

AUDIO STEGANOGRAPHY USING BIT MODIFICATION

Kaliappan Gopalan

Department of Engineering, Purdue University Calumet, Hammond, IN 46323

ABSTRACT

A method of embedding a covert audio message in a cover utterance for secure communication is presented. The covert message is represented in a compressed form with possibly encryption and/or encoding for added security. One bit in each of the samples of a given cover utterance is altered in accordance with the data bits and a key. The same key is used to retrieve the embedded bits at the receiver. The results, based on cover signals from a clean TIMIT utterance and a noisy aircraft cockpit utterance, show that the technique meets several major criteria for successful covert communication.

1. INTRODUCTION

Audio steganography is a useful means for transmitting covert battlefield information via an innocuous cover audio signal. The two primary criteria for successful embedding of a covert message are that the stego signal resulting from embedding is perceptually indistinguishable from the host audio signal, and the embedded message is recovered correctly at the receiver. Other requirements such as robustness of embedding, data recovery without the original cover signal, etc. may depend upon the type of applications. Additionally, for covert communication of key information, the embedded information must withstand channel noise and intentional attacks or jamming on the signal. Also important is the resilience of the hidden information to stay hidden to pirates during their attempts at detection.

If the embedded message is another audio signal, one may perceive the intended message even if some errors are encountered during the recovery. This is particularly important in covert communication as the stego signal may be received with intentional and unintentional changes due to noise in the channel or attacks. The primary goal here is to convey the concealed message albeit with reduced speech quality. Additionally, some degradation in the perceptual quality of the stego signal from that of the original host signal may be

acceptable. If the host used for carrying the covert message is not a common or familiar audio signal, the degraded quality of the stego signal may not be noticeable by attackers; hence, the presence of information hidden in the stego may stay imperceptible and, consequently, impervious. Based on this premise, a method of data embedding by manipulation of the host samples in accordance with the covert information is described in this paper.

2. STEGANOGRAPHY USING BIT MODIFICATION OF HOST SAMPLES

Data embedding by bit modification of host samples is a common technique in image embedding [1 - 3]. Typically, the least significant bit of every pixel in the host image is altered in accordance with the data bits and a key. With the low sensitivity of the human visual system to luminance – about one part in 30 for random patterns and approximately one part in 240 in uniform regions of an image – modification of the least significant bit has been shown to meet both the criteria of being perceptually indistinguishable and being able to correctly retrieve the embedded data. In addition, the payload capacity of the method is large with one bit for every host sample or pixel.

In general, direct extension of the bit modification technique to host audio signals is precluded by the higher sensitivity and dynamic range of the human auditory system (HAS) compared with the visual system. With a large power and dynamic range, the human ear can detect a change in an audio file as low as one part in 10 million. In addition, the HAS can perceive a frequency range of one thousand to one [2]. Thus, any change due to embedding in an audio file must be extremely small to prevent the detection of the existence of the hidden information; alternatively, the change must be occurring at points that are masked out by their strong neighbors in the original host audio. Since bit modification of a sample cannot be directly related to a particular frequency, concealing of embedding by frequency masking is not possible in all cases. Instead, modification of bits at lower significant levels is likely to result in an imperceptible stego signal if the dynamic range of the host is large.

Alternatively, more significant bits in host samples can be modified with a tolerable level of degradation in speech quality if the payload is small. These are presently under investigation and the initial results are presented here.

3. EXPERIMENTAL RESULTS

In the first experiment a female utterance from the clean TIMIT database was used as the host or cover signal. The covert message consisted of 2,800 bits obtained from the standard Global System for Mobile communication half-rate coder (GSM 06.20) for a short male utterance. The host samples are first normalized to 16-bit unipolar values. With 53,556 samples of 16 bits each in the host speech that was used, the covert message bits can clearly be embedded anywhere in the utterance. For further security, a key of 256 bits is used at the transmitter and a copy of the key is made available at the receiver.

3.1 Direct Embedding

The GSM-coded covert speech data of 2,800 bits are embedded in the middle of the host utterance using the key as follows. At each sample of the host starting at 25,379, the least significant bit (lsb) is modified as

$$b_s(n) = b_o(m) \oplus b_k((n \bmod 256)+1), \quad (1)$$

for $m = 1$ to 2800 and $n = 25,379$ to 28,178, where $b_s(n)$ is the new value of the lsb of the n^{th} (stego) sample, $b_o(m)$ is the value of the m^{th} message bit, and b_k is the key bit from the 256-bit key at $(n \bmod 256) + 1$. (\oplus refers to the exclusive OR operation.)

The bit-modified host becomes the data embedded stego signal that is transmitted with or without overlapping frames.

At the receiver, the embedded bits $\{ b_r(m) \}$, are recovered from the stego samples $\{ b_s(n) \}$ starting at $n = 25,379$ by reversing the embedding procedure as

$$b_r(m) = b_s(n) \oplus b_k((n \bmod 256)+1), \quad (2)$$

for $n = 25,379$ to 28,178 and $m = 1$ to 2,800.

The above embedding and recovery procedures were carried out for (a) the lsb, (b) 5th least significant bit, and (c) a higher order (the 10th or 11th least significant) bit of the host samples. In all cases, the data recovery was correct in all bits; hence, the feasibility of error-free covert communication was established for noise-free transmission environment with no intentional or unintentional

operations on the stego signal. In the lsb modification, the perceptual quality of the stego signal could not be distinguished from that of the original host utterance in informal listening tests. Additionally, the spectrograms of the host and stego were also indistinguishable, as seen in Fig. 1. We note that with the exclusive OR operation, not all of the 2,800 of the host samples may have changed in their least significant bits. Hence, the histogram of the received stego signal amplitudes – a common technique in steganalysis to determine embedding when the host is available for comparison – also looked identical to that of the host signal. (The histograms of the amplitudes were obtained with the 16-bit range of 0 to 65535 divided into 256 equally spaced values.)

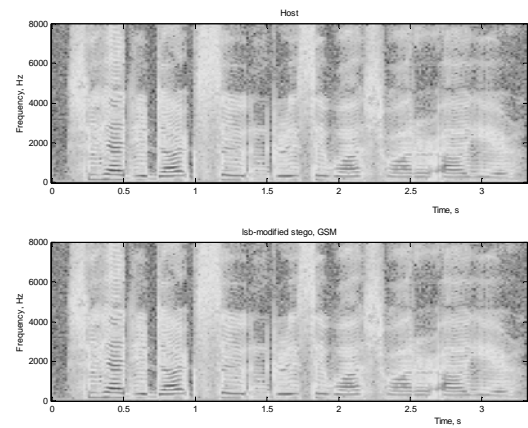


Fig. 1 Spectrograms of host (top) and lsb-modified stego (bottom) with 2800 bits of data

As observed, it is not surprising that the stego signal cannot be perceived any differently from the host audio with only the lsb modified. In practice, however, the lsb of the received stego may not remain the same because of filtering. To minimize the loss of covert data, therefore, higher order bits of the host were modified for embedding. In the case of bit 5, the spectrogram of the stego still could not be distinguished from that of the host. The expanded spectrogram of the stego where bits were modified (in samples of 25,379 to 28,178) showed a barely discernible difference from that of the host. This difference, even if the host is available for comparison, may not be enough to cause any suspicion about the presence of covert data. A 256-bin histogram of sample amplitudes of the stego was also indistinguishable from that of the host.

Modifying the 10th (or the 11th) significant bit, however, altered the spectral characteristics of the stego signal noticeably by adding high frequency noise in the region of embedding as seen from Fig. 2. In particular, the spectrum of a frame where the 10th bit was modified (Fig. 3) shows significant high frequency components relative to the unmodified host frame spectrum.

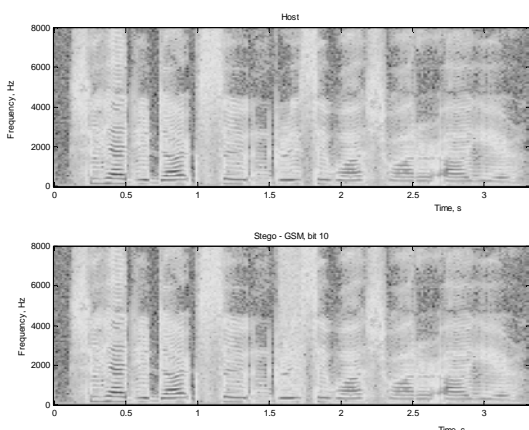


Fig. 2 Spectrograms of host (top) and 10th bit-modified stego (bottom) with 2800 bits of data

While the change in the spectral content is noticeable in comparison with that of the host, it is not significant enough to cause much audible difference. Here again, the embedded data and the randomly generated key may be such that not many samples were affected at the 10th bit position to alter the perceptual quality. In the case of a noisy host (cover) utterance from a database of cockpit voice recordings (the Greenflag database), the change in the spectral content or in the speech quality of the stego was too insignificant to be noticed even at higher order bit modification.

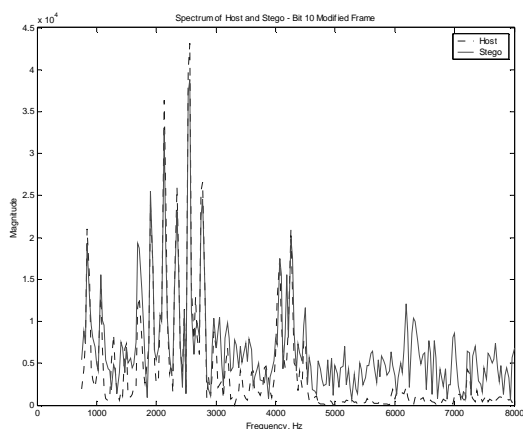


Fig. 3 Spectra of a frame of host and stego where bit 10 was modified

To reduce the concentration of spectral changes which may enable detection of embedding, modification may be carried out at every 1st host sample. Fig. 4 (bottom) shows the spectrogram of the stego in which the 10th bit of every 15th sample was modified. Clearly, little difference is discernible between the stego and host (top) spectrograms. Additionally, a comparison of the

histograms of the host and stego also showed no significant difference that can be used in steganalysis.

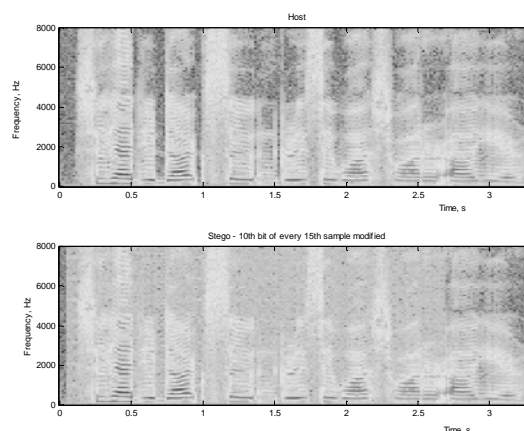


Fig. 4 Spectrograms of host (top) and 10th bit of every 15th sample modified stego (bottom)

Spreading of embedding, of course, is possible only if the data size is very small relative to the number of host samples. Therefore, only compressed speech can be used as hidden data in covert message communication.

3.2 Spread Spectrum Embedding

Because of the large payload available – one bit in every host sample – embedded data can be made secure and robust by encryption or encoding. To retain data under noisy transmission conditions, for instance, each bit may be repeated many times before embedding. The receiver can retrieve each correct bit by a majority voting. Alternatively, error detection and correction bits may be appended and the augmented, or coded data, can be embedded. Security and data integrity can be further increased by spreading the data using a spread spectrum technique. In all these cases, however, the size of data to be embedded increases; hence, the spectral distortion and perceptual difference is likely to become significant, particularly if a higher order bit is modified.

As an example, the GSM-coded data (covert speech) bits were repeated 15 times and these spread data were used to embed in (a) bit 5 and (b) bit 10. (Host samples in the middle were used for embedding.) Although the payload was $2800 \times 15 = 42000$ bits, giving $42000/53556 = 78.42$ percent of embedding capacity, a change in bit 5 did not result in any noticeable distortion in spectrogram. Modifying bit 10 clearly showed a significant difference in the spectrogram and perceptual quality of the stego by introducing wide band noise across the utterance duration. Fig. 5 shows the three spectrograms for comparison. This degradation in

perceptual quality may be detrimental to covert communication in which a known cover signal is used as a

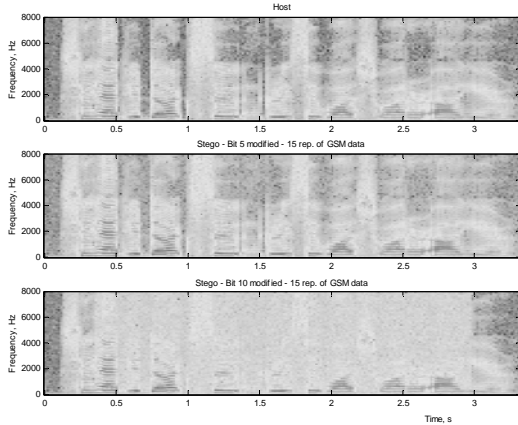


Fig. 5 Spectrograms of host (top), stego with 5th bit modified (middle), and 10th bit modified (bottom) in over 78 percent of host samples

carrier and a higher order bit is modified. In practice, however, the repetition (data spread) rate, bit order and host length can be varied to conceal the existence of embedded information.

3.2 Robustness of Stego

Additive noise on the transmitted stego has less effect on the speech quality compared to that of the original stego if a higher order bit was modified. This is to be expected because of the significantly large effect of bit modification relative to the added noise. Correct data retrieval, in all cases including repetitive bit embedding, remained at 100 percent. At lower significant bit modification – at bit 5, for instance – bit errors resulted after adding zero-mean Gaussian noise to the stego while the perceptual quality of the noisy stego stayed the same as that without the added noise. Thus the stego is robust to retaining the embedded data if a more significant bit – such as bit 11 – is used for modification at a cost of degraded speech quality. The trade off between perceptually indistinguishable and data-robust stego can be seen in Fig. 6. The spectrogram of the noise-added stego with bit 5 modified in accordance with 15 repetitions of the data is indistinguishable from that of the host while the stego with bit 11 altered shows wideband noise spread across the entire duration. Recovered data, however, had over 130 bit errors with bit 5 modification while bit 11 modification resulted in exact recovery.

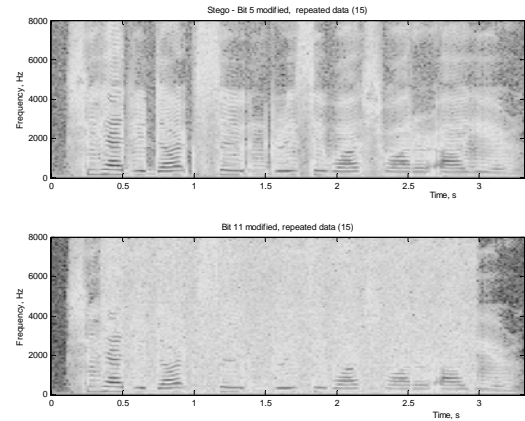


Fig. 6 Spectrograms of noise-added stego with 15 repetitions of 2800-bit data with embedding in (a) bit 5 (top) and (b) bit 11 (bottom)

4. CONCLUSION

A method of audio steganography by modifying host audio samples in the time domain has been presented. By altering one bit in each sample of a host utterance in accordance with a given covert message and a key, the payload capacity can be made quite large. The stego is robust to retaining the embedded data at low levels of additive noise. While the bit alteration is not specifically carried out in perceptually masked regions, imperceptibility of embedding in the stego has been observed for the combination of host utterances and covert information tested. For general covert communication with a large size of encrypted data, the bit index of a given host utterance to modify may have to be determined experimentally. Additionally, host samples that are perceptually masked – in the temporal or spectral domain – can be used for modification of a lower significant bit. These topics are presently under investigation.

5. REFERENCES

- [1] M.D. Swanson, M. Kobayashi, and A.H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proc. IEEE*, Vol. 86, pp. 1064-1087, June 1998.
- [2] W. Bender, D. Gruhl, N. Morimoto and A.Lu, "Techniques for data hiding," *IBM Systems Journal*, Vol. 35, Nos. 3 & 4, pp. 313-336, 1996.
- [3] L.M. Marvel, C.G. Boncelet, Jr. and C.T. Retter, "Spread spectrum image steganography," *IEEE Trans. Image Proc.*, Vol. 8, No. 8, pp. 1075-1083, 1999.