

# Audio Steganography using Bit Modification – A Tradeoff on Perceptibility and Data Robustness for Large Payload Audio Embedding

Kaliappan Gopalan

Department of Electrical and Computer Engineering  
Purdue University Calumet  
Hammond, IN 46323, U.S.A.  
gopalan@calumet.purdue.edu

Qidong Shi

Department of Electrical and Computer Engineering  
Purdue University Calumet  
Hammond, IN 46323, U.S.A.  
qidong.shi@gmail.com

**Abstract**—Audio steganography using bit modification of time domain audio samples is a simple technique for multimedia data embedding with potential for large payload. Depending on the index of the bit used to modify the samples in accordance with the data to be hidden, the resulting stego audio signal may become perceptible and/or susceptible to incorrect retrieval of the hidden data. This paper presents some results of the tradeoff between the conflicting requirements of data robustness, payload and imperceptibility. Experimental results on both clean and noisy host audio signals indicate that while the payload can be as high as over 3000 bits/s – much higher rate than common audio data embedding techniques – noticeability of embedding is decreased and noise tolerance increased by using higher bit indices than the traditional least significant bit. Bit error rates of below one percent were observed for data retrieved from noise-added stego audio signals with 39 dB of SNR for an embedded payload of over 10 Kbits in a 3.3 s host audio.

**Keywords** - Audio steganography, data embedding, bit modification; perceptual quality; stego audio

## I. INTRODUCTION

Audio steganography is concerned with hiding information in a cover (host) audio signal in an imperceptible way. Hidden information from the stego, or data-embedded audio signal, is retrieved using a key similar to (or, in most cases, the same as) the one that was employed during the hiding phase. Audio and multimedia data embedding is a useful means for transmitting covert battlefield information via an innocuous cover audio signal. Other applications include transmitting confidential data such as banking transactions and biometric data with information safety and security. Watermarking, a subset of steganography, is yet another application in which a small amount of predefined information is hidden indiscernibly in a host audio signal for copyright protection or authentication of the host. While cryptography is widely used for secure data transmission by encrypting the data, stego transmission employs existing unclassified channels carrying innocuous audio signals with no additional bandwidth. Furthermore, the very use of encrypted messages arouses suspicion about secure communication via unsecure channels. To realize the advantages of an unsuspecting host carrying covert data via an

unclassified channel, a selected steganography technique must meet the primary criteria of imperceptibility of the stego, accurate recovery of data even under adverse conditions, and a strong key-based embedding and recovery of the data. Other requirements such as high payload and data recovery without the original cover signal (oblivious recovery) depend upon the type of applications. Generally, it is preferable for the stego to carry a large volume of data that can be extracted using the key but without employing the original host for comparison.

Since imperceptibility is of paramount importance in covert communication – to avoid discernibility of hidden information by data pirates – audio steganography commonly employs the psycho acoustic masking phenomenon of the human auditory system [1 – 3]. Modifying the amplitude or phase at the masked frequencies in accordance with the data to be hidden ensures imperceptible stego. The technique, however, is algorithmically complex and the payload is low; the payload is also dependent on the host audio.

Modifying a bit in each sample of a cover audio can increase the payload at the cost, depending on the index of the modified bit, of perceptual difference between the host and the stego. If the least significant bit (LSB) of each sample – a simple and earlier technique used for image embedding – for example, is altered in accordance with covert data, the modification may not be perceivable. At other indices, discernibility of the resulting stego depends on the index of the bit and the payload. The following sections describe some experimental observations of audio steganography by bit modification at selected bit indices for different payloads of embedded data and the resulting perceptual quality changes. Robustness of the hidden data with noise added to the stego audio at various SNR is presented for several cases.

## II. AUDIO 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 (or, selected pixels) in a 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 available 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 to the human 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. 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 frequency 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 were investigated and preliminary observations were presented in [4]. Present experimental results describe quantitatively the effect of modifying higher bit indices on the payload, perceptibility and data robustness.

### III. BIT MODIFICATION ALGORITHM

The process of embedding a bit of data on the  $k$ th bit of an  $N$ -bit audio sample ( $k = 1, 2, \dots, N$ , with  $k = 1$  as the LSB) proceeds with the selection of a  $K$ -bit key. Typically, a key of 256 bits – longer key for higher level of security – is selected at the transmitter and a copy of the key is made available at the receiver.

Let  $b_i(m)$ ,  $m = 1, 2, \dots, M$  denote the  $m$ th data bit of an  $M$ -bit data to be embedded, and let  $l$  be the index of audio samples where the data will be embedded. We note that

$$l = L, L+1, L+2, \dots, L+M-1$$

for the case where  $M$  successive audio samples starting at sample  $L$  will carry the hidden data.

Then, for each audio sample  $d_l$  for  $l = L, L+1, L+2, \dots, L+M-1$ , its  $k$ th bit  $b_l(k)$  is modified as

$$b_l^t(k) = b_0(m) \oplus b^k((l \bmod K)+1), \quad (1)$$

where

$b_0(m)$  is the  $m$ th data bit and  $b^k(i)$ , for  $i = 1, 2, \dots, K$ , is the  $i$ th key bit. ( $\oplus$  refers to the exclusive OR operation.) The modified bit  $b_l^t(k)$  replaces the  $k$ th bit  $b_l(k)$  of the  $l$ th audio sample.

To retrieve the data bit  $b_l^d(k)$  hidden in the  $k$ th bit of the  $l$ th stego audio sample, the process is reversed as

$$b_l^d(k) = b_l^r(m) \oplus b^k((l \bmod K)+1), \quad (2)$$

where  $b_l^r(k)$  is the bit received in the  $k$ th bit position of the stego sample. Clearly, if  $b_l^r(k) = b_l^t(k)$  and the  $l$ th stego sample received is the same as the  $l$ th stego sample transmitted, the exclusive OR operation in (2) correctly recovers the embedded bit. The next section describes the experimental results observed for perceptibility, payload and data resiliency at different values of the bit index  $k$ .

### IV. EXPERIMENTAL RESULTS

Initially, a noisy speech utterance from the Greenflag database that contains ground-to-aircraft communications was used for payload vs. perceptual quality testing. This utterance has 40338 samples with a sampling rate of 8000 samples/s for a length (duration) of 5.0423 s of speech by a male speaker. With change in perceptibility of bit modification as the criterion, extreme cases of embedding (a) all 1's, and (b) all 0's in selected speech samples were considered first. For small-size data (of up to 1000 bits or so), each bit was embedded once in every  $N$ th sample. For 1000 bits of data, for example, each bit may be embedded in samples at intervals of 40 samples starting from the first sample. This interleaving of data in the host audio can contribute to imperceptible embedding depending on the bit index used for modification. As the payload increases, number of samples left unmodified needs to be reduced to accommodate all the data bits.

#### A. Perceptibility and Payload

Perceptibility of embedding can be determined from the spectrograms of host and stego audio signals and by listening, which is a subjective measure. To objectively measure the distortion introduced by bit modification due to embedding, a perceptual quality measure based on Bark spectral distortion (BSD) was employed. The BSD measure is based on the assumption that speech quality is directly related to speech loudness, which quantifies auditory sensation in the psychoacoustical domain. The measure gives the average Euclidean distance between estimated loudness of a host (original) audio and that of a stego (or, modified) audio using Bark scale spectra in each critical band. A variation of this measure using 15 loudness components to calculate the loudness difference between two utterances has been shown to give improved results over basic BSD measure that correlates well with subjective listening tests and their mean-opinion scores [5].

Table I shows the results of embedding payload of 100 bits to as high as 6000 bits on the noisy host. Correct bit recovery was achieved in each case with no noise added to the stego audio. The third column on the table lists the number of samples skipped between embedding successive data bits, and the column on data type indicates if a random sequence of bits,

or all bits of 1's or 0's were used. Figs. 1 and 2 show the spectrograms of the original and stego audio for the case of modifying the 10<sup>th</sup> bit of the original audio samples in accordance with a dataset of 1000 bits of 1's with ex-OR operation using a 256-bit key. By modifying one in every 40 samples, the embedding is spread out. While this resulted in a slightly noticeable change in the spectrogram of the stego, the difference was not audible; nor was it detectable as a measureable value using BSD. It is, of course, possible that the result of the ex-OR operation of data and key bits (here, the complement of the key bits) may have been the same as the original sample bits at the 10<sup>th</sup> position so that not all 1000 samples were modified. Still, considering that all data bits are unlikely to be the same for a useful payload, the result of indiscernible embedding even at the 10<sup>th</sup> position is quite significant.

When the bit index for embedding was changed to 15, however, considerable distortion occurred, as expected, due to the large changes in sample values. Similar results of imperceptible embedding were observed for payloads as high as 10,000 bits, as indicated in Table I. It should be noted that even at the high payload of 10,000 bits, with only one sample interleaved, embedding did not show any significant difference in perceptual quality in subjective listening or in the objective test using the BSD criterion. It is particularly noteworthy that when embedded in bit 5, the stego was indiscernible. This may, quite possibly, be due to the choice of the host audio and its bit values at the embedded bit index, as stated earlier. Additionally, the high noise level present in the host renders modification of a part of the samples undetectable in listening and objective tests.

To further study the effect of modifying sample values on the discernibility of the audio, a noise-free utterance from the clean TIMIT database was next employed as the host audio. This utterance, spoken by a female speaker, was sampled at 16,000 Hz and has a length of 3.2826 s. Samples were modified using a randomly generated, fixed length key, as with the noisy host. At a low payload of 1000 bits, the stego showed little or no noticeable difference in spectrogram or speech quality when bit index 5 was used for modification with all 1's, 0's, or a random array of bits, as indicated in Table II. With bit index 10, however, the same payload of 1000 bits caused a slightly noticeable change, as expected, even with 50-sample interleaving. At the high payload of 10,000 bits, the change in perceptual quality or spectrogram was extremely small. This, again, is due to the specific values of the host audio as evidenced by the large changes when bit 15 was modified with 1000 bits of 1's with 50-sample interleaving.

The EMBSD measure is particularly significant in quantifying the perceptual difference between the noise-free host audio and the data-embedded stego audio. With the 1000-bits embedded stego, for example, two listeners could not notice any audible difference in the stego while the Bark spectral distortion in the BSD measure showed a value of 0.1. Although this is an insignificant value, it must be taken into consideration for strong data hiding as in covert communication. Comparison of perceptual quality, of course,

requires the original host audio, which may not be available readily for steganalysis. Similar quantitative measures were also observed for the difference between host and stego audio signals using mel frequency distributed cepstral coefficients.

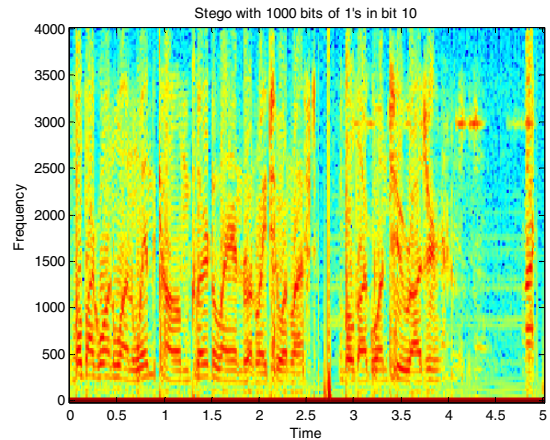


Figure 1. Spectrogram of a noisy host audio

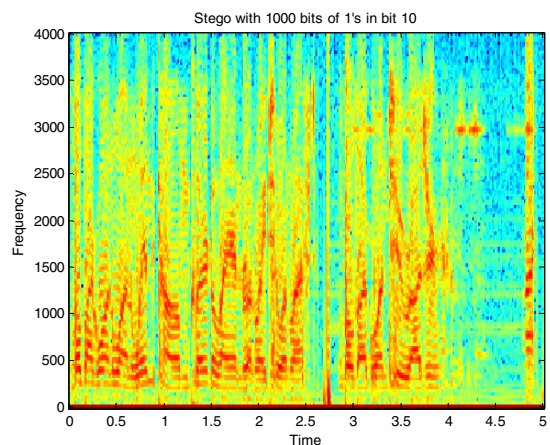


Figure 2. Spectrogram of stego using the noisy host (Fig. 1) carrying 1000 bits in the 10<sup>th</sup> bit position of 1000 samples with 40-sample interleaving

Fig. 3 shows the spectrogram of the original noise-free host and Figs. 4 – 6 show examples of spectrograms of bit modified stego for the cases of embedding (a) 1000 bits of 1's in sample bit index 5 with 50-sample interleaving (Fig. 4), (b) 10,000 bits of 1's in index 5 with 5-sample interleaving (Fig. 5), and (c) 10,000 bits of 0's in index 10 with every sample modified starting at the 100<sup>th</sup> sample in the host audio (Fig. 6). It is clear that visibility and audibility of embedding depend on the type of data and the number of samples skipped between embedded samples, besides the type of host audio.

### B. Data Robustness

Bit error rate (BER) due to noise added to the stego, clearly, depends on the modified bit index and the noise power level. To obtain a measure of robustness, white Gaussian noise was added with its power level monitored for getting a retrieved BER of less than one per cent for a payload of 10,704 bits (20

per cent of the host samples). Table III shows the BER as a function of SNR and the bit index for a few cases.

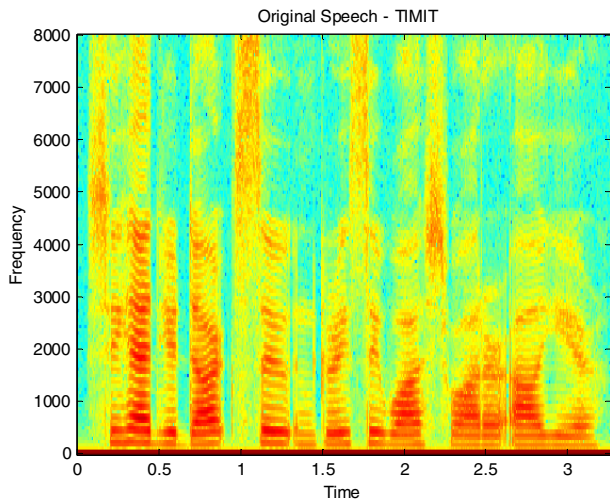


Figure 3. Spectrogram of a noise-free host audio

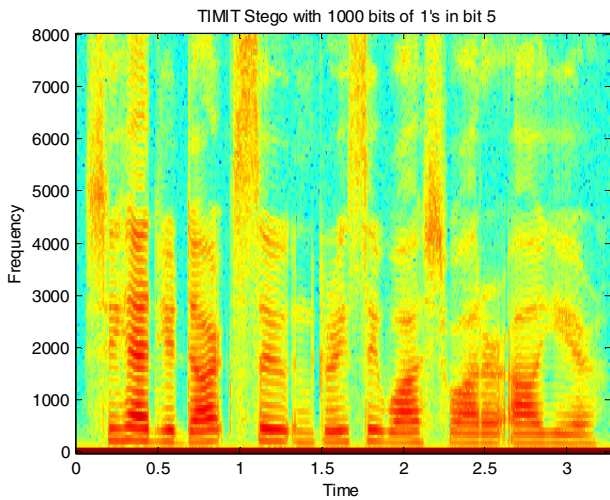


Figure 4. Spectrogram of stego carrying 1000 bits of 1's in bit index 5 in the clean host of Fig. 3

When lower bit indices are used for embedding, the resulting stego is susceptible to noise and hence, retrieved data may incur bit errors even at significantly high signal-to-noise ratios. Embedding at a high bit index can, clearly, tolerate high noise level without resulting in a large BER. This robustness – with 20 per cent of the samples carrying concealed data at the rate of 3260.8 bits/s in the result – shows that the bit modification is a viable technique even under adverse transmission conditions. We note that the robustness (or imperceptibility of stego) at higher bit indices does not require any subsequent bit modification, as is the case with previously reported work [6].

With capability for large payloads, BER can be reduced by repeating the data sequence during embedding and employing a majority voting method for retrieval. Although this process reduces payload, it can be an advantage for secure embedding

for covert communication. Additionally, the strength of key for embedding and retrieving can be increased by a larger size key and/or by using different keys for different sample ranges.

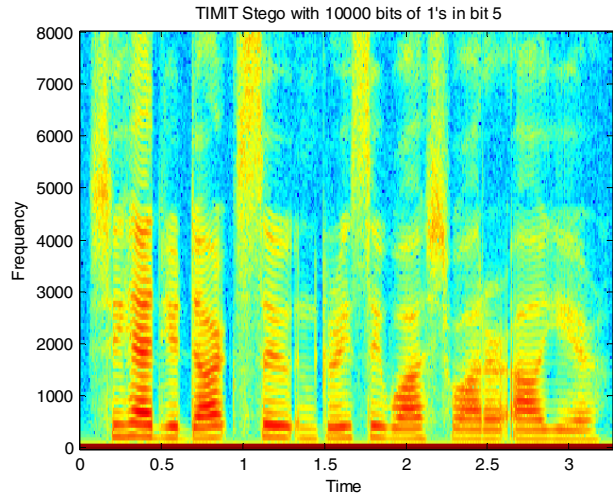


Figure 5. Spectrogram of stego carrying 10,000 bits of 1's in bit index 5 in the clean host of Fig. 3

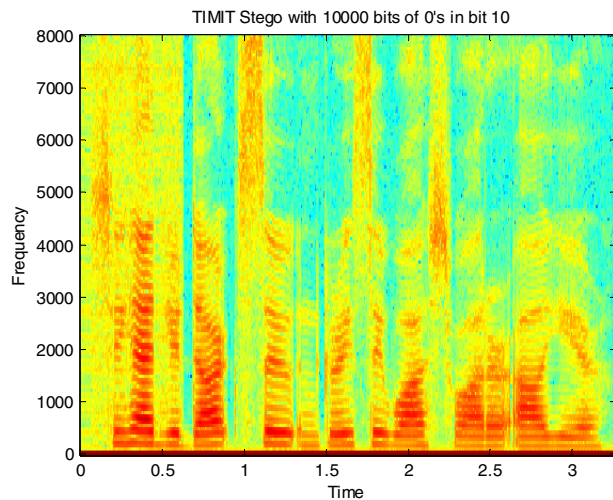


Figure 6. Spectrogram of stego carrying 10,000 bits of 0's in bit index 10 in the clean host of Fig. 3

## V. CONCLUSION

We presented some results of a study on the conflicting requirements of payload, perceptibility and data robustness in bit modification audio steganography. The study demonstrated the capability of the technique for hiding a potentially large payload of data with robustness using high bit indices for embedding. A tradeoff between noise tolerance and payload, both of which depend on higher bit indices, is needed for a reasonably imperceptible embedding.

ACKNOWLEDGMENT

The authors gratefully acknowledge the help and code provided by Professor Robert Yantorno of Temple University, Philadelphia, PA, for the Bark spectral distortion measure.

REFERENCES

- [1] 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.
- [2] R.J. Anderson and F.A.P. Petitcolas, "On the limits of steganography," *IEEE J. Selected Areas in Communications*, Vol. 16, No. 4, pp.474-481, May 1998.
- [3] M.D. Swanson, M. Kobayashi and A.H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proc. IEEE*, Vol. 86, pp. 1064-1087, June 1998.
- [4] K. Gopalan, "Audio Steganography Using Bit Modification," *Proc. of the IEEE 2003 International Conference on Multimedia and Exposition (ICME 2003)*, July 2003
- [5] W. Yang, M. Dixon and R. Yantorno, "A modified bark spectral distortion measure which uses noise masking threshold," *Proc. of the IEEE Speech Coding Workshop*, pp. 55-56, Pocono Manor, 1997.
- [6] N. Cvejic, and T. Seppanen, "Increasing robustness of LSB audio steganography using a novel embedding method," *Proc. of the International Conference on Information Technology: Coding and Computing (ITCC 2004)*, vol.2, issue 5-7, April 2004.

TABLE I. PAYLOAD VERSUS PERCEPTIBILITY AS A FUNCTION OF BIT INDEX FOR A NOISY HOST

Bit Index $k$	Payload $M$	Samp. Skip. *	Data Type	Spec. Visib.	EMBSD Meas.	Aud. Diff.
5	1000	40	1	No	0.0	No
5	1000	40	0	No	0.0	No
5	1000	40	R	No	0.0	No
10	1000	40	1	Slight	0.0	No
15	1000	40	1	High	2.9	Yes
15	1000	1	1	Mod.	1.5	Yes
5	10,000	1	0	No	0.0	No
10	10,000	1	0	Slight	0.1	No

R – Random array of bits; Mod. – Moderately noticeable change in spectrogram compared to host spectrogram.

\* Number of samples between data-embedded samples

TABLE II. PAYLOAD VERSUS PERCEPTIBILITY AS A FUNCTION OF BIT INDEX FOR A CLEAN HOST

Bit Index $k$	Payload	Samp. Skip.	Data Type	Spec. Visib.	EMBSD Meas.	Aud. Diff.
5	1000	50	0	Slight	0.0	No
5	1000	50	1	No	0.0	No
5	1000	50	R	No	0.0	No
10	1000	50	1	Slight	0.1	No
10	1000	50	0	Slight	0.1	No
15	1000	50	1	High	6	Yes
15	1000	1	1	Mod.*	1.3	Yes
5	10,000	1	0	No	0.0	No
5	10,000	1	1	No	0.0	No
10	10,000	5	0	Slight	0.1	Min.
10	10,000	5	1	Slight	0.3	Min.
10	10,000	5	R	Varies	0.2 – 0.3	Slight <sup>+</sup>

R – Random array of bits; \* Visible and audible only at the beginning; + An audible low level static

TABLE III. SNR FOR BER BELOW 1 PERCENT FOR DIFFERENT EMBEDDED BIT INDICES FOR A PAYLOAD OF 10,704 BITS

<b>Bit Index</b>	<b>SNR, dB</b>	<b>BER, in Percent</b>
1	106	0.1588
2	104	0.8595
5	99	0.7661
8	86	0.8688
10	75	0.9996
15	39	0.9342