Lecture Notes in Computer Science, 2776, 2003: 383-394

Data Hiding in Digital Audio by Frequency Domain Dithering

Shuozhong Wang, Xinpeng Zhang, and Kaiwen Zhang

Communication & Information Engineering, Shanghai University, Shanghai 200072, China
shuowang@yc.shu.edu.cn zhangxinpeng@263.net ztszkwzr@sh163.net

Abstract. A technique that inserts data densely into short frames in a digital audio signal by frequency domain dithering is described. With the proposed method, large embedding capacity can be realized, and the presence of the hidden data is imperceptible. Synchronization in detection is achieved by using a two-step search process that accurately locates a PN sequence-based pilot signal attached to the data during embedding. Except for a few system parameters, no information about the host signal or the embedded data is needed at the receiver. Experimental results show that the method is robust against attacks including AWGN interference and MP3 coding.

1 Introduction

As a result of the rapid development of digital technology and computer networks, digital multimedia materials are widely used and disseminated. Since digital information is easy to copy, protection of intellectual property rights has become a serious concern. As a means of IPR protection, watermarking [1-2] has attracted much attention. In addition, information-hiding techniques have also found applications in covert communication, or steganography [3]. The aim is to convey information under the cover of an apparently innocuous host material, which differs from traditional encryption as not only the contents of the transmitted data are kept unintelligible to eavesdroppers, but also the very fact that communication is taking place is hidden. Clearly, a sufficient data capacity is an important factor in covert communication. This is in contrast to the IPR protection-oriented watermarking in which robustness is a primary specification.

Watermarking in digital audio has also received considerable research interests. Many techniques [4~6] have been proposed based on the characteristics of digital audio signals and the human auditory system (HAS). Some time-domain methods can hide a large amount of data but are not robust enough. Among the frequency-domain techniques, phase coding makes use of the insensitivity of HAS to the absolute phase in the Fourier transform coefficients. Inaudible embedding is achievable with a small phase change representing the embedded data. In echo data hiding, the hidden data is carried by parameters of an introduced echo (reverberation) close enough to the original signal.

This paper describes a data hiding approach that uses an audio signal as the host. Since HAS is very sensitive to slight distortion, tiny changes in the audio signal may be perceptible to normal listeners. Consequently, the achieved embedding capacity in early works was relatively low. For example, the hiding rate of phase coding is around 8~32bps, while DSSS only allows 4bps [5]. This, of course, makes the techniques impractical to be used in covert communication. The objective of this work is to develop a method that can hide a substantial quantity of data into a host audio without causing audible distortion. The proposed scheme makes use of the psychoacoustic masking both in the time domain and in the frequency domain to choose a series of candidate frames. Data are inserted into the spectrum of selected short frames of the host waveform using a technique of dither modulation. The approach can also be viewed as an application of orthogonal frequency division multiplexing with part of the subcarriers being the original audio spectral components modified by the stego data. In order to acquire synchronization in detection, a pilot signal is appended to the stego data. Knowledge about the host signal and stego data is not required in extraction.

The rest of the paper is organized as follows. Section 2 discusses the methodology, including data embedding, generation of the synchronization pilot, and extraction of the embedded data. Section 3 describes the experiments and presents the results. Section 4 concludes the paper. In the following discussion the two terms *data hiding* and *watermarking* will be used interchangeably.

2 Methodology

2.1 Selection of Candidate Audio Frames

There are two types of approach for data insertion in terms of the distribution of the hidden information. First, the embedded data are spread relatively evenly across a long period of time or over the entire image space. The simplest LSB approach, for example, replaces the least significant bits in all digital samples with an embedded sequence. Although the data capacity is large, this method is susceptible to attacks. Another example is the quantization index modulation in which several quantizers are used to introduce perturbations to a large number of samples [7]. In a time-domain technique, an audio signal is divided into segments, and all segments are watermarked with the same chaotic sequence having the same length as the segments [6].

The second type is to modify brief signal segments in an audio waveform or small areas in an image that are sparsely scattered over the entire signal. For example, a patchwork technique [5] statistically modifies randomly chosen small image patches according to the embedded data bit. In an audio watermarking system designed for encoding television sound, data were embedded into selected segments distributed over the signal [8].

The method proposed in this paper belongs to the latter category. Candidate frames in the host audio signal are first selected and discrete Fourier transformed. Watermark embedding is performed in the frequency domain. Studies on the HAS [1,9] indicate that slight distortion in the neighborhood of a high volume sound is inaudible. The masked period after a laud sound is generally longer than that prior to it. Therefore the candidate frame is selected in a relatively quiet segment immediately after a loud sound. The chosen segment must not be too quiet, though, in order to accommodate sufficient strength of the embedded signal. Meanwhile, the frequency domain masking is also utilized. Spectral components adjacent to large peaks, especially on the high frequency side, are less perceptible. Therefore the candidate frame should be chosen in segments that contain a significant amount of low frequency components. Based on these considerations, a search routine can be developed, which identifies a series of candidate frames in the host. Assume that each frame contains N samples: $\mathbf{s} = \{s(0), s(1), \dots, s(N-1)\}$, where N is chosen according to the required data rate and the imperceptibility requirement.

The highest frequency component in the signal is $W = f_s/2$ where f_s is the sampling frequency. Let w be the width of the frequency band occupied by watermark. $B=w/W$ is the normalized watermark bandwidth. Each watermark unit takes a portion in the spectrum of a signal frame lasting $T = N/f_s$ seconds. When f_s = 44.1 kHz and N = 1024, for example, T is 23.2ms.

In the present study, a band $[f_0, f_0+w]$ where $f_0=w=W/4$ is used. Before embedding, a test is performed to make sure that most of the spectral components in the band are below an auditory masking threshold [10,11]. If this is not satisfied, the frame is skipped, and the next frame that meets the condition is chosen.

2.2 Data Embedding

The proposed method uses dither modulation in the frequency domain. It may be viewed as an application of the multi-carrier modulation technique OFDM. An OFDM signal is composed of many equally spaced subcarriers within the occupied band, which are modulated using various modulation schemes. Suppose there are N symbols, $X(n)$, $n = 0, 1, \dots, N-1$, modulating N subcarriers respectively. The spacing between subcarriers, Δf, is chosen such that the subcarriers are mutually orthogonal within one symbol period, T. The requirement of orthogonality is satisfied if $\Delta f = 1/T$. Thus, subcarrier frequencies are $f_n = n/T$, $n = 0, 1, \dots, N-1$, and the OFDM signal in a symbol period is expressed as

$$x(t) = \sum_{n=0}^{N-1} X(n) \exp\left[j2\pi \frac{n}{T} t \right] \qquad 0 \le t \le T \ . \tag{1}$$

Sampling this waveform at intervals $\Delta t = T/N$ (sampling frequency $f_s = 1/\Delta t$) yields

$$x(k) = \sum_{n=0}^{N-1} X(n) \exp\left[j2\pi \frac{nk}{N} \right] \qquad k = 0, 1, \dots, N-1 . \tag{2}$$

So, $x(k)$ and $X(n)$ from a DFT pair. This means that the baseband OFDM waveform can be obtained from IDFT of the N modulating symbols. To obtain a real waveform in the time domain, the complex symbol series $\mathbf{X}=\{X(1), X(2), \dots, X(N-1)\}$ is extended to the negative frequencies to give a new series, $Y(n)$, of length $N_2=2N$ that is conjugate-symmetrical:

$$Y(n) = \begin{cases} X(n) & 0 \le n \le N-1 \\ X^*(N_2 - n - 1) & N \le n \le N_2 - 1 \end{cases}. \qquad (3)$$

IDFT of the vector **Y** is now a real vector of length N_2. To avoid redundant computation, two real vectors of length N_2 may be used as real and imaginary parts respectively to form a complex vector of the same length. Nonetheless, the negative frequency components are omitted for simplicity in the following discussion.

Among the N spectral lines (subcarriers) of the candidate frame, only $N/4$ are modified, using a combination of QAM and dither modulation as illustrated in Fig.1, by the embedded data while the other components are unchanged. The original n-th spectral component $\mathbf{C}_n$ is first quantized to $Q[\mathbf{C}_n]$ in the complex plane. The introduced distortion is determined by the quantization step Δ. The n-th watermark vector $\mathbf{W}_n$ is then added to $Q[\mathbf{C}_n]$ to produce a dithered spectral component $\mathbf{C}'_n$:

$$\mathbf{C}'_n = Q[\mathbf{C}_n] + \mathbf{W}_n \qquad (4)$$

where $\mathbf{W}_n$ is obtained using QAM, representing D=2 bits of the stego-data. Schemes other than QAM can also be used with different embedding capacity and robustness. Let the magnitude of $\mathbf{W}_n$ be $\Delta/2\sqrt{2}$ so that all coded data are located at centers of the grid quadrants as indicated by the circles in Fig.1. In the extreme case where $\Delta = \max[|\mathbf{C}_n|]$ thus $Q[\mathbf{C}_n]=0$, the host spectral components in the selected band is completely replaced by $\mathbf{W}_n$.

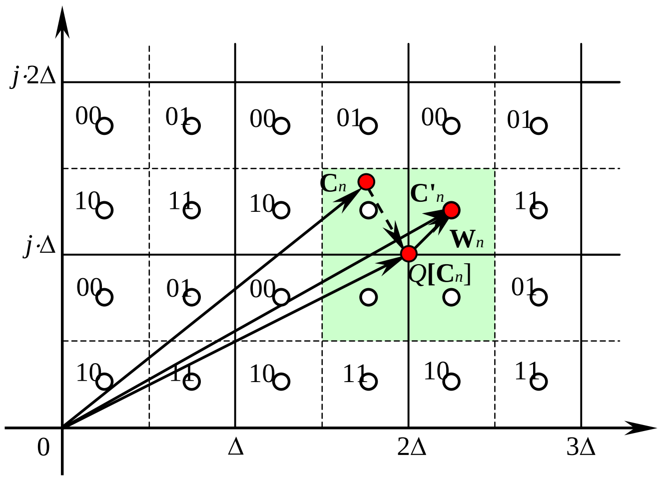

Fig. 1. Dither modulation in the complex frequency plane

2.3 Synchronization Pilot

Synchronization is essential to correctly recover the embedded data. A search process is used in the watermark detector to locate the encoded frame. For this purpose, a pilot signal is attached to the data as a part of the embedded sequence. The pilot must not take too large a portion of the watermark band and be easy to track. In the present system, it is composed of a number of symbols (1+*j*) and −(1+*j*) corresponding to an *m*-sequence of length *L* and occupies the lower part of the watermark band. The pilot is inserted into the signal spectrum in the same way as the mark symbols.

2.4 Structure of the Coded Signal Spectrum

The watermarked frame is composed of the preserved audio components *s*'(*k*), the embedded mark *m*(*k*), and a synchronization pilot *p*(*k*). The preserved audio contains most of the important frequency contents essential for imperceptibility, and the rest carries both the stego-data and the pilot. The spectrum of the composite signal is

$$X(n) = S(n)\,W_S(n) + \{M(n) + Q[S(n)]\}\,W_M(n) + \{P(n) + Q[S(n)]\}\,W_P(n) \tag{5}$$

where *S*(*n*), *M*(*n*), and *P*(*n*) are the signal spectrum, the mark symbols, and the pilot, respectively, and $0 \le n \le N-1$. Windows for the mark, the pilot and the preserved audio signal are defined, respectively, by

$$W_M(n) = \begin{cases} 1 & (N/4) + L \le n \le (N/2) - 1 \\ 0 & 0 \le n \le (N/4) + L - 1 \text{ or } (N/2) \le n \le N-1, \end{cases} \tag{6}$$

$$W_P(n) = \begin{cases} 1 & (N/4) \le n \le (N/4) + L - 1 \\ 0 & 0 \le n \le (N/4) - 1 \text{ or } (N/4) + L \le n \le N-1, \end{cases} \tag{7}$$

and

$$W_S(n) = 1 - W_M(n) - W_P(n) \qquad 0 \le n \le N-1 \ . \tag{8}$$

The mark, the pilot, and the quantization operator $Q[\cdot]$ are designed such that

$$Q\{M(n) + Q[S(n)]\} = Q[S(n)] \ , \tag{9}$$

and

$$Q\{P(n) + Q[S(n)]\} = Q[S(n)] \ . \tag{10}$$

Since the mark window can accommodate ($N/2-N/4-L$) symbols, and each complex symbol represents *D* bits (*D*=2 when using QAM), the number of stego-bits is $(N/4-L)D$. In the above example where the sampling frequency f_s=44.1kHz, N_2= 1024, and *T*=23.2ms, the watermark band can accommodate a total of $N_2/8 = N/4$ = 128 symbols including the mark and pilot when *B*=1/4. With QAM and *L* = 31, for example, the data capacity is 194 bits representing 27 ASCII characters.

2.5 Watermark Detection

The first step in watermark detection is to locate the encoded frame. This can be done by cross-correlating the pilot sequence with the spectral lines in $W_P(n)$ for each frame of the audio waveform. A correlation peak indicates that synchronization is achieved.

To speed up the search process, a replacement scheme may be used for the pilot sequence instead of dither modulation, and the search is carried out in the time domain. The price paid is a slight increase of distortion. In this method, a candidate frame is band-pass filtered to suppress spectral contents outside the embedded band, and then correlated with a locally generated pilot waveform that is the time domain representation of the pilot. A two-step search procedure is adopted: A coarse search is first carried out to quickly approach the peak, and then a fine search accurately locates the encoded frame. Since the pilot is a narrow band signal composed of a series of sinusoids, the correlation output $d(m)$ oscillates rapidly with m, as shown in Fig.2. Therefore the search is likely to fail since the possibility of falling into a pit between two peaks is high. To resolve the problem, magnitude of the correlation "envelope" may be used instead, which is obtained by taking difference between the maximum and minimum among several consecutive samples in $d(m)$. As soon as it exceeds a predefined threshold, a fine search is invoked.

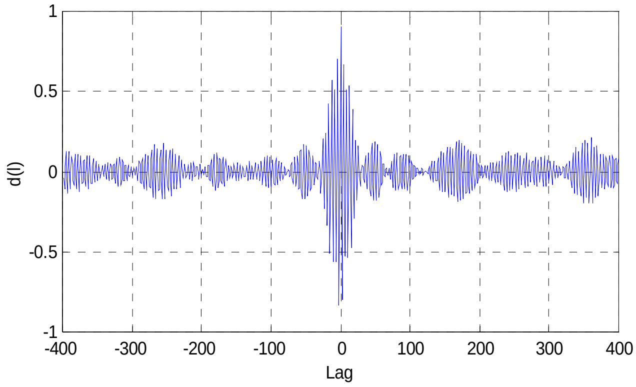

Fig. 2. Correlation between local pilot and the band-pass filtered audio with embedded data.

The key to an efficient search is an appropriate choice of search step size, determined by the correlation radius of the pilot obtainable from IDFT of the power spectrum density, $E^2(n)$, $n = 0, 1, \ldots, N_2-1$. Since the pilot in the frequency domain is composed of L spectral lines corresponding to an m-sequence, $E^2(n)$ is rectangular shaped whose width is given by

$$w_P = L\Delta f = \frac{f_S}{2N_2}L \quad . \tag{11}$$

So $e(m)$ is a sinc function. Define the half-width of the main lobe as correlation radius:

$$K = \frac{1}{w_P \Delta t} = \frac{2N_2}{L} \quad . \tag{12}$$

Thus, letting the search step be K, and choosing a threshold greater than, say, twice the highest sidelobe will ensure a reliable search. To further speed up the search process, the pilot is weighted with a Hamming window so that the main lobe is significantly broadened.

Having identified the encoded frame, the embedded symbols are recovered:

$$M(n) = S(n)W_M(n) - Q[S(n)W_M(n)] \; . \tag{13}$$

The information needed at the receiver includes the frame length N_2, the modulation technique used (here QAM), the mark band allocation, the quantization step, and the pseudo-random sequence for generating the pilot. These may form part of the key.

3 Experimental Results and Performance Study

3.1 Experimental System

A block diagram of the experimental system is shown in Fig.3. The OFDM subcarriers consist of the dither modulated spectral components and the unmodified signal spectral lines outside the embedding band. The complex watermark stream is obtained from a binary sequence using QAM. A 31-bit *m*-sequence is used as the pilot. After IFFT, a frame of marked waveform replaces the selected frame in the host.

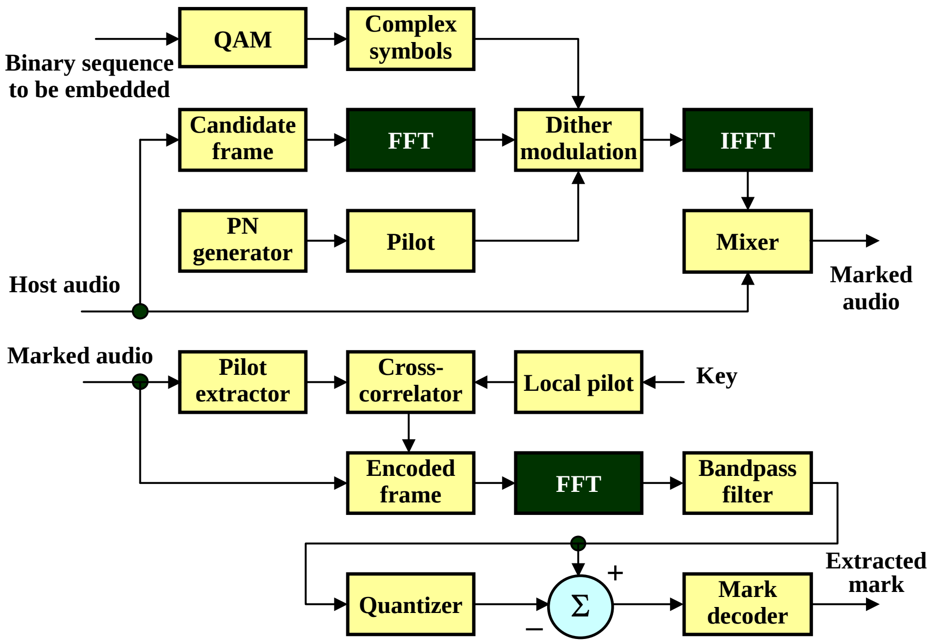

Fig. 3. Block diagram of the experimental system

At the receiver, search is performed either in the frequency domain or in the time domain. In the time domain approach, signal components in the known band is extracted with a 5th-order type I Chebyshev band-pass filter, and then cross-correlated with a locally generated pilot waveform. In order to obtain an accurate alignment, a dual-direction filter is used in the fine search to preserve the phase.

3.2 Embedding Capacity and Imperceptibility

Embedding capacity is a function of watermark bandwidth B, frame length N_2, bit number represented by each symbol, D, and length of the pilot, L_m, as given by

$$J = \left(\frac{N_2 B}{2} - L_m \right) D . \quad \textbf{(14)}$$

With the proposed technique, it is possible to embed several hundred bits into a single segment lasting 20~30ms. For a given embedding capacity, the higher the sampling frequency hence the bandwidth is, the shorter the required frame length. For high fidelity music, the frame length can be very short so that the effect of data hiding on sound quality is small.

Watermarks are usually embedded into a number of frames in the host audio. These frames are organized into groups, and a chained structure is use to avoid lengthy searches. The synchronization pilot was only inserted into the first frame of each group with position information of the next frame contained in the embedded data. The band assigned to the pilot was thus used to carry a pointer for the subsequent frame. Choosing the minimum spacing between frames as $32N_2$ where N_2 = 1024, a total of 20 candidate frames were identified in a segment of Radetsky March lasting 23.77s, with f_s = 44.1 kHz, 16 bits per sample, and embedding bandwidth $w = W/4$. This resulted in a payload of more than 3,800 bits when using QAM, that is, more that 540 ASCII characters, or nearly 25 characters per second.

The embedding induced distortion is a function of the quantization step Δ. Fig.4 shows waveforms of a signal frame before and after embedding, with Δ= max(|C_n|)/8, where max(|C_n|) is obtained from a representative signal section. The Δ value should be included in the key. The two waveforms are hardly distinguishable. The difference between them, very close to the horizontal axis, is also shown. Table 1 presents SNR of the marked audio frame from Radetsky March with different quantization steps.

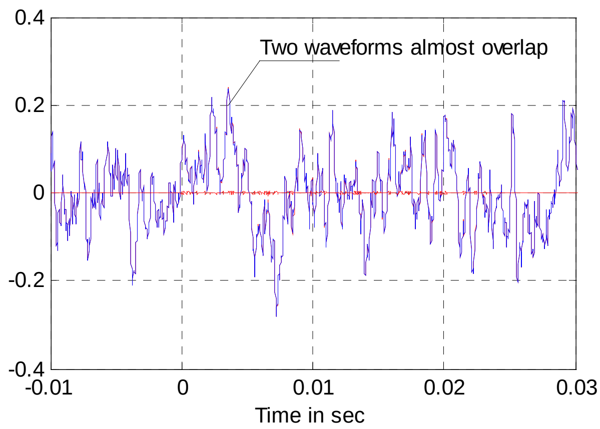

Fig. 4. Waveforms of a signal frame before and after embedding, and their difference

Table 1. Signal-to-noise ratio of the embedded frame

Δ	max(|C_n|)	max(|C_n|)/2	max(|C_n|)/4	max(|C_n|)/8	max(|C_n|)/10
SNR(dB)	18.68	24.28	29.99	35.98	38.12

Signal-to-noise ratios of the entire music, as a metric to assess imperceptibility, are listed in Table 2. The largest quantization step was used in this experiment, (complete replacement of the spectral lines within the embedding band). In the table, f_s is the sampling frequency, N_q number of bits per sample, T length of the music, N_f number of embedded frames, and N_b the total number of embedded bits. Even with the largest quantization step, the introduced distortion is inaudible. A subjective test on several music clips was carried out. In each piece, a number of frames were identified using the HAS criterion, and data were embedded into them with various quantization steps. Using a procedure based on the ABX method [12], a group of 10 people were independently asked to listen to the original and the modified versions (A and B) of each piece in a random order, and then listen once more to a randomly chosen one (X). They were asked to tell whether X is A or B. The rates of correct identification were roughly 50%, indicating that the data embedding is imperceptible. In contrast, adding white Gaussian noise at similar levels is clearly audible to most listeners.

Table 2. SNR of embedded pieces. Dither steps: Δ_1=max(|C_n|), Δ_2=max(|C_n|)/8

Host audio	f_s (kHz)	N_q (bits)	T (sec)	N_f	N_b	SNR (dB)	
						Δ_1	Δ_2
I: Classic	44.10	16	23.77	36	9,216	32.02	39.43
II: Classic	44.10	16	47.74	70	17,920	42.13	47.24
III: Pop	44.10	16	25.52	45	11,520	32.66	41.29
IV: Pop	44.10	16	46.83	87	22,272	31.63	41.32
V: Speech	22.05	8	3.47	8	2,048	31.80	41.20
VI: Speech	22.05	8	2.51	6	1,536	33.21	36.45

3.3 Robustness Test

Tests for robustness against attacks such as AWGN interference and MP3 coding were performed on audio pieces watermarked with the largest quantization step.

Additive Noise Interference. AWGN was added to the marked audio. Fig.5 shows the constellation of the extracted stream with QAM watermark data and a Hamming windowed pilot. Ideally, the watermarks should all appear at four points in the complex plane: $1+j$, $-1+j$, $-1-j$, and $1-j$, as indicated by the thick dots. The scattered circles represent a noise-contaminated signal at SNR=30dB referenced to the average power of the waveform. Clearly, synchronization and accurate decode of watermark symbols can be achieved as long as the symbols remain on the correct quadrants.

Progressively increasing noise caused errors to occur, until the search or decoding failed. Fig.6 gives the relation between SNR and the bit error rate. Three types of signals were used in the experiment: (1) hi-fi music with f_s = 44.10 kHz, (2) speech with f_s = 22.05 kHz, and (3) low quality music or speech with f_s = 8.0 kHz. The embedding bandwidth was W/4. The results show that noise tolerance mainly depends on the modulation scheme, essentially the number of bits contained in a symbol, D, and to a much less extent on sampling frequencies and the particular type of signal.

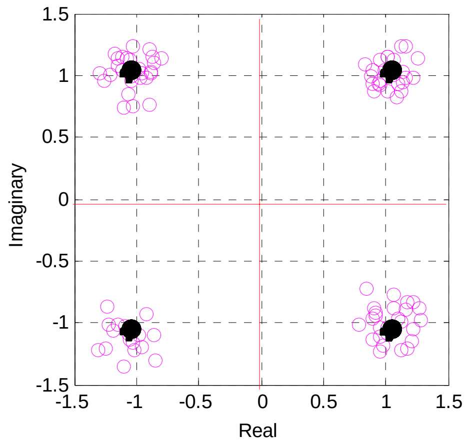

Fig. 5. Constellation of OFDM symbols. Scattered circles are the noise-contaminated signal.

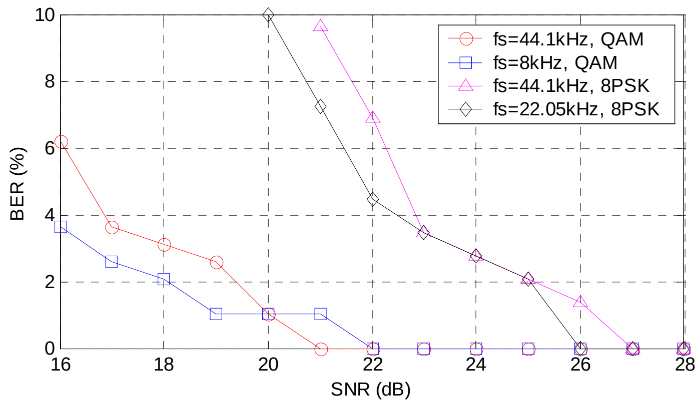

Fig. 6. Relationship between SNR and BER

Linear Filtering. The watermarked signal was passed through a low-pass filter prior to detection. A 9th-order Butterworth filter was used. For any type of signal and modulation scheme, error-free recovery of the embedded data was achieved provided the cut-off frequency was above the high end of the watermark band.

MP3 Coding. Robustness against MP3 coding is important for audio watermarking. Five pieces of hi-fi music (I~III: classic music, IV and V: pop songs) were tested in the experiment. Parameters were the same as that in Table 2, with = 1. In Table 3, the bit error rates obtained at different compression rates and for different music are given. Two BER values are shown in each case, where the left and right values correspond to the watermark bands [W/4, W/2] and [W/4, 3W/8], respectively.

It is concluded from this experiment that, when using QAM, the system is robust against MP3 at bit rates as low as 64 kbps～80 kbps, depending on the assignment of the watermark band. With the narrower band, the embedded data was extracted without error at MP3 bit rate of 64 kbps per sound channel. When using BPSK, error-free extraction was achieved for all the 5 tested host signals even at the MP3 bit rate of 56 kbps and with a wider watermarking band.

Table 3. Robustness against MP3: BER(%) at different MP3 bit rates. Left and right BER values were obtained with watermark bands [W/4, W/2] and [W/4, 3W/8] respectively.

MP3 bit rate	128 kbps	112 kbps	96 kbps	80 kbps	64 kbps	56 kbps
I	0/0	0/0	0/0	0.52/0	2.06/0	5.67/1.55
II	0/0	0/0	0/0	0/0	2.06/0	1.55/0
III	0/0	0/0	0/0	0/0	3.09/0	3.61/1.03
IV	0/0	0/0	0/0	0/0	0.52/0	2.58/0.52
V	0/0	0/0	0/0	0/0	0.5 /0	2.06/0

4 Conclusions

Using a frequency domain dithering technique, a substantial amount of information can be embedded into a digital audio signal. In this technique, a data sequence is encoded and inserted into the spectrum of short frames of the signal. A high degree of imperceptibility is achieved by utilizing the HAS both in the time domain and in the frequency domain. With a large quantization step, the system is sufficiently robust against additive white Gaussian noise and MP3 compression coding.

When the quantization step becomes small, better transparency, but less robustness, results. This is considered to be suitable for covert communication applications, and should be subject to both perceptive and statistic analysis. It has been found that, with a small quantization step, say, $\max(|C_n|)/8$, the modifications to the waveform of the affected frame as shown in Fig.4 is in fact well beyond several least significant bits. Therefore, LSB based steganalytic techniques cannot be used to detect the presence of the data embedding. Moreover, since the frames are sparsely scattered, locating signal segments that likely contain secrete information without the knowledge of the synchronization pilot is extremely difficult. Further study in this aspect is required.

A number of parameters can be varied to meet different requirements. For example, choosing a short frame length and a narrow watermark band toward lower frequencies can make the watermark more robust. The embedded data can be repeated in the candidate frames over the host signal if only a few data are to be embedded. On the other hand, if data capacity is important, a longer frame should be used, and a more efficient modulation scheme such as 16QAM (D=4) can be chosen. Error correction techniques may also be introduced with a moderate reduction of payload.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 60072030), and Key Disciplinary Development Program of Shanghai (2001-44).

References

1. M. D. Swanson, et al., "Multimedia Data-Embedding and Watermarking Technologies," *Proc. IEEE*, vol. **86**, 1998: pp.1064-1087.
2. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for images, Audio and Video," *IEEE Trans. Image Processing*, vol.**6**, 1997: pp.1673-1687.
3. N. F. Johnson, et al., *Information Hiding: Steganography and Watermarking — Attacks and Countermeasures*. Kluwer Academic Publishers, 2000
4. M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust Audio Watermarking Using Perceptual Masking," *Signal Processing*, vol. **66**, 1998: pp.337-355.
5. W. Bender, et al., "Techniques for Data Hiding," *IBM System Journal*, vol.**35**, 1996: pp.313-336
6. P. Bassia, I. Pitas, and N. Nikolaidis, "Robust Audio Watermarking in the Time Domain," *IEEE Trans. Multimedia*, vol.**3**, 2001: pp. 232-241
7. B. Chen, and G. Wornell., "Quantization Index Modulation: a Class of Provably Good Method for Watermarking and Information Embedding," *IEEE Transactions on Information Theory*, vol. **47**, no.4, 2001: pp.1423-1443
8. J. F. Tilki, "Encoding a Hidden Digital Signature Using Psychoacoustic Masking," Thesis submitted to the Faculty of the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, June 9, 1998
9. R. V. Cox, et al., "Scanning the Technology: On the Applications of Multimedia Processing to Communications," *Proc. IEEE*, vol.**86**, 1998: pp.755-824.
10. J. D. Johnston, "Transform Coding of Audio Signal Using Perceptual Noise Criteria," *IEEE J. Select. Areas Commu,.*, vol.**6**, 1998: pp.314-323
11. D. Tsoukalas, J. Mourjopoulos, and G. Kokkinakis, "Speech Enhancement Based on Audio Noise Suppression," *IEEE Trans. Speech Audio Processing*, vol.**5**, 1997: pp.497-514.
12. E. Brad Meyer, "ABX Tests and Testing Procedures," *Boston Audio Society Speaker*, vol.**19**, no.3, 1990 http://bostonaudiosociety.org/bas_speaker/abx_testing.htm