received: 2008-03-26; revised: 2008-05-13

* Supported by the Chuanxin Foundation (No. 110109001) and the Basic Research Foundation of Tsinghua National Laboratory for Information Science and Technology (TNList)

** To whom correspondence should be addressed.
E-mail: xusz@mail.tsinghua.edu.cn; Tel: 86-10-62771708

However, effective data hiding in audio can create robust and imperceptible data within audio signals, thus allowing property rights protection or secret data exchange. Data hiding acts as an intentional noise inserted into digital audio signals. Data hiding is a good method for preventing illicit copying and dissemination of copyrighted audio and can also be used to transfer secret information in convert communications.

A number of systems are currently available to transmit a watermark in a hidden channel within uncompressed audio signals. However, compressed audio, using a perceptual audio encoder to save storage space as well as to minimize the transmission bandwidth and time, is very popular and widely used over the Internet and in wireless communications. Current research has mainly focused on the pulse code modulation (PCM) domain for AAC files[2], where the hidden information disturbs the signal during quantization in the audio

encoder. There is also some research that hides data directly into the compressed audio[3]. The disadvantage of this method is that the hidden data treated as an additional disturbance to the quantization noise, results in obvious degradation of the audio quality. This paper presents a method for data hiding during the encoding process of the MPEG-4 AAC audio. The scheme has flexibility in hiding the data and obtains a good balance between robustness, security, and capacity.

1 State of the Art

1.1 Advanced audio coding in MPEG-4

AAC, which is included in the MPEG-4 audio standard, is a compatible successor to the MPEG-2 audio coding[4]. As perceptual encoders, they exploit knowledge of human perception to shape the noise distribution introduced by the irrelevancy reduction to achieve the best possible audio quality. The basic structure of the perceptual encoder used in the MPEG audio family is shown in Fig. 1. AAC encoding consists of frequency transform, quantization, entropy coding, and bitstream multiplexing steps. AAC typically employs modified discrete cosine transform (MDCT) with 1024 samples per frame. The 1024 frequency samples in each time frame are separated into 49 frequency bands. Within the same frequency band, samples are considered to have similar perceptual effect to the human ear and, hence, share the same quantization step size. Perceptual modeling is applied to the MDCT coefficients to estimate the maximum amount of distortion that can be withstood by each coefficient. The quantization step size is iteratively modified until the rate is below the target bit rate and the distortion is below the maximum acceptable value obtained from the perceptual model. Huffman coding is used to encode the quantized coefficients and the quantization step size. Finally, the coded bits with some side information are multiplexed into a single bitstream to be transmitted or saved.

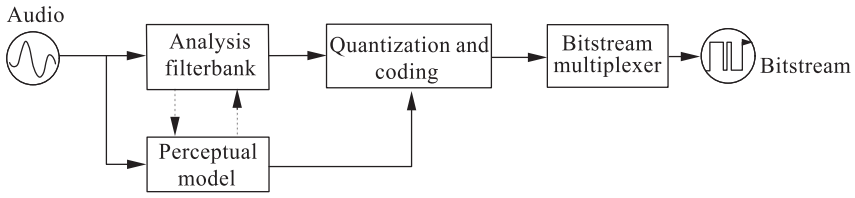

Fig. 1 Basic structure of a perceptual encoder

The main task of the perceptual encoding is to compress the digital audio as much as possible such that the sound quality of the reconstructed signal is exactly the same as or as close as possible to the original audio signal. There are also some other requirements like low complexity and flexibility for a wide application area. The modern highly developed perceptual models and efficient redundancy reduction audio encoding systems offer a multitude of coding tools like temporal noise shaping (TNS), joint stereo coding, and perceptual noise substitution (PNS) to fulfill these requirements.

1.2 Data hiding

Various schemes are used for data hiding in audio, such as echo hiding[5], time domain modification[6], and spread spectrum technology[7]. Data hiding in audio must satisfy at least the three constraints of security or imperceptibility, robustness, and capacity. These terms are commonly used to describe the properties of different data hiding schemes.

Security Data hiding in audio is also called inaudibility. In most cases, security, including perceptual transparency of the hidden data, is considered to be the most important issue. In other words, the noise introduced by the hidden data should be almost inaudible and should not degrade the audio quality. The statistical properties of the embedded audio signal should be exactly the same as the original audio to ensure that the hidden data is imperceptible and undetectable by third parties.

Robustness The algorithm should be robust enough to withstand unintentional or intentional attempts such as removal or alteration of the hidden data. Even with unfavorable conditions such as bad wireless channels which degrade the audio quality, the hidden data should be recovered successfully.

Capacity Often, the capacity of the hidden data is also a very demanding aspect. Capacity refers to the number of bits per second that can be transmitted by the data hiding system. This depends on the underlying

technology and the choice of parameters for the hiding scheme. At present, the data rates reach several hundred bits per second.

Security, robustness, and capacity have contradictory arguments so they cannot be adjusted independently. For instance, increases of the data hiding capacity will degrade the robustness and security. These trade-offs form the triangle shown in Fig. 2 with an appropriate operating point within the limits of the triangle chosen for different applications.

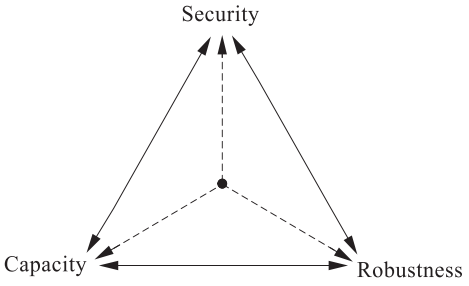

Fig. 2 Trade-off between security, robustness, and capacity

2 Data Hiding Scheme

2.1 Basic structure

The present scheme alters the quantization and coding process by modifying the scale factor to produce more bits, with the redundant bits replaced by hidden data. The basic structure of the system including all of the relevant parts of the MPEG-4 AAC perceptual encoder is shown in Fig. 3.

The encoding scheme is very efficient. The uncompressed input audio signal is processed by the analysis filterbank with parameters gathered from both the signal itself and the perceptual model. The data is first compressed, encrypted, and then hidden in the AAC bitstream. The hiding process takes place at the heart of the Layer III encoding process in the inner loop. The inner loop quantizes the input data and reduces the quantization step size until the quantized data can be coded with the available number of bits. Another loop checks that the distortions introduced by the quantization do not exceed the threshold defined by the psychoacoustic model. Finally, the multiplexer produces a bitstream including data and side information to complete the data hiding process. The adherence to the psychoacoustic model makes it very difficult to "hear" the hidden data.

For example, during the traditional quantization and coding process, the compressed signal can be represented as

$$X_{\mathrm{s}} = 2^{\frac{3T}{16}} X_i^{\frac{3}{4}} = \left(2^{\frac{T}{4}} X_i\right)^{\frac{3}{4}} \tag{1}$$

where T is the scale factor calculated by the perceptual model and X_i (i=1, …, 960) represents the quantized data. For the same signal X_{s}, the scale factor T is reduced, the quantization step 2^T becomes smaller, and X_i becomes larger, and has to be represented by more binary bits. If the scale factor T is 4 which means that the masking threshold for psychological hearing is $2^T = 16$, and the signal X_{s} is 128, then X_i is 8 (the binary form is 1000). If the scale factor

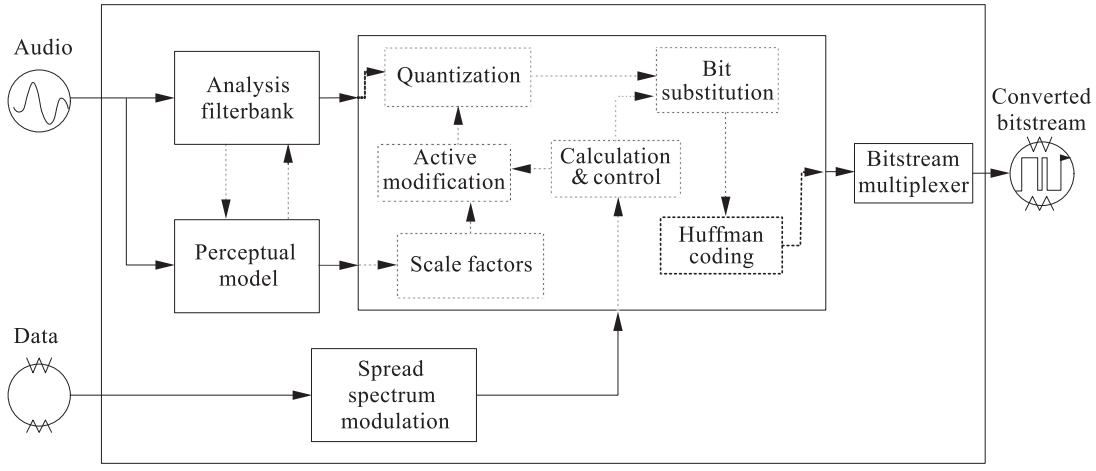

Fig. 3 Basic data hiding structure for the AAC perceptual encoder

T is reduced from 4 to 2, then X_i will become 16 (10 000). If the least significant bit (LSB) of X_i is replaced by the hidden data, then the hidden data will sound like noise in the audio signal with the introduced noise energy still below the masking threshold level so it will not affect the audio quality. This active modification method can flexibly increase the hidden data capacity by varying the scale factors and the number of spectral subbands. The robustness and security of the hidden data are achieved by the spread spectrum and the pseudo selection of the subbands for data hiding.

2.2 Data extraction

The extraction process should be implemented before reconstructing the MDCT coefficients because the information is embedded in the quantized coefficient. Blind extraction introduces two problems into the data extraction process. A normal audio frame which does not have any information may be treated as a special audio frame when the circle redundant check (CRC) byte is passed with a very low probability. However, a special audio frame may be treated as a normal audio frame when mistakes occur in the channel. These two problems are common in wireless communications. Thus, the information should be changed to a special data format with a CRC byte inserted into the information for wireless transmissions.

2.3 Characteristics

Compared with current data hiding schemes, the algorithm presented here has the following advantages:

(1) The scheme enables optimal coordination between the quantization strategy of the encoder and the data hiding process. All information about the modification of the quantization parameters is taken into account by the data hiding process.

(2) The scheme can flexibly vary the data hiding capacity by adjusting the number of coefficients for the quantization, and thus this scheme provides adaptive data hiding. When the embedded data bit rate needs to be changed, the scheme can adaptively change the quantization parameters.

(3) The scheme assures high data hiding security. The coefficients which carry the hidden data are picked up by a pseudo random mechanism which is unlikely to be intercepted by a third party.

3 Test Results

3.1 Audio quality

The quality of the data hiding AAC signals was carefully assessed by subjective listening tests. The AAC test items represent a variety of typical materials from digital radio broadcasting source signals[8] and the test used a MUSHRA listening test[9]. The quality was evaluated by ten listeners who were all experienced and familiar with the set of test items. Figure 4 shows the mean values and 95% confidence level for the critical test items. The results show that there is no statistically significant degradation between the original AAC encoded items and the modified signals for all hidden data capacities varies from 10% to 100%.

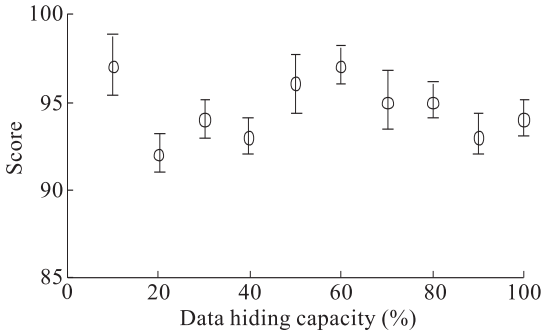

Fig. 4 Evaluated quality of original MPEG-4 AAC audio and modified AAC audio

Signal-to-noise ratio (SNR) is a typical objective measure of audio quality. In the compression domain, the signal is defined as the original uncompressed audio and the noise is defined as the distortion between the decompressed audio and the uncompressed audio. The SNR of MPEG-4 AAC compression is typically 18-20 dB. The hidden data will introduce some quantization noise which will slightly reduce the SNR value. Statistical calculations with large test sets show that there is only about 0.5-1.5 dB loss in the SNR when hiding messages at the maximum bit rate. In addition, the use of the psychoacoustic model guarantees that the noise introduced by the data embedding process does not exceed the threshold and only slightly degrades the audio quality.

3.2 Hiding data capacity

Since the active modification scheme requires more bits to represent the signal, the compressed audio will be enlarged. Figure 5 shows how the size of the carrier

of the AAC file is increased as the hidden data increases, with the increased size of the AAC file being much smaller than that of the hidden data. For instance, when the hidden data size is increased by 60%, the ratio of the hidden AAC file to the original file size is 132%, only increased by 32%. There are also some flat regions in the curve which indicate that the hidden data size do not affect the AAC file size. Since the strategy is to modify the scale factors of some spectral subbands whose length are fixed, different hidden data capacities may require the same number of modified subband scale factors.

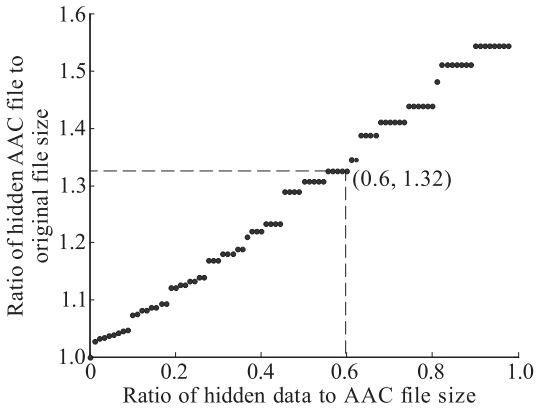

Fig. 5 File sizes of the original MPEG-4 AAC and converted AAC audio files

3.3 Time and spectrum waves

Tests were also run to evaluate whether the hidden data in an AAC audio can be detected through more qualitative methods, and the characteristics of the changes. Further research could yield a set of analytical formulas describing the relationships between the hidden data, the voice or carrier data, the measurements, and the results. This research only analyzed the sound waveforms in the time and frequency domains. CoolEdit2000 was used for the time and spectral analyses. The original AAC audio waveforms are compared with the recording containing the encrypted hidden data. The resulting waveforms in the time and frequency domain are shown in Figs. 6 and 7.

Both the time and frequency analyses do not show any distinguishing differences between the original recording and the recording with the hidden data. However, there seems to be a slight amount of signal loss and a slight increase in distortion due to the hidden data, but the overall measurement scores are very good. The very minor alterations of the original recording when the data is hidden are so small as to not be perceptible.

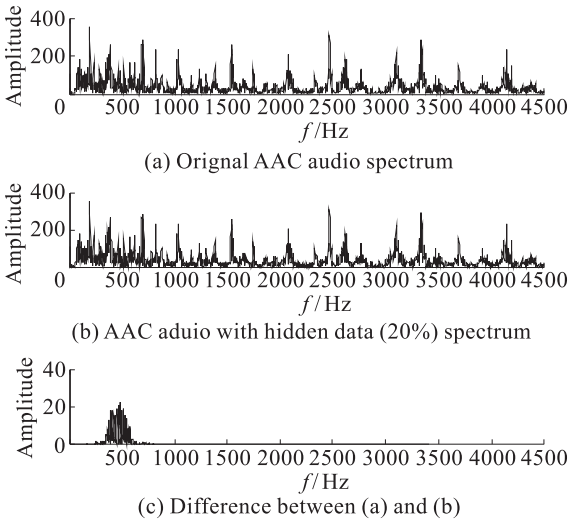

Fig. 6 Frequency domain waveform of the MPEG-4 AAC and converted AAC audio streams

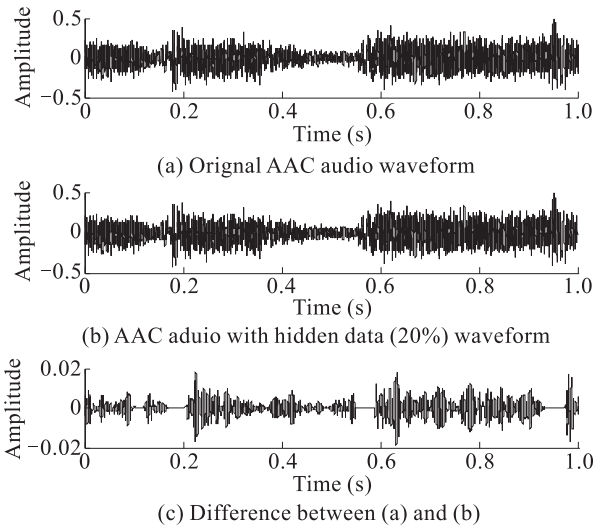

Fig. 7 Time domain waveform of the MPEG-4 AAC and converted AAC audio streams

3.4 RS and SPA detection

The regular and singular groups method (RS) introduced by Fridrich et al.[10,11] and the sample pair analysis (SPA) introduced by Dumitrescu et al.[12] have proven to be the most useful tools for detecting least significant bit (LSB) embedding[13]. The RS estimates the length of the embedded data by counting the length variation of the regular set and the singular set. When the data is embedded in a random order, the length of the embedded data can be accurately estimated by this

method. The SPA accurately estimates the length of the embedded data when more than 3% of the data is embedded in LSB. Andrew[13] verified the credibility of the RS and the SPA in many tests.

The present scheme is an LSB embedded method. However, the scheme can effectively resist the RS and SPA detections since the embedded process works in the frequency and compression domain. The AAC V2 correlation is much higher than that of LSB by RS detection algorithm, and the correlation becomes better as the hiding data increased. Figures 8 and 9 show the results of using RS and SPA to analyze sequence sets embedded data by a normal LSB scheme and the current AAC V2 scheme, which indicate that the later has better performance of security.

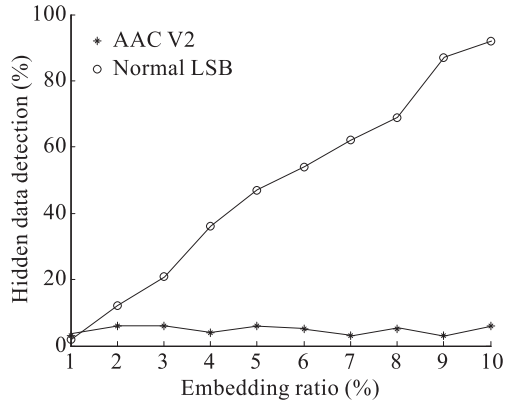

Fig. 8 RS analysis results for detecting hidden information

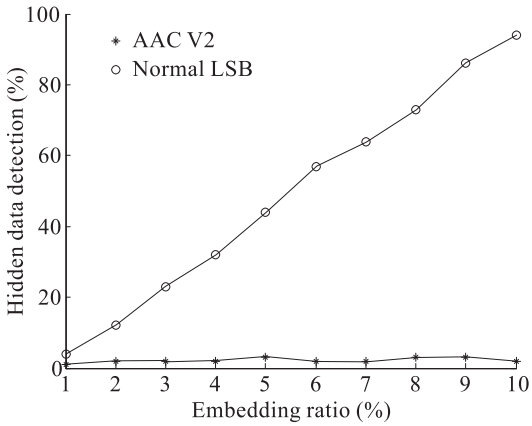

Fig. 9 SPA analysis results for detecting hidden information

3.5 Detection analysis

Steganalysis can be charaterized as classification of two elements. The carrier and the stego are objectives to be classified with the condition that the carrier and the stego are both blind to the classifier. A judgment result very close to 50% indicates high security with an absolutely secure system having a result of 50%.

The classifier structure is shown in Fig. 10. The distance measurement module calculates the characteristic difference between the converted and the original AAC files. Since the analysis does not know what kinds of characteristics have been changed during the data hiding process and to assure the generality of the analysis, the system analyzes as many characteristics as possible.

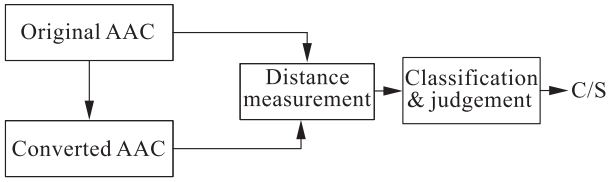

Fig. 10 Basic classifier structure

Three kinds of characteristics are considered in the classifier[14,15].

(1) Time domain measurements include the SNR, segment SNR (SNRseg), Czenakowski distance (CAD), and Minkowsky measurement (MM).

(2) Frequency domain measurements include the logarithm likelihood ratio, Itakura-Saito distance, COSH distance, inverse spectrum distance, multitude spectrum distortion, and phase spectrum distortion.

(3) Perceptual domain measurements include the BARK distortion.

Linearity regression, classification tree, vector support, and projection were selected as the classification algorithms in the classifier.

Generally speaking, the two classifier results are detection and false detection. The testing sets include 200 groups of AAC files compressed from 22.05 kHz to 48 kHz, 16 bits, mono and stereo PCM bitstreams. The test results are shown in Table 1.

The data in Table 1 shows that the measurement probability of the AAC data hiding algorithm is almost 50% while the possibilities for other hiding algorithms, LSB, ECHO, Hide4PGP, and Steghide, are much higher. Thus, the AAC data hiding algorithm has comparatively higher security.

Table 1 Performance comparison of different data hiding algorithms

Algorithm	Probability of detection/probability of false detection			
	Linearity regression	Vector support	Classification tree	Projection
LSB	0.72/0.38	0.79/0.42	0.70/0.60	0.82/0.44
ECHO	0.78/0.26	0.81/0.36	0.84/0.38	0.79/0.22
Hide4PGP	0.75/0.40	0.86/0.38	0.60/0.43	0.82/0.32
Steghide	0.67/0.44	0.71/0.37	0.69/0.35	0.70/0.39
AAC(BPF*=8)	0.52/0.41	0.66/0.42	0.56/0.42	0.48/0.44

* BPF: Byte per frame.

4 Concluding Remarks

This paper first gives an overview of data hiding compressed audio signals and the state-of-the-art technologies for source compression especially AAC formats. A data hiding scheme for MPEG4 AAC is presented based on the human auditory system. The data hiding occurs in the quantization process of the MDCT procedure of the AAC coding, and the hidden data capacity can be adaptive to the original data rate by active modification of scale factors without degradation of audio quality. The implementation is very efficient with good robustness achieved by enabling optimum coordination of the AAC quantization and data hiding processes. Tests confirm the potential of this new technology and show it to be an attractive solution for property rights protection and covert communications.

References

[1] Petitcolas F A P, Anderson R J, Kuhn M G. Information hiding—A survey. In: Proceedings of the IEEE Special Issue on Protection of Multimedia Content, 1999, **87**(7): 1062-1078. http://www.petitcolas.net/fabien/publications/ieee99-infohiding.pdf.

[2] Cheng S, Yu H, Xiong Z. Enhanced spread spectrum watermarking of AAC audio. In: Proceeding ICASSP'02. Orlando, FL, USA, 2002: 505-510.

[3] Neubauer C, Herre J. Audio watermarking of MPEG-2 AAC bitstreams. In: 108th AES Convention, Audio Engineering Society Preprint 5101. Paris, France, 2000: 1153-1161.

[4] Koenen R. Overview of the MPEG-4 standard. Document, M-arz 2002. ISO/IEC JTC1/SC29/WG11 N, 2001.

[5] Gruhl D, Lu A, Bender W. Echo hiding. *Lecture Notes in Computer Science*, 1996, **1174**: 295-315.

[6] Xu Chansheng, Wu Jiankang, Sun Quibin, et al. Applications of digital watermarking technology in audio signals. *Journal of Audio Engineering Society*, 1999, **47**(10): 805-812.

[7] Garcia R A. Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory. In: 107th AES Convention. New York, USA, 1999: 2713-2720.

[8] Digital Radio Mondiale (DRM) System Specification (ETSI) 201980, V2.1.1, 2005.

[9] Stoll G, Kozamernik F. A method for subjective listening tests of intermediate audio quality. ITU Working Party 10-11Q, Sept. 2000.

[10] Fridrich J, Goljan M. Practical steganalysis of digital images — State of the art. In: Proceedings of SPIE: Security and Watermarking of Multimedia Contents IV. 2002: 1-13.

[11] Fridrich J, Goljan M, Du R. Reliable detection of LSB steganography in color and grayscale images. In: Proceedings of the ACM Workshop Multimedia Security. Ottawa, Canada, 2001: 27-30.

[12] Dumitrescu S, Wu Xiaolin, Wang Zhe. Detection of LSB steganography via sample pair analysis. *IEEE Transactions on Signal Processing*, 2003, **51**(7): 1995-2007.

[13] Andrew D K. Improved detection of LSB steganography in grayscale images. *Lecture Notes in Computer Science*, 2004, **3200**: 97-115.

[14] Ozer H, Avcibas I, Sankur B, et al. Steganalysis of audio based on audio quality metrics. In: Security and Watermarking of Multimedia Contents. Santa Clara, USA, 2003: 55-66.

[15] Yang Bo. Extensive bark spectral distortion measurement for objective speech quality assessment. *Journal of University of Electronic Science and Technology of China*, 2006, **35**(3): 343-345.