

# Audio Steganography Using Tone Insertion Technique

Suhaip A. Yousif  
Karrary University  
Khartoum, Sudan

Talaat M.wahbi  
Sudan University of  
Science and Technology  
Khartoum, Sudan

Mohamed H. Sayed  
The National Ribat  
University  
Khartoum, Sudan

---

**Abstract:** This paper presents a new technique of embedding text data into an audio file using tone insertion method. The new technique generates two frequency  $f_1$  and  $f_2$  and inserts them into audio file in a suitable power level according to specific table (stego-table). The proposed technique aimed to increase the payload capacity of the audio file using two bits in the frame without increases the number of the inserted frequencies to four frequencies, as well as using another convoy frequency (CF) for specific pattern. The proposed method conceals the English text into the .wave audio. The performance of the proposed method has been checked by spectrogram, MSE and PSNR.

**Keywords:** Audio steganography, Tone insertion, spectrogram, PSNR, MSE.

---

## 1. INTRODUCTION

Everyday internet access becomes available for a lot of people. The business company and banks and other organization which looks for more customers found the internet is the cheapest and the easiest way to increase customers, but they always fear of secret data. Cryptography and steganography appear to solve those fears.

Steganography is Greek word means secret writing. It is the part of information hiding science which focuses on concealing the data (secret data) into an object (covert object). The new object is the stego-object which sends through unsecure connection to the receiver. Any interception to the stego-object appears normal (without changing in the covert object) to the observer. The receiver can extract the secret data from the stego-object safely.

Nowadays the digitalized multimedia attracts the attention of information hiding scientist. Steganography can use text, image, audio, video and protocols as covert media. Steganography and cryptography are closely related. Cryptography scrambles messages so they cannot be understood. Steganography on the other hand, will hide the message in a way that hide the first place of the message. This property is so useful especially into battlefields and banks systems. Audio steganography takes advantage of the psycho acoustical masking phenomenon of the human auditory system [HAS]. Psycho acoustical, or auditory, masking is a perceptual property of the HAS in which the presence of a strong tone renders a weaker tone in its temporal or spectral neighborhood imperceptible. This property arises because of the low differential range of the HAS even though the dynamic range covers 80 dB below ambient level [1].

There are many techniques of Audio Steganography can be implemented into audio. They are temporal domain and transformed domain. Several methods of transformed domain contain the frequency domain, wavelet domain, Encoder domain. It is clearly that the tone insertion method comes under transformed domain into frequency domain [2]. It relay into frequency masking property. Any new steganography algorithm should get the three steganography triangle sides imperceptible (security), payload capacity (bit rate) and robustness. The weaknesses of this method are lack of transparency and capacity, but it has good embedded bits rate 250 bps and imperceptible and concealment of embedded data [2].

## 2. RELATED WORKS

K.Coplan et al published two papers in tone insertion technique [3][4]. They try to improve the capacity of tone insertion method. They use utterances from TIMIT (Texas Instruments Massachusetts Institute of Technology) database as host sample. TIMIT is noise-free database. Firstly they take the utterance "she had your dark suit in greasy wash water all year" which is available as 16 bit sample at the rate 16000 per second. With 208 frames, a random data of 208 bits were embedded. They generated two tones  $f_0$  (set at 1875Hz) and  $f_1$  (set at 2625Hz) to embedding bit 0 and bit 1 respectively. They divided host utterances to non overlapped segment. Every segment is 16ms in time. For every frame the compute  $P_e$  (frame power) and embedding only one bit into frame. If the embedding bit is 0 the power of the  $f_0$  is set to 25% of the  $P_e$  and the power of  $f_1$  is set to 0.001 of  $P_e$  and vice versa if the embedding bit is 1 the power of  $f_1$  set to 25% of the  $P_e$  and  $f_0$  is set to 0.001 of  $f_1$ . For recovery the every frame power is computed more over the power  $P_0$  and  $P_1$  for  $f_0$  and  $f_1$  respectively is computed. Then calculate the ratio of the power  $P_e/P_0$  and  $P_e/P_1$  if the first term greater than the second one then the embedded bit was 0 otherwise the embedded bit was 1.

Secondly they use noise host. It is applied in the Greenflag database consisting of noisy recordings of air traffic controllers, as host or cover audio samples. Successive frames for embedding were overlapped with 50 percent to further increase the payload capacity. After test the imperceptible of hidden data the technique had extended for use in covert battlefield communication in which the hidden information can be another utterance. They use speech utterance "seven one" said by male speaker as covert message and represented it in GSM half rate (GSM 06.20) coding schema resulting in compact form of 2800 bit. They concatenated two TIMIT utterances as cover audio to accommodate the large covert data. Each utterance with 16 bit samples and 16,000 samples/s. Tones for insertion were selected at frequencies of 687.5 Hz, 1187.5 Hz, 1812.5 Hz, and 2562.5 Hz. These frequencies were either absent or weak in the host frames. One of the frequencies is set to 25% of  $P_e$  the other frequencies are set to negligible power. To embedding set of (0, 0) set the power of  $f_0$  to 25% of the frame power and do the same for (0, 1), (1, 0) and (1, 1) for suitable frequency. For

recovery the receiver know the frequency order, so he can extract the data by computing the minimum power rate to the frame power. Another level of security may add by use frequency hopping in these four frequencies and use 4 bit key for every frame. The second experiment increase payload capacity.

### 3. PROPOSED TECHNIQUE

In this thesis the levels of the power of the inserted tones to three levels was increased and the carrier tone was decreased to two tones rather than four tones. By using this way two bits can be embedded in the frame. Moreover, we use a third inserted tone as controlled tone (convoy frequency CF) to increase the capacity more than two bits into the frame. Here an English text was used to conceal it into an audio file. So to find the best way to use convoy frequency CF we learnt carefully the ANSII code of the typed characters statistically. we decided to use CF for the 0110 pattern because this pattern appears 26 times in the typing character in ANSII at these letters (a, b, c, d, e, f (twice), g, h, I, j, k, l, m, n, o, v, F, V, X, Y, Z, [ , ' , &, 6). Moreover we will determine the benefits of CF for that random set in percentage. Note any ANSII character is coded into 8 binary bits, so we fetch the pattern only into the complication of two position (the ANSII position is 0,1,2,3,4,5,6,7) because the S.T table takes pairs of bit any time. Note that the 0110 pattern appears into the vowel letters except u in the lower case and clear that the vowel. letters repeats continually into words. This is raised the feasibility of using the pattern 0110. In the discussion part I study the percentage of using 0110 pattern in random set text written by different people. I use the stego-table (S.T table) that shows the frequencies and the power of the frequencies related to the embedded bits. Figure 1 shows the S.T table. Senders and receivers must know the S.T table in order to conceal and recovery the data safety. (Fe) is the fame energy.

**Table 1. The S.T table**

Tone frequency	Level of power	Embedded data
F1	1 (25% of fe)	0,1
F2	1 (25% of fe)	1,0
F1 and F2	2 (15% of fe)	1,1
F1 and F2	3 (0,01 of fe)	0,0

In case of (0,1) and (1,0) the power of F1,F2 is raised to 25% of fe respectively(level 1 of power). The F1,F2 together raised to 15% of fe(level 2 of power) when the embedded data is 1,1. The F1,F2 together raised to 0,01% of fe(level 3 of power) when the embedded data is 0,0 is set to level three power. Table 2 shows how to use the CF frequency in order to increase the embedded bits more than two into a frame. Note that any character is eight bits in ANSII and I embedded two bits in a frame so the bits to insert are complication of two.

**Table 2. The CF table**

Convoy frequency CF	Level of power	Embedded bits
F3	1 (25% of fe)	The following bit is 0110

In the embedded process after convert text data into binary checked, the 0110 pattern into the binary file if it is found raised the CF to 25% of fe. In the recovery process firstly check the CF if it is 25% of fe extract the 0110 pattern else check the f1 power and f2 power according to S.T table.

### 3.1 The Embedded Algorithm

1. Convert text file into binary file according to ANSII code.
2. Divide the audio file into frame.
3. Computer the power of the fame (fe).
4. Use the S.T table and CF table to raise the suitable frequency related with embedded bits.
5. Go to step3 until the end of the text file.

### 3.2 The Recovery Algorithm

1. Divide the stego-object into frame.
  2. Compute the power of the frame fe.
  3. Filter the cf frequency into frame.
  4. Check the power of cf  
If it is 25% fe output 0110.  
Go to step 1.
  5. Check the power of f1 (f1p)  
If it is 25% of fe  
Output 0,1  
If it is 15% of fe  
  
Output 1,1  
If it is 0,01% of fe  
Output 0,0  
Else if check f2p  
If it is 25% of fe  
Output 1,0.
  6. Go to step 1 for second frame until the end of the audio.
- Note: check f1p is enough to determine the hidden data but you can check f2p to be certain.

## 4. EXPERIMENTAL RESULTS and DISCUSSION

To determine the feasibility of using the pattern (0110) . We take random set from different persons( twelve text ) the median length of these text is about 840 words(into node bad). These text has been send to specific program in order to determine the size of the text after applied the s.t table and how many times the pattern 0110 appears on that text. Finally we compute the feasibility of using the pattern 0110. Table 3 shows these results.

**Table 3. The Feasibility of Using 0110 Pattern**

File name	Items of S.T	No-Of-Pattern	Benefit ratio
File 1	7432	502	15.6192%
File 2	16720	964	5.7656%
File 3	10679	1725	16.1532%
File 4	4312	568	13.1725%
File 5	7390	1118	15.1286%
File 6	7495	1101	14.6898
File 7	6772	988	14.5895
File 8	5321	783	14.7153
File 9	5492	720	13.1100
File 10	2115	353	16.6903
File 11	366	26	7.1038
File 12	87	17	19.5402

The best value of the benefit ratio is file10 and file 3. These files include just small letters without number. So to increase the benefit ratio of the pattern detection 0110 the senders must have two different audio. For every audio I change the text file(use two text file length) and the segment length (use three different length).

The first experiment I use 8 sec audio witch available into 16 b/sec sample rate and 16 sample as host audio file. Divide this audio into segments each one is 256 sample (16 m.sec). and use text data with 130 byte to hide it into the audio. Also I use the same file to hide 518 byte in it. (the audio was very small to hide this data). Moreover I do the same experiments but this time I change the segment length to 128 sample (8 m.sec) and 64 (4 m.sec).

The second experiment I use 14 sec audio witch available into 32000 b/sec sample rate and 8 sample as host audio file. Divide this audio into 256 sample length(8 m.sec). Also I change the segment length to 128 sample (4 m.sec) and 64 (4m.sec). Every time I use the same text files above to hide it into this audio. Table 4 shows all experiments.

**Table 4. The Feasibility of Using 0110 Pattern**

Audio	Text file In byte	Stego-name	Available capacity in byte	Note
Audio 1	130	Steg1_1_256	117	
Audio 1	130	Steg1_1_128	234	
Audio 1	130	Steg1_1_64	468	
Audio 1	518	Steg1_2_256	117	Big text
Audio 1	518	Steg1_2_128	234	Big text
Audio 1	518	Steg1_2_64	468	
Audio 2	130	Steg2_1_256	848	
Audio 2	130	Steg2_1_128	1696	
Audio 2	130	Steg2_1_64	3393	
Audio 2	518	Steg2_2_256	848	
Audio 2	518	Steg2_2_128	1696	
Audio 2	518	Steg2_2_64	3393	

The name of every stego in the stego\_name column has three value after the word stego. The first value indicate witch audio has been used. The second value means witch text file is hidden into this stego\_file. The last value indicate the segment length.

Note that stego1\_2\_256 and stego1\_2\_128 wasn't created because the audio file is too short to hide text file2 (518 byte) but when we use 64 segment length the audio can hide the text file. On the other hand note that stego1\_1\_256 and stego1\_2\_64 has been created although the available capacity was too short to hide the certain text file that because the use of the pattern detection.

## 5. SECURITY METRICS

We use two ways to ensure the security of this method these methods are spectrogram and PSNR,MSE. spectrogram of the audio and stego file it is clearly there are no visual differences between the two figure for all experiments.

Figure 1 to figure 5 shows the spectrograms of the first original audio and it is stego according to experiments

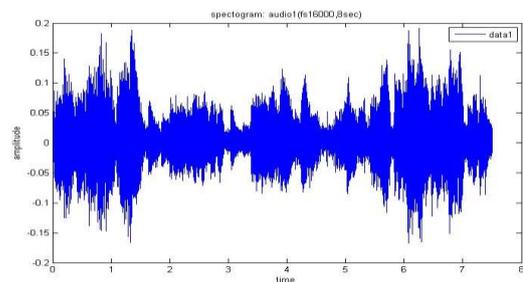


Figure 1 the original audio1

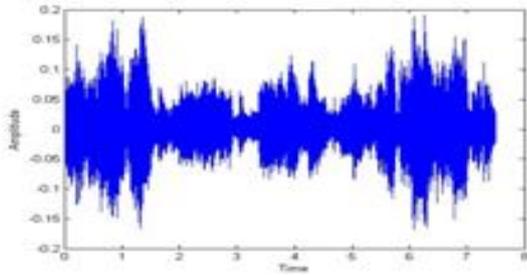


Figure 2 the stego1\_1\_256

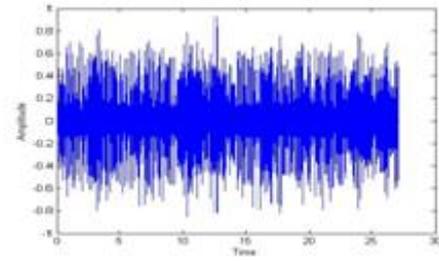


Figure 7 the steg2\_1\_256

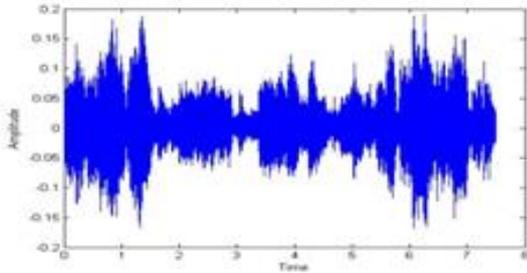


Figure 3 the stego1\_1\_128

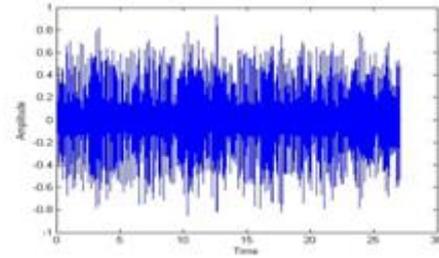


Figure 8 the stego2\_1\_128

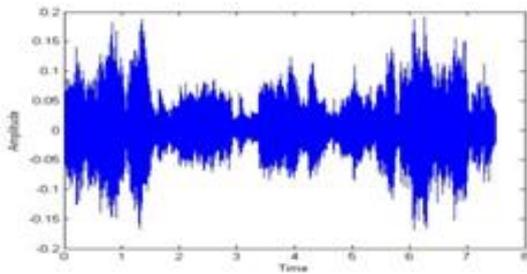


Figure 4 the stego1\_1\_64

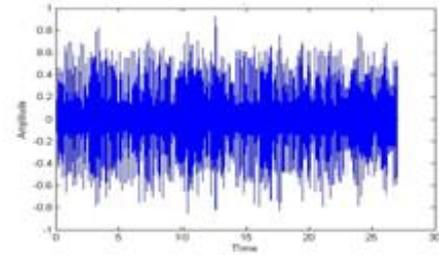


Figure 9 the stego2\_1\_64

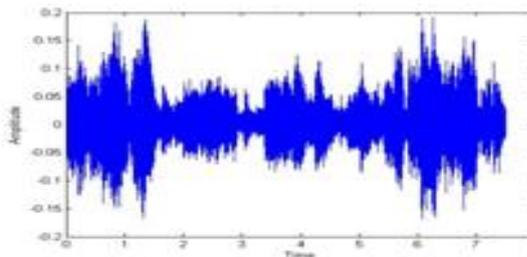


Figure 5 the stego1\_2\_64

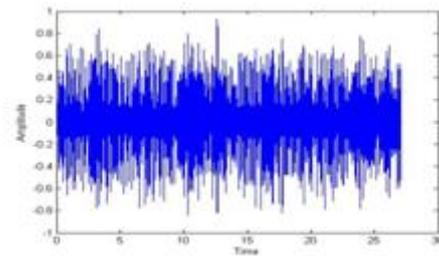


Figure 10 the stego2\_2\_256

Figure 6 to figure 12 shows the spectrograms of the second original audio and it is stego according to experiments.

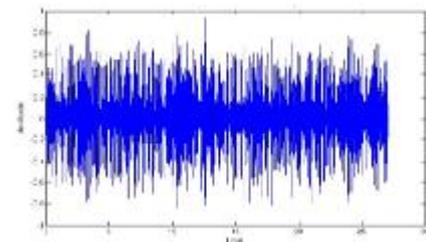


Figure 6 the original audio 2

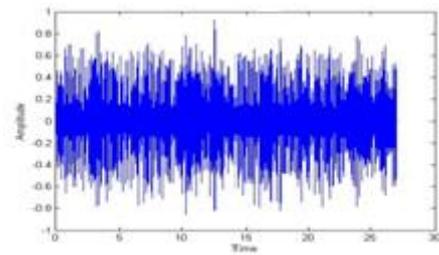


Figure 11 the stego2\_2\_128

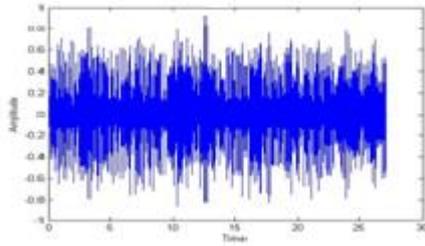


Figure 12 the stego2\_2\_64

The term peak signal-to-noise ratio (PSNR) is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. In statistics, the mean squared error (MSE) of an estimator measures the average of the squares of the "errors", that is, the difference between the estimator and what is estimated. Table 5 shows the PSNR and MSE values for experiments.

**Table.5 The PSNR and MSE Values**

Audio_name, stego_name	PSNR values	MSE values
Audio1,stego1_1_256	121.6298	0.000003
Audio1,stego1_1_128	125.7655	0.000003
Audio1,stego1_1_64	125.9546	0.000003
Audio1,stego1_2_64	122.0325	0.000003
Audio2,stego2_1_256	103.8790	0.000003
Audio2,stego2_1_128	111.1388	0.000001
Audio2,stego2_1_64	114.8495	0.000001
Audio2,stego2_2_256	98.2985	0.000010
Audio2,stego2_2_128	102.4410	0.000004
Audio2,stego2_2_64	104.9162	0.000002

From table 5 the PSNR increase when the segment length decrease that because of the signal stationary. When the segment length decrease the audio host is being more stable. So it is clear from table 4 and table 5 that small segment means increasing in security and in capacity. This depend somewhere into the host audio signal .

## 6. CONCLUSION AND FUTURE WORK

Audio Steganography is very interesting field. The tone insertion technique is imperceptible but the main limitation is the low capacity. Using the proposed method we can increase the capacity to 70% compared with the latest related work.(note that the related work use 2 frequency to insert one bit into frame we use 2 frequency to insert two bit into frame that is 50%. Moreover we use the pattern detection. The feasibility of the patter may arrived to 20% if we applied the mentioned recommendation). The host audio must be carefully choosed. If the host audio has empty sample at the beginning or at the end of it. We have to process it before using it. To guarantee the imperceptible tone and write extraction. We recommended also the host audio must analysis to determine the best segment length. Also use the mid band tone frequency is better. In the future we will concatenate the pattern detection with space character (note that the space character is frequently used). Also we use the human voice as a host audio. Use frequency hopping with known key to deny the frequency tracking and to achieve more security. Increase security by using suitable encryption algorithm for the plain text file before steganography.

## 7. REFERENCES

- [1] E. Zwicker and H. Fastl, 1990 "*Psychoacoustics*", Spriger-Verlag, Berlin.
- [2] Fatiha Djebbar, Beghdad Ayad, Karim Abed Meraim and Habib Hamam, 2012 "Comparative study of digital audio steganography techniques" ,EURASIP Journal on Audio, Speech, and Music Processing.
- [3] K Gopalan, S Wenndt, July 8–10, 2004 "Audio Steganography for Covert Data Transmission by Imperceptible Tone Insertion", WOC 2004, Banff, Canada.
- [4] K. Gopalan, S. Wenndt, A. Noga, D. Haddad, and S.Adams, , March 2003" Covert Speech Communication Via Cover Speech By Tone Insertion", Proceeding of IEEE Aerospace Conference, Big Sky, Montana.