Spread Spectrum Audio Steganography using Sub-band Phase Shifting

Hosei Matsuoka
Research Laboratories, NTT DoCoMo, Inc.
matsuoka@nttdocomo.co.jp

Abstract

This paper presents an improvement of spread spectrum audio data hiding methods. We introduce phase shifting in audio signals to reduce the correlation with PN signal per each sub-band. It allows easy detection of the embedded data signal from audio when de-spreading the compound signal. The paper reports the subjective test results and the measurements of noise resiliency. The proposed method generates the quality degradation at the same level of NMR +3dB, but accests +6dB noise, therefore, the method has 3dB benefits.

1. Introduction

Because of the increasing popularity of mobile handheld devices that offer high level audio functions, there is a strong possibility of using sound for data transmission from ordinary loudspeakers to the handheld devices. While the data rates for sound are relatively low compared to media such as radio, sound links can establish an attractive interface for service access. That is, a commercial on the TV or radio can be used to load the URL of the company selling the item into the user's mobile device. Sound is attractive for such applications since no additional hardware infrastructure is needed.

We propose a method of audio steganography based on a spread spectrum data hiding technique. The method uses frequency masking effect based on the psycho acoustic model to embed spread spectrum data signal into audio materials. However, at the receiver side, it is difficult to retrieve the data signal from the received audio signal when the correlation between the embedded data signal and the original audio signal is not low. Our proposal is to apply phase shifting to the original audio material to lower the correlation with the data signal, thus reduce the error probability of receiving data. This paper evaluates the degradation of audio quality and the reduction of error probability of data transmission resulted from the proposed phase shifting, and reports the 3dB benefits of the noise resiliency in total.

2. Audio Data Hiding

The main factor of audio data hiding is that the encoded audio material is acoustically indistinguishable from the original audio. Several audio data hiding techniques based on the HAS (Human Auditory System) have been proposed in recent years.

In the method of Echo Hiding [1], motivated by the fact that the HAS (Human Auditory System) cannot distinguish an echo from the original when delay and amplitude of the echo are appropriately controlled, it employs two different delay times to carry binary information. Phase Coding [2] embeds data by altering the phase in a predefined manner. To a certain extent, modifications of the phase of a signal cannot be perceived by the HAS. This can be done by using all-pass filters to modify the phase without changing the magnitude response. In the method of using spread spectrum [3], data signal is spread across a wide frequency and embedded in the audio by using the frequency masking effect by which a faint sound becomes inaudible in the presence of another louder sound. The decoder can retrieve the embedded information by dispreading the received signal with the same PN (Pseudo-random Noise) code as is used in encoding. [4] reports the audio quality and data transmission rate of the above techniques.

The second important factor is the robustness of the embedded information. There are many kinds of transmission environments that a signal might experience on its way from encoder to decoder. The robustness means to keep the embedded information over such transmission environments. [5] categorizes the transmission environments to 4 patterns as described in Fig. 1. The first is the digital end-to-end environment where signals are never modified in any way (A). This class has the least constraints on data-hiding methods. The next class is when a signal is re-sampled to a higher or lower sampling rate, but

Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 IEEE

remains digital throughout (B). This class preserves the absolute magnitude and phase, but changes the temporal characteristics of the signal. The third case is when a signal is played into an analog state, transmitted on a clean analog line and re-sampled (C). Absolute magnitude, sample quantization are not preserved, but phase will be preserved. The last case is when the signal is played into the air and re-sampled with a microphone (D). The signal will be subjected to possibly unknown non-linear modifications resulting in phase changes, amplitude changes, frequency drifts, echo, etc.

Watermarking generally supports the robustness up to (C) to protect copyrights or prohibit unauthorized copying. However, steganography should support the robustness up to (D) for aerial data transmission.

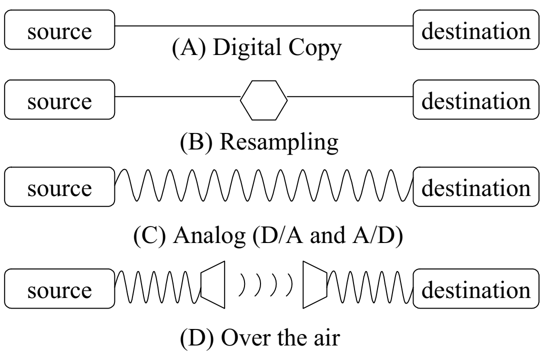

Fig.1. Signal Transmission Environments

3. Spread Spectrum Audio Steganography

Basically this method calculates the frequency masking threshold using psycho acoustic model, data signal is spread by a M-sequence code, and the spread signal is embedded in audio below the frequency masking threshold. M-sequence codes have good autocorrelation properties where the autocorrelation function has peaks equal to 1 at 0, N, 2N… (approximately 1/N elsewhere). Because of these periodic peaks, the M-sequence code is self-clocking, so the receiver can easily synchronize the data frame and retrieve the embedded data by de-spreading with the same M-sequence code.

Fig.2. shows an example of the original audio spectrum and the frequency masking threshold. The frequency masking threshold is calculated at each critical band based on the psycho acoustic model [6].

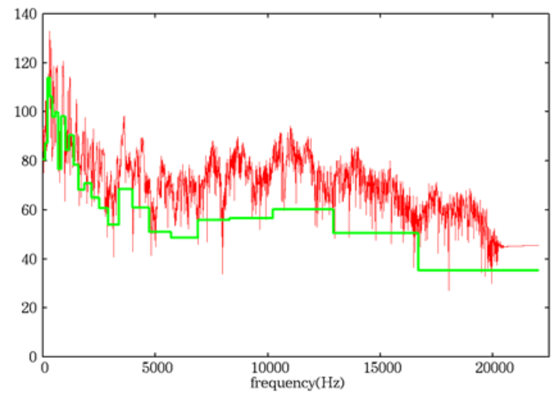

Fig.2. Frequency Masking Threshold.

Next, data signal is spread by an M-sequence code. The spectrum of the spread signal is shown as Fig.3. The spectrum is almost flat in the wide frequency range. Compared to the masking threshold, some bands have higher spectrum of the spread data signal than the masking threshold, other bands have much lower spectrum than the threshold. The higher spread signal is audible to human ears, and the lower spread signal can be still increased up to the threshold.

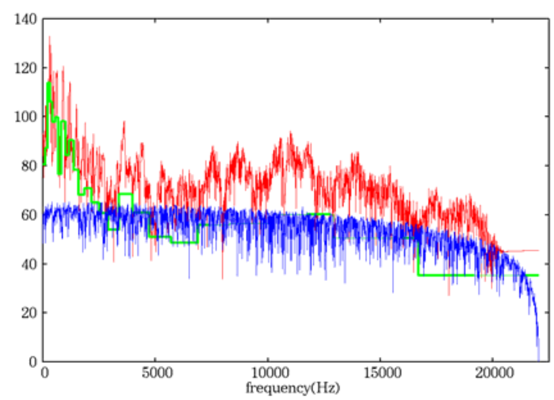

Fig.3. Spread Data Signal.

The spread signal of the flat spectrum is filtered with the masking filter that has the same amplitude response as the masking threshold. The spectrum of the output signal has the same spectrum as the masking threshold. Then the output signal is power-adjusted below the masking threshold and added to the original audio. Fig.4 shows the spectrum of the output signal. In this way, the power of the spread signal can be maximized below the frequency masking threshold.

Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 IEEE

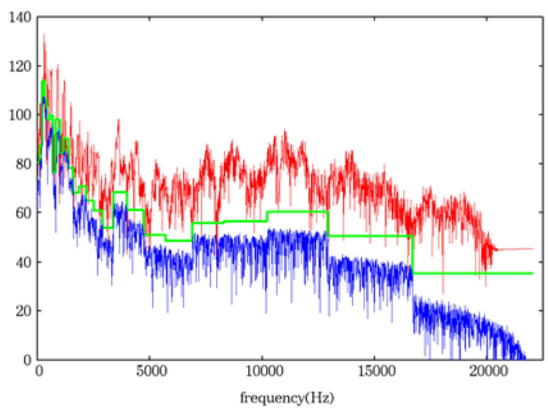

Fig.4. Colored Spread Signal.

4. Sub-Band Phase Shifting

The spread spectrum audio steganography reduces the error probability by increasing the spreading rate and coding gain. However, high spreading rates take long time to send 1 bit information, so the data transmission speed decreases. Therefore, in order to improve the performance, the encoding method with a low spreading rate has to provide the good robustness of the embedded information. Table.1. shows the relations between spreading rate and transmission speed. The sampling rate is 44.1kHz, and the chip rate of the spread spectrum is 22.05kHz. In the case that the spreading rate is 511, 1 bit data can be embedded in a frame of 1022 samples, and the transmission speed is approximately 40bps. To recover bit errors at the receiver side, the redundant bits are necessary for error correcting. The coding rate of the error correcting code is 8.2% to provide the robustness of the embedded information against the aerial transmission at a distance of 1m. Therefore, the effective data transmission speed becomes 3.3bps.

Table.1. Spreading Rate and Transmission Speed.

Spreading Rate	511	1023	2047	4095
Coding Rate	8.2%	35.6%	70.1%	93.0%
Trans. Speed	40bps	20bps	10bps	5bps
Effective Trans. Speed	3.3bps	7.1bps	7.0bps	4.6bps

When the correlation between original audio signal and the spread data signal is low, even if the spreading rate is low, it can keep the low error probability and increase the effective transmission speed. This paper proposes a method of decrease the correlation by using phase shifting to each sub-band signal of original audio, and verifies its effect.

4.1. Correlations in each sub-band.

Fig. 5 shows the correlations in each sub-band separated by 1kHz bandwidth. For each sub-band, the sub-band signal is shifted from 0 to 5 samples and calculated the correlation with the spread data signal. At the first sub-band (0-1000Hz), the correlation is likely to high and the phase shifting is not so effective to reduce the correlation. However, at the upper sub-bands, the correlations are not so high and the phase shifting can adjust the correlations. Therefore, for the sub-bands higher than 1kHz, the encoder should apply the phase shifting to reduce the correlations, and embed the spread data signal in the synthesized audio signal after the phase shifting. Then, at the receiver side, the recorded signal lower than 1kHz should be eliminated and the signal remained should be decoded.

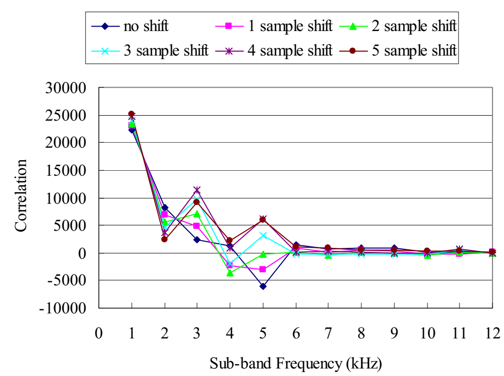

Fig. 5. Correlations at each sub-band

5. Subjective Quality Testing

We evaluated the audio quality degradation of the proposed scheme with the subjective testing. The testing employs the comparison between the original audio and the following materials.

(1) Data embedded audio without the phase shifting.
(2) Data embedded audio with the phase shifting.
(3) Bandwidth limited original audio to 7kHz.
(4) Bandwidth limited original audio to 3.5kHz.

For the case of (1) and (2), the testing is made by setting the NMR (Noise Masking Ratio) that is the embedded data signal ratio compared to the masking threshold to -3dB, 0dB, +3dB, +6dB. 20 examinees had a listen and categorized the quality degradation for each audio material according to the DCR

(Degradation Category Rating) indication. Fig. 6 shows the average score of each audio material.

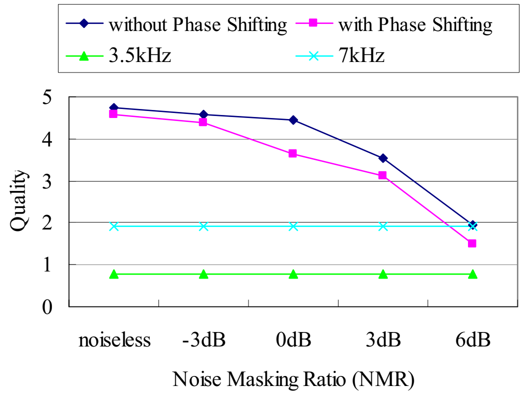

Fig. 6. Subjective Testing Results

In the case of (3) and (4), since the spread data signal is not embedded, the scores were respectively 1 and 2 regardless of the NRM. Comparing (1) and (2), because of the phase shifting, the quality of (2) is (1) lower than that of (1) at every NMR. The quality difference is almost equal to the degradation of +3dB NRM. Therefore, if (2) has higher noise tolerance than +3dB of (1), (2) has better performance in total. The measurement of the noise tolerance is described in the next section below.

6. Noise Tolerance

We compared the noise tolerance of the encoded materials (1) (without phase shifting), and (2) (with phase shifting), where the spreading rate is 511 and the NMR is 0dB. We added white gaussian noise to both audio materials to simulate the noise in the transmission environments, and evaluated the relations between S/N and BER (Bit Error Rate) shown in Fig. 7. The value of S/N is the signal ratio of the encoded audio materials compared to the added white noise. The BER of (2) is lower than that of (1) in the same S/N. The gain of the noise tolerance is approximately +6dB. In the previous section, (2) has the quality degradation which is equal to +3dB NMR. Therefore, even if (2) suppresses the spread data signal to -3dB NMR to provide the same audio quality as (1), the noise tolerance is still +3dB higher than that of (1).

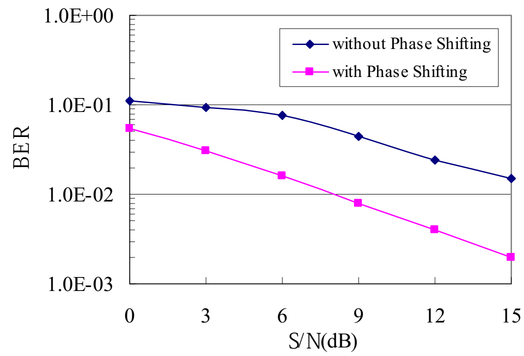

Fig.7. S/N and BER.

7. Conclusions

This paper proposes the sub-band phase shifting as a method of processing the original audio signal so that the data signal can be easily retrieved at the receiver. Except the proposed scheme, it is considered that the pre-processing methods where the original audio signal is processed before embedding is effective. In the proposed scheme, the total gain of 3dB was verified from the results of the subjective quality testing and the noise tolerance simulation. The proposed scheme can provide the robust aerial data transmission compared to the traditional schemes.

References

[1] D.Gruhl, A.Lu and W.Bender, “Echo Hiding”, Information Hiding 1996, pp.295-315.

[2] Y.Yardimci, A.E.Cetin and R.Ansari, “Data hiding in speech using phase coding”, ESCA. Eurospeech97, pp.1679-1682, Greece, Sept. 1997.

[3] L.Boney, A.H.Tewfik and K.N.Hamdy, “Digital watermarks for audio signals”, IEEE Intl. Conf. on Multimedia Computing and Systems, pp.473-480, March 1996.

[4] J.D.Gordy and L.T.Bruton, “Performance Evaluation of Digital Audio Watermarking Algorithms”, IEEE MWSCAS 2000.

[5] W.Bender, D.Gruhl, N.Morimoto and A.Lu, “Techniques for data hiding”, IBM Systems Journal, vol 35, no. 3-4, pp.313-335, 1996.

[6] J.Johnston, “Transform Coding of Audio Signals Using Perceptual Noise Criteria”, IEEE Journal on Selected Areas in Communication, vol.6, pp.314-323, February, 1988.

Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 IEEE