A Robust Bit Modification Audio Steganography for Covert Communication

Kaliappan Gopalan
Department of Electrical and Computer Engineering
Purdue University Calumet
Hammond, U.S.A.
e-mail: gopalan@purduecal.edu

Jiajun Fu
Department of Electrical and Computer Engineering
Purdue University Calumet
Hammond, U.S.A.
e-mail: fu77@purduecal.edu

***Abstract*—Sample bit modification for data embedding on a cover audio signal has been shown as a viable technique for steganography and watermarking. Depending on the sample bit index chosen for carrying the embedded data, there is a tradeoff between viability of the data in the presence of noise, robustness and imperceptibility. Although a high threshold of audio samples can carry data at higher bit indices thereby raising robustness, it can be susceptible to noise even at low levels and, with sample amplitudes changed significantly, embedding becomes conspicuous, both detrimental for covert or secure communication. In this paper, modification of the high threshold sample embedding is shown to increase noise immunity with correct data retrieval at a lower payload, but without sacrificing indiscernibility. Experimental results using a noise-free utterance (from a corpus of read speech) and a noisy utterance (between air traffic controllers and pilots) show zero to low bit error rate of hidden data recovery at added noise levels of 50 decibels of signal-to-noise ratio.**

Keywords— Audio steganography; data embedding; bit modification; perceptual quality measure; noise robustness; stego audio.

I. INTRODUCTION

Secure and covert communication using unsecured network relies on steganographic techniques employing audio, image and video as host or cover carriers. Applications of such secure communication abound in battlefield data transmission and civilian transmission of banking, medical and employment data, to name a few. Steganography in general, and audio steganography more specifically, can supplement and enhance encrypted digital data for added security and privacy.

While the challenge of meeting all the key criteria of high payload, low or no perceptibility of embedding and high data integrity in the presence of noise is hard, applications with different requirements can readily be satisfied with tradeoff in one or more criteria. Watermarking of speech for copyright protection or authenticity verification, for example, may not need as much payload as for transmitting confidential medical data. Additionally, music copyright and/or transmission requires high level of indiscernibility. Covert communication may need to carry a reasonably high volume of information with little noticeability of the presence of embedding. Efficacy of techniques of data hiding can, therefore, be different with varying degrees of fulfilling the criteria. Additionally, use of the original host, or cover, audio signal for retrieving the embedded information may not be a limitation in watermarking applications; for covert communication, however, this type of escrow detection of hidden data may be an impediment requiring the use of the same host signal at the receiver and transmitter. It also may cause suspicion about the audio signal hiding information. Oblivious retrieval, on the other hand, needs some property of the host signal to remain the same in the stego, or data-embedded signal. A generally used invariant property is the psychoacoustic masking phenomenon of the human auditory system that renders spectral changes in an audio signal that are below its global masking threshold imperceptible. If the embedding procedure leaves the resulting spectral changes below the masking threshold, the stego becomes indiscernible from the host. In addition, the same masking threshold can be used at the receiver to retrieve the embedded data. Based on these key advantages, a number of techniques have been developed for audio steganography with oblivious detection [1]-[4].

The paper is organized as follows. Section II provides a brief review of audio sample bit modification for embedding. In Section III the proposed bit modification technique is described. Experimental results observed and a discussion of these results are given in Section IV. Conclusions drawn from the work form Section V.

II. AUDIO STEGANOGRAPHY EMPLOYING TIME DOMAIN SAMPLE BIT MODIFICATION

An alternative to hiding data in the spectral domain of host audio that exploits the auditory masking property of human perception is to alter time-domain samples in according with the data. Time-domain sample modification maintains imperceptibility if small changes are made to a few samples that are in the neighborhood of relatively large samples, for example. An early sample modification technique replaced the least significant bit (lsb) of each of a selected set of host audio samples with the data to be embedded. Such a simple technique, clearly, is susceptible to loss of data due to noise and also to illegal removal or replacement of the lsb. Several higher order bit modification techniques carry data on samples that are large enough but at bit indices that contribute to relatively small changes so that audibility of embedding is reduced. While lower bit indices generally cause less noticeability of embedding with higher

 ISBN: 978-1-61208-411-4

payload, it is also more susceptible to noise [5]-[7]. In this paper, we report an imperceptible bit modification steganography that can recover hidden data in the presence of noise on the stego.

III. SAMPLE BIT MODIFICATION AT SIGNIFICANT SAMPLES

Employing high bit indices for carrying hidden data can alleviate noticeability of modification if the samples are large in amplitude and the modified bit is relatively small. While this may reduce payload for a given host audio – due to non-availability of a large number of high amplitude samples – it can help mask auditory perception and contribute to higher noise robustness. With this premise, the following bit modification procedure was carried out on a noise-free and a noisy audio signal, as an extension to previously reported bit modification steganography [8].

Samples of a given host audio signal are selected for carrying hidden data based on a threshold M, where $M = 2^{l1} + 2^{l2} + 2^{l3}$, so that only amplitudes a that satisfy $|a| \geq M$ are used for modification. For a 16-bit audio with full dynamic range, a typical threshold can use l1 = 10, l2 = 11 and l3 = 12 (with LSB at index 1) so that $M = 3584$; hence, only samples with magnitudes of at least 3584 are considered potential samples available for bit modification.

To reduce the significance of change due to one of the 16 bits modified in the set {S}, sample bit $k < l1$, the smallest index used for the threshold, is used for modification in accordance with data to be hidden if $\frac{2^k}{M} \leq r$.

This criterion ensures that the modified sample is different from the original host audio by no more than $100r$ %. By a choice of r, this empirical rule can result in minimal changes in stego while affording different higher order bit indices for embedding in larger sample values. Although a large bit index k may raise data robustness to noise, it can also cause noticeable change in spectrogram and audibility, both resulting in conspicuousness of embedding. A reasonable choice for the index k is, therefore, below the lowest threshold bit index $l1$, in general. Test results are shown in the following section for different values of k.

IV. EXPERIMENTAL RESULTS

The first test used a noise-free utterance (from the corpus of phonemically and lexically transcribed speech of American English speakers) available at a sampling rate of 16000 Hz. Using a threshold of $M = 3584$, i.e., with l1 = 10, l2 = 11 and l3 = 12, effect of modifying different bit indices of samples satisfying the threshold was studied for different levels of noise added to the stego. The host audio was windowed into 320 samples (20 ms) of non-overlapping segments. Only those segments that had a significant number of potential amplitudes (at least 10) were considered for carrying hidden data. To increase data robustness in the presence of noise, if a frame had at least 10 potential samples, each of these samples was modified at its k^{th} bit with the same single bit of data (or, data bit exclusive-ORed with a key) to be embedded. By using a majority of the 10 (or more) recovered bits, probability of correct bit recovery was increased at the cost of reduced payload. As an example, Figure 1 shows the spectrograms of the original (host) audio and the stego carrying 71 bits of data in each of the 71 frames. The host with 51544 samples had 161 frames with 71 frames having 10 or more samples that were larger than the threshold of $M = 3584$. Each of the first 10 significant samples in a frame was modified at its 7th bit (lsb = 1) with the same data bit. The data bit index value corresponded to the embedded frame index – frame 16 that satisfied both the threshold and the number of samples, for example, carried data bit 16 in all of its 10 or more samples. Thus, with 71 frames of the host audio, the stego carried 710 bits with 71 bits of data.

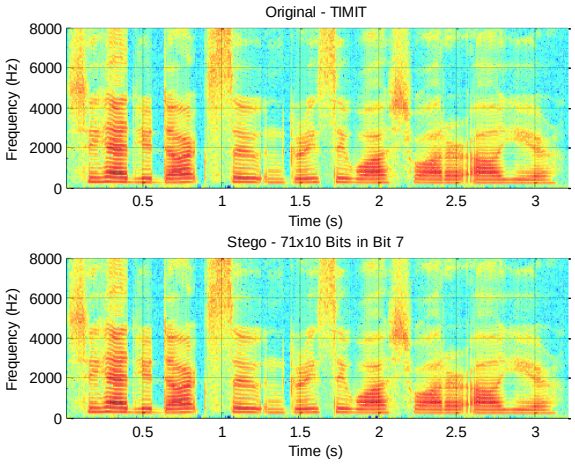

Figure 1. Spectrograms of original (noise-free) host (top) and stego carrying 71x10 bits at sample index 7 (lsb = 1)

Retrieval of the data bits was carried out by first segmenting the stego audio into samples of 320 and determining those samples that were above the threshold of 3584 in magnitude. If a segment had at least 10 such samples, then the k^{th} bit of each of the first 10 of these samples was obtained (with the same key as the one used for embedding). A majority of the 10 recovered bits was considered the correct embedded bit. With this procedure, all of the 71 embedded bits were correctly recovered.

To study the robustness of data with noise, zero-mean Gaussian noise at various levels of signal-to-noise ratio (SNR) was added to the stego. Data-retrieval from the noise-added stego was proceeded in the same manner as above. If the level of noise was such that the threshold was unaffected, the samples in which a bit in each was embedded remained the same; the noise, however, could have affected the k^{th} bit in some cases. By a majority voting of the recovered bits from the first 10 samples of each embedded frame, error due to noise was reduced. Figure 2 shows the original host audio and the noise-added stego at 50 dB of SNR. With majority voting, all 71 bits embedded in the stego were correctly recovered.

 ISBN: 978-1-61208-411-4

As the noise level was increased, either more samples were affected at the k^{th} bit of embedded samples, or worse, the noise altered the embedded samples so that threshold was not satisfied at the same frames as those used for embedding; both cases led to errors in data retrieval. Changing the bit index k for sample modification of the host, similarly, caused errors with lower levels of noise as k was decreased. Table I shows the data bit error rate (BER) as a function of modified bit index k and SNR. Each row corresponds to the maximum SNR for the k^{th} bit used for embedding. As the bit index k was reduced, noise tolerance became smaller and BER increased, although a BER of zero was achieved for the stego without any noise.

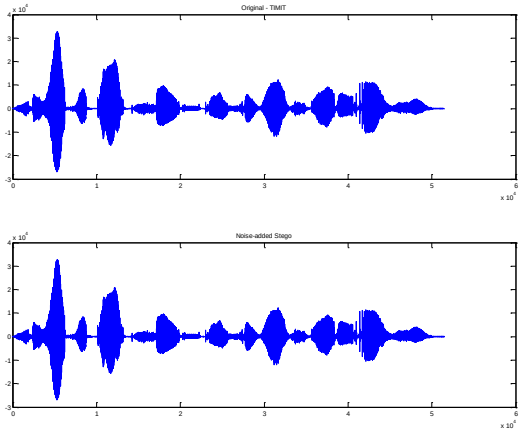

Figure 2. Waveforms of original (host) audio (top) and the noise-added stego carrying a total of 710 bits.

TABLE I. EMBEDDED BIT INDEX VS. NOISE VS. BER

k	SNR	BER, %
9	50	0
8	50	0
7	50	0
6	50	1.4085

All cases correspond to the same threshold of 3584 and 71 bits of data with each bit repeated 10 times in a frame with 10 or more samples above the threshold.

In the second test, a noisy audio was used as a practical example of host to carry hidden information. This audio from the Greenflag database of communication between fighter aircraft pilots and their air traffic controllers has 80150 samples obtained at the rate of 8000 per second. With 160 samples (20 ms) per frame, there were 500 frames and 302 of these frames satisfied the same threshold of 3584 with 20 or more samples. Choosing to repeat the same bit 20 times (the first 20 in each embeddable frame), all 302 bits were recovered from the total of 6040 bits. Correct data recovery was also achieved with noise at 50 dB SNR added to the stego. Figure 3 shows the spectrograms of the original host audio and the noise-added stego audio carrying data at sample bit index 7. At higher levels of noise, BER started to show up. Similar results were observed for embedding indices of 8 and 9, again with noise added at 50 dB or higher SNR.

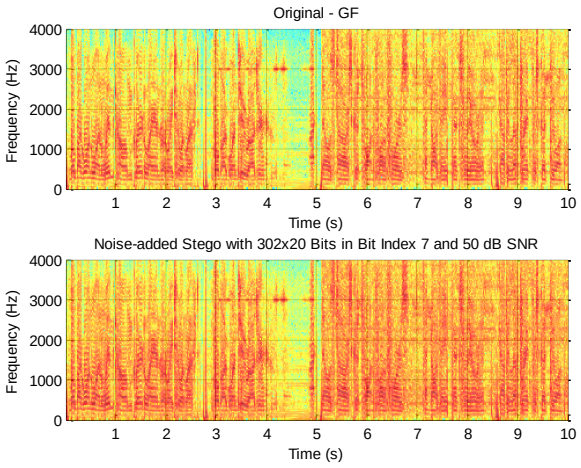

Figure 3. Spectrograms of host (noisy) audio (top) and noise-added stego carrying 302x20 bits at sample index 7

When the repetition rate of embedding the same bit in a frame was reduced to 10, there were 335 frames available for a data payload of 335 bits. At this slightly increased payload, one to three bits were incorrectly recovered at the noise level of 50 dB for a BER of 0.2985 to 0.8955, while no error resulted in the absence of added noise. This shows that the majority voting contributes to correct data recovery when noise is present in the stego. Similar results were observed at other indices for k, with BER increasing with noise level at lower indices.

From the two examples of audio considered, we may observe that a noise-free host audio is likely to have fewer samples satisfying a large threshold; payload, consequently, is reduced. A more realistic host audio with ambient and other type of noise, on the other hand, may have a high number of samples that can be modified with data without causing any perceptual or other difference.

V. CONCLUSION

An improved audio steganography employing time-domain sample bit modification at high bit indices has been proposed. Results observed on a clean and a noisy host audio signal show the viability of the technique in imperceptible embedding, and oblivious and error-free retrieval of the embedded data. By using a higher bit index of selected samples, and with a majority voting, hidden data bits can be correctly extracted even in the presence of added noise. The tradeoff for robust data recovery is low payload. The proposed method may be suitable for covert

 ISBN: 978-1-61208-411-4

communication of battlefield information or for secure transmission of medical and other data. Based on the imperceptibility of embedding, audio watermarking and authentication can also use this method.

REFERENCES

[1] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data embedding and watermarking technologies," Proc. IEEE, Vol. 86, June 1998, pp. 1064-1087.

[2] J. F. Tilki and A. A. Beex, "Encoding a Hidden Auxiliary Channel onto a Digital Audio Signal Using Psychoacoustic Masking , IEEE Southeastcon 97, April 1997, pp. 331-333.

[3] K. Gopalan, "Audio Steganography Using Bit Modification," Proc. of the IEEE 2003 International Conference on Multimedia and Exposition (ICME 2003), July 2003, pp. I-629-632.

[4] N. Cvejic and T. Seppanen, "Increasing robustness of LSB audio steganography by Reduced Distortion LSB Coding," Journal of Universal Computer Science, vol. 11, no. 1, pp 56-65, 2005.

[5] K. Gopalan and Qidong Shi, "Audio Steganography using Bit Modification – A Tradeoff on Perceptibility and Data Robustness for Large Payload Audio Embedding," Proc. of the 19th International Conference on Computer Communications and Networks (ICCCN 2010) -- Workshop on Multimedia Computing and Communications, Zurich, Switzerland, Aug. 2010, pp. 1-6.

[6] S. Rekik, D. Guerchi, S. A. Selouani, and H. Hamam, "Speech steganography using wavelet and Fourier transforms," EURASIP Journal on Audio, Speech, and Music Processing 2012, no. 1, Aug 2012, pp. 1-14.

[7] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative study of digital audio steganography techniques," EURASIP Journal on Audio, Speech, and Music Processing 2012, no. 1, Oct 2012, pp. 1-16.

[8] K. Gopalan and Jiajun Fu, "An Imperceptible and Robust Audio Steganography Employing Bit Modification," to be presented at the IEEE International Conference on Industrial Technology 2015, Seville, Spain, March 2015, pp. 1635-1638.

 ISBN: 978-1-61208-411-4