SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF GRADUATE STUDIES

# Audio steganography using tone insertion technique

A Thesis Submitted in Partial Fulfillment of the Requirements of Master Degree in Computer Science

BY:

SUHAIB ALI ABD ELGAGER YOUSIF

SUPERVISOR

Dr. TALAAT MAHI ELDEEN WAHBI

**September 2014**

# الآيـــــــــــــــة

قال تعالي :( مُّحَمَّدٞ رَّسُولُ ٱللَّهِۚ وَٱلَّذِينَ مَعَهُۥٓ أَشِدَّآءُ عَلَى ٱلۡكُفَّار
رُحَمَآءُبَيۡنَهُمۡۖ تَرَىٰهُمۡ رُكَّعٗا سُجَّدٗا يَبۡتَغُونَ فَضۡلٗا مِّنَ ٱللَّهِ وَرِضۡوَٰنٗاۖ
سِيمَاهُمۡ فِي وُجُوهِهِم مِّنۡ أَثَرِ ٱلسُّجُودِۚ ذَٰلِكَ مَثَلُهُمۡ فِي ٱلتَّوۡرَىٰةِۚ
وَمَثَلُهُمۡ فِي ٱلۡإِنجِيلِ كَزَرۡعٍ أَخۡرَجَ شَطۡـَٔهُۥ فَـَٔازَرَهُۥ فَٱسۡتَغۡلَظَ فَٱسۡتَوَىٰ
عَلَىٰ سُوقِهِۦ يُعۡجِبُ ٱلزُّرَّاعَ لِيَغِيظَ بِهِمُ ٱلۡكُفَّارَۗ وَعَدَ ٱللَّهُ ٱلَّذِينَ
ءَامَنُواْ وَعَمِلُواْ ٱلصَّٰلِحَٰتِ مِنۡهُم مَّغۡفِرَةٗ وَأَجۡرًا عَظِيمَۢا ) الفتح 29

# الحمـــــــــــــــــــد

الحمد لله حمداً طيباً مباركاً فيه يليق بجلال وجهه وعظيم سلطانه. الحمد لله الذي بنعمته تتم الصالحات. احمد الله عز وجل أن وفقني إلي إتمام هذا البحث راجياً من الله أن يجعله في ميزان حسناتي وان ينتفع به غيري وان يزدني علماً ...

# DETICATION

To the soul of my mother, the first person who care and teach me.

To my wise father, for his advices and prayers for me.

To my wife, for care and support all the time.

And to my children's, Nabeil and Mohamed with hope for bright future.

# ACKNOWLEDGEMENT

There are a list of people I would like to thank them. First of all I would like to express enough thanks to my supervisor Dr. TALAT WAHBI for continued support. Besides my advisor; I would like to thank Dr. Ali Alfaki. Also I thank my fellows specially Mohamed Yousif and Mutwakel Faisal.

My completion of this project could not have been accomplished without the support of my family. When I have special problem with the thesis my father is always ask me very important questions that inspire me to solve the problem. So I would like to thank him for that and for supporting me spiritually throughout my life.

Last but not the least to my caring, loving, supportive wife. I would like to thanks her for continued encouragement and for care.

# Abstract

There are many techniques of Audio Steganography can be implemented into audio. This thesis focuses on the frequency masking in audio by using tone insertion method. Inserting tones at known frequencies and at low power level (depends on the original audio power) then modulate the secret massage into this inserted tone. The hidden information is imperceptible so a listener is unable to distinguish between the cover- and the stego-audio signal. So the cover is audio and the hidden data (secret massage) is the English language text document.

The proposed method focuses on the payload of the host audio with no disruption of robustness and imperceptible. The payload has been increased using new algorithm relay on new stego-table and using frequent pattern detection into ANSI code.

There are twelve experiments in this thesis. Two cover audio has been used to conceal different text size (130 and 518 byte). Every time the host audio has been changed using different sample rate, different recording environment (clean or noise) and different audio segmentation length. The total numbers of experiments are twelve. The proposed method has been tested using security metrics peak signal-to-noise ratio (PSNR) and mean squared error (MSE). The experiments showed that the proposed method increased the payload by 165% compared with the latest related works.

# المستخـــلص

يمكن تطبيق عدة أنواع من إخفاء البيانات(steganography)علي الصوت. يركز هذا البحث علي إخفاء البيانات داخل الأصوات مستخدماً طريقة إدراج نغمة في الصوت(tone insertion method). ويتم ذلك بإدراج نغمة معلومة التردد بمستوي قوة صوت منخفضة اعتماداً علي قوة الصوت الرئيسي. بهذه الطريقة نضمن أن النغمة المدرجة غير مسموعة وذلك لتغطيتها بصوت اعلي منها في القوة. وبذلك يمكن تحميل البيانات السرية المراد إخفاءها في هذه النغمة المدرجة. هذا البحث اعتمد علي أن الأصوات هي الغطاء للبيانات السرية المتمثلة في نصوص انجليزية مكتوبة علي محرر الويندوز.

هنالك ثلاث معايير رئيسية لعملية إخفاء البيانات هي السرية, سعة البيانات التي يمكن تحملها وقوة تحمل الصوت الناتج ضد الهجمات لتدمير البيانات المخفية داخله. هذا البحث يركز أساساً علي زيادة سعة البيانات مع عدم الإضرار ببقية المعايير. تم زيادة سعة البيانات بطريقتين الأولي باستخدام خوارزمية جديدة تعتمد علي جدول اخفاء جديد(stego-table) والثانية باستخدام اكتشاف الانماط المتكررة ( frequent Pattern detection) في البيانات الثنائية للحروف الانجليزية.

هنالك 12 تجربه  في هذا البحث. تم استخدام صوتين مختلفين كغطاء لنصوص  بأحجام مختلفة (130 أو 518 بايت). في كل تجربة يتم تغيير حجم مقاطع الصوت(segment), ومعدل ترميز الصوت( sample rate). وقد تم اختبار هذه الخوارزمية ببعض معايير السرية لإخفاء البيانات(PSNR,MSE). التجارب اظهرت ان السعه تزيد ب  165%  مقارنة بالبحوث ذات الصلة.

# TABLE OF CONTENTS

الآيـــــــــــــــــــــــة ........ II

الحمـــــــــــــــــــــــد ........ III

DETICATION ........ IV

ACKNOWLEDGEMENT ........ V

Abstract ........ VI

المستخـــــلص ........ VII

TABLE OF CONTENTS ........ VIII

LIST OF TABLES ........ X

LIST OF FIGURES ........ XI

Chapter 1

introduction ........ 1

1.2 Problem statement ........ 3

1.3 Research scope ........ 4

1.4 Research methodology and tools ........ 5

1.5 Objectives ........ 6

1.6 Research questions ........ 6

1.7 Research organization ........ 6

Chapter 2

Literature review andrelated works ........ 7

2.1 Introduction ........ 8

2.1.1Types of covert media ........ 9

2.1.2 Audio steganography ........ 9

2.1.3 Tone insertion technique ........ 10

2.1.4 Key steganography criteria ........ 11

2.2 Related works ........ 12

Chapter 3

Proposed method and Tools ........ 17

3.1 Overview ........ 18

3.2 Proposed method ........ 18

3.3 Embedding algorithm ........ 19

3.4 Extraction algorithm ........ 20

3.5 Application and tools ........ 22

Chapter 4

RESULT AND DESCUSSION ........................................ 24

4.1 Results ........................................ 25

4.2 Discussion ........................................ 27

Chapter 5

Conclusion and Recommendation ........................................ 31

5.1 Recommendation ........................................ 32

5.2 Future works........................................ 32

5.3 Conclusion ........................................ 33

REFERENCES........................................ 34

# LIST OF TABLES

Table 2.1:Related works comparison ........................................................................................ 16
Table 3.1 the S.T table ............................................................................................................... 19
Table 3.2 the CF table ................................................................................................................ 19
Table 4.1: The feasibility of using the pattern(0110) ....................................................................... 26
Table 4.2 PSNR and MSE value ...................................................................................................... 29
Table 4.3 The experiments detail ................................................................................................ 30

# LIST OF FIGURES

Figure 1.1 : fundamental schema of steganography process............................................................ 3
Figure 1.2:  Key Steganography Criteria ............................................................................................. 4
Figure 2.1: Types of steganography ..................................................................................................... 8
Figure 2.2:  Tone insertion method using two frequencies ............................................................ 13
Figure 3.1: Diagram of embedding algorithm .................................................................................... 20
Figure 3.2: Diagram of extraction algorithm ...................................................................................... 21

# Chapter 1

## INTRODUCTION

# 1.1 Introduction

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 throw 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 stgo-object.

Steganography and cryptography are closely related. Both make the data confidentiality but there is difference between them. Cryptography scrambles messages so they cannot be understood. Steganography on the other hand, will hide the message so there is no knowledge of the existence of the message in the first place. In some situations, sending an encrypted message will arouse suspicion while an "invisible” message will not do so. For battlefield communication, in particular, hiding the information using steganography is crucial. Both sciences can be combined to produce better protection of the message. In this case, when the Steganography fails and the message can be detected, it is still of no use as it is encrypted using cryptography techniques.

Now at day multimedia is becomes in digital form. Information and communication revolution helps multimedia spreading. Watermarking solves the problem of copy rights for Media Company. Watermarking always hide a pattern in target. The main difference between watermarking and steganography is that steganographic information must never be apparent to a viewer unaware of its presence. This feature is optional when it comes to watermarking.

Secrets can be hidden inside all sorts of cover information: text, images, audio, video and more. Any media have threshold capacity for steganographic information without changing the noticeable properties of the cover. The second property in any stego-object is the robustness against passive and active attacks. Figure 1 illustrate the general idea of steganography techniques

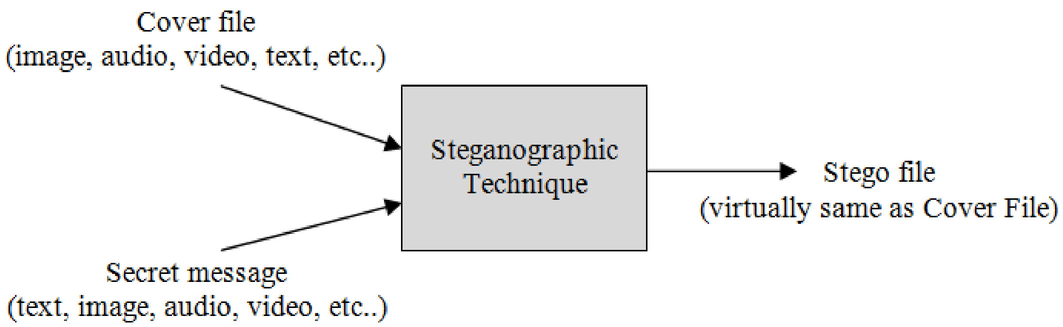

Cover file
(image, audio, video, text, etc..)
Steganographic Technique
Stego file
(virtually same as Cover File)
Secret message
(text, image, audio, video, etc..)

**Figure 1.1 :fundamental schema of steganography process**

Steganographic techniques have been used for centuries. The first known application dates back to the ancient Greek times, when messengers tatted messages on their shaved heads and then let their hair grow so the message remained unseen. In the 20th century, invisible inks where a widely used technique. In the Second World War, people used milk, vinegar, fruit juices and urine to write secret messages. Germans developed a technique called the microdot. Microdots are photographs with the size of a printed period but have the clarity of a standard type written page. The microdots where then printed in a letter or on an envelope and being so small, they could be sent unnoticed.

Steganographic techniques have been used with success for centuries already. However, since secret information usually has a value to the ones who are not allowed to know it, there will be people or organizations who will try to decode encrypted information or find information that is hidden from them. Governments want to know what civilians or other governments are doing, companies want to be sure that trade secrets will not be sold to competitors and most persons are naturally curious. Many different motives exist to detect the use of steganography, so techniques to do so continue to be developed while the hiding algorithms become more advanced.

# 1.2 Problem statement

Observers always suspect on encrypted messages, but they don't suspect on Steganography because the cover media appears normal to them. The main problems in Steganography are robustness of the stego-object against passive and active attacks, the available capacity on the covert media always small. And the transparency of the stego-object to the interaction entity(machine or human). In this thesis I will try to increase the available capacity on .wave audio using tone insertion method. Moreover, I will design new algorithm to address these above problems beside strength the algorithm against passive and active attacks like noise and filtering.

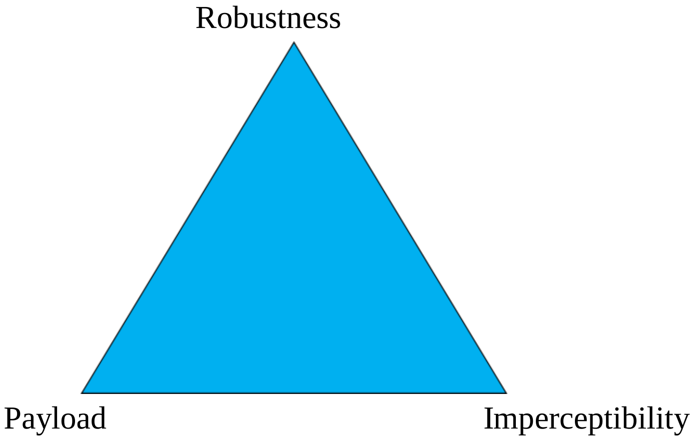


**Figure 1.2: Key Steganography Criteria**

# 1.3 Research scope

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 [1]. This property arises because of the low differential range of the HAS even though the dynamic range covers 80 dB below ambient level [1]. put faint tone immediately before or immediately after strong tone make the first tone undetectable by the HAS. Additionally, a weak pure tone is masked by wide-band noise if the tone occurs within a critical band. We must note that the masked sound becomes inaudible in the presence of another louder sound; the masked sound, faint it may be, is still present, however. This property of inaudibility of weaker sounds is used in different ways for embedding information [2].

The masking effect can be understood like a kind of interference with the audibility of a sound (called probe or “maskee”) caused by the presence of another sound (called masker), when both these sounds are close enough to each other in frequency and occur simultaneously or closely to each other in time. If a lower level probe sound is inaudible because of a simultaneous higher level masker, the effect is referred to as the simultaneous masking or frequency masking. Masking is typically described by the minimum shift if the probe intensity level above its threshold of audibility in quiet, necessary for the probe to be heard in the presence of masker. The Threshold of Masking corresponds to a limit case when the masking conditions are such that the “maskee” is in the threshold between audibility and inaudibility [3].

This thesis focuses on the frequency domain in audio (covert media) by making small changing in the frequency of the audio signal. I will use the tone insertion method in order to conceal secret text massage in this tone and cover it by modulate it on wave audio signal. The secret text massage is the English language in the windows text document under windows.

## 1.4 Research methodology and tools

First of all I will read the literature review and related works in this field of tone insertion method in audio in order to determine the strength and weakness of each algorithm. Then I will design a new algorithm to cover the common weakness and code it (I focus on the capacity of the stego-object and robustness of it). Convert an audio file to digital mode (AtoD) then insert the secret data in it and convert it analog mode(DtoA). Then   Finally I will test the algorithm efficiency using spectrogram, PSNR metrics and if possible the professional phonetics and sounds people. And compare it with other related works.

## 1.5 Objectives

**The main objectives of this thesis are:**

- Concealing massage into .wave audio imperceptible using tone insertion method.
- Try to increase payload of the audio using tone insertion technique.
- Extract the message safely from the stego-object.

**The sub objectives of this thesis are:**

- Compare the proposed  algorithm with other related algorithms.
- Frequent pattern detection into ANSI code in order to find much 4 bit repeated.
- Finding much expected text formula in order to focus on it.

## 1.6 Research questions

- Is there a relation between covert audio frequency and tone frequency?
- Can I use the carrier tone as insertion tone which take the modulate text?
- What is the maximum capacity can be taken into one minute audio without notice from listener?
- How can securely extract the hidden text from .wave audio?
- Does the sampling method effect in the audio steganography?
- What is the suitable audio segment to process it.
- How can I test my new algorithm.

## 1.7 Research organization

Chapter one is introduction which highlight a brief history of tone insertion Steganography and the basics of modulation and demodulation of tone in covert media. Recently literatures review and related works appears in chapter two. Chapter three contains the new algorithm and comparison with other algorithms. Chapter four is result and discussion. Conclusion and recommendation will write in chapter five.

# Chapter 2

## LITERATURE REVIEW ANDRELATED WORKS

## 2.1 Introduction

With steganography you can send messages without anyone having knowledge of the existence of the communication. There are many countries where it is not possible to speak as freely as it is in some more democratic countries. Steganography can be a solution which makes it possible to send news and information without being censored. Secrets can be hidden inside all sorts of cover information: text, images, audio, video and more and without the fear of the messages being intercepted and traced back to you.

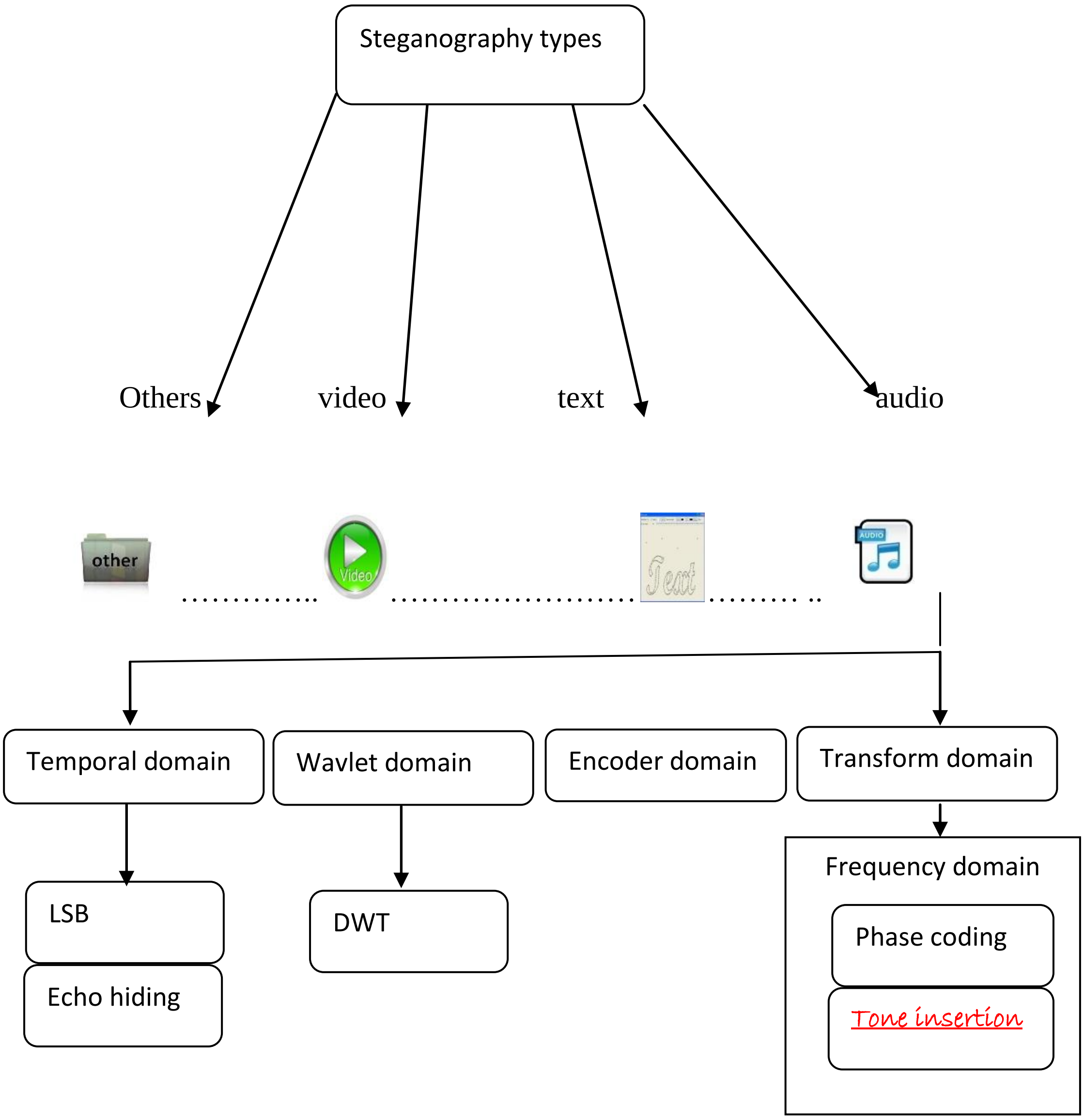


**Figure 2.1: Types of steganography**

### 2.1.1Types of covert media

Hiding information in plain text can be done in many different ways. Example the first-letter algorithm (take the first letter from each ward) but it isn't secure. Hiding information inside images is a popular technique nowadays. An image with a secret message inside can easily be spread over the World Wide Web or in newsgroups. The earliest and easiest way into image steganography is LSB method [4]. Audio stenography is masking, which exploits the properties of the human ear to hide information unnoticeably. An audible, sound can be inaudible in the presence of another louder audible sound .This property allows to select the channel in which to hide information. Video files are generally a collection of images and sounds. So most of the techniques used in images files and audio files can be applied to video files also. The great advantages of video is that large amount of data that can be hidden inside and since it is a moving stream of images and sounds, the continuous flow of information otherwise noticeable distortions might go unobserved by humans beings. The protocol steganography is used to embed information within the network protocols, such as TCP/IP. Information can be hidden in the header part of a TCP/IP packet and in some fields which are either optional or are never used.

### 2.1.2 Audio steganography

Audio files and signals make appropriate mediums for steganography due to the high data transmission rate and the high level of redundancy.  Text steganography is believed to be the hardest type of steganography because of the low degree of redundancy in text as compared to image, audio or video. Redundancy can be described as the bits of a media, signal or file that offer accuracy more than needed for the object’s use [5]. The redundant bits of an object may also be defined as those bits that can be easily altered without this change being; but hiding data in audio signals is not a simple mission**.**

There are many techniques of Audio Steganography can be implemented into audio. They are temporal domain and transform domain. The concept of hiding data into transformation domain comes from the human auditory system. It has certain peculiarities that must be exploited for hiding data effectively. The”masking effect” phenomenon masks weaker frequencies near stronger resonant ones. Several methods in the transform domain contain the frequency domain, wavelet domain, Encoder domain. The main techniques under temporal domain are Least Significant Bit (LSB) Coding, Parity Coding, and Echo data hiding. The wavelet domain based on Discrete Wavelet Transform (DWT). The encoder domain when considering data hiding for real time communications, speech codec such as: AMR, ACELP, SILK at their respective coding rate are employed. Passing through one of the codec, the transmitted signal is coded and compressed according to the codec rate then decompressed at the decoder end. The Frequency domain contain the following method tone insertion, Phase encoding, Spread spectrum, Amplitude modification, Cepstral domain [6].

### 2.1.3 Tone insertion technique

From figure 2.1 it is clearly that the tone insertion method comes under transform domain into frequency domain Frequency. It relay into frequency masking property. Masking property is exploited in tone insertion method. A weak pure tone is masked in the presence of a stronger tone. This property of inaudibility is used in different ways to embed information. By inserting tones at known frequencies and at low power level, concealed embedding and correct data extraction are achieved. The hidden information is imperceptible [7] if a listener is unable to distinguish between the cover- and the stego-audio signal.

The first problem that all data-embedding using tone insertion method needs to address is that of inserting data in the digital signal without deteriorating its perceptual quality. Of course, we must be able to retrieve the data from the edited host signal, i.e., the insertion method must also be invertible. Since the data-insertion and data-recovery procedures are intimately related, the insertion scheme must take into account the requirement of the data-embedding application. In many applications, we will need to be able to retrieve the data even when the host signal has undergone modifications, such as compression, editing, or translation between formats, including A/D and D/A conversions.

Audio masking is the effect by which a faint but audible sound becomes inaudible in the presence of another louder audible sound. The masking effect depends on the spectral and temporal characteristics of both the masked signal and the masker. Frequency masking refers to masking between frequency components in the audio signal. If two signals that occur simultaneously are close together in frequency, the stronger masking signal may make the weaker signal inaudible. The masking threshold of a masker depends on the frequency, sound pressure level, and tone-like or noise-like characteristics of both the masker and the masked signal. It is easier for a broad-band noise to mask a tonal signal than for a tonal signal to mask out a broad-band noise. Moreover, higher frequency signals are more easily masked. The human ear acts as a frequency analyzer and can detect sounds with frequencies that vary from 10 to 20 000 Hz. The HAS can be modeled by a set of band pass filters with bandwidths that increase with increasing frequency. The bands are known as the critical bands. The critical bands are defined around a center frequency in which the noise bandwidth is increased until there is a just noticeable difference in the tone at the center frequency. Thus, if a faint tone lies in the critical band of a louder tone, the faint tone will not be perceptible. Frequency-masking models are readily obtained from the current generation of high-quality audio codec's, e.g., the masking model defined in the International Standards Organization (ISO)-MPEG Audio Psychoacoustic Model 1 for Layer I. The Layer I masking method is summarized as follows for a 32-kHz sampling rate. The MPEG model also supports sampling rates of 44.1 and 48 kHz. The frequency mask is computed on localized segments (or windows) of the audio signal. The final step into this model consists of computing individual and global masking thresholds [8].

### 2.1.4 Key steganography criteria

The key steganography are imperceptible (security), payload capacity (hiding rate), and robustness[9]. The imperceptibility is the most important requirement of a steganography system, as the strength of steganography system depends on its ability to be unnoticed by the human senses (visually or acoustically. Robustness defines how strong the used steganographic technique against changes. It measures the capability of the embedded secret data to endure different types of intentional and unintentional modifications such as signature manipulation, compression. The payload capacity is the size of embedded data that can be hidden into a particular innocent cover medium relative to the size of this medium. The real challenge is how to hide as much secret data as possible while keeping the quality of the medium untouched and without infringe the imperceptibility requirement. Generally, increasing the embedding capacity makes the secret hidden information more conspicuous in viewing. To calculate the embedding capacity of a particular steganography system, the size of the embedded secret message is divided by the total size of the cover medium.

# 2.2 Related works

There is a somewhat limited amount of research in the field of audio steganography and audio stego_analysis. This has to do with the fact that many audio steganography schemes are quite advanced and the nature of high-bandwidth audio streams makes it difficult to produce consistent analyzing tools[10].

Embedding secret messages in digital sound is usually a more difficult process than embedding messages in other media, such as digital images . There are two kinds of tone insertion timing domain tone insertion and frequency domain tone insertion the first one looks like the echo hiding because the weak frequency is added very close after or before the louder voice that make the HAS cant perceive the weak one. The frequency masking witch I will focus on the weak tone is masked cover by high frequency tone.

KaliappanGopalan and Stanley Wenndt[11] proposed paper in tone insertion steganography called "Audio steganography for covert data transmission by imperceptible tone insertion". They rely on the frequency masking. When the HAS cannot perceive the weak tone which cover by high power tone because the low deferential range of the HAS even through the dynamic range covers (80 dB). They use tow deference environments as cover environments.

In the first one experience they use utterances from TIMIT (Texas Instruments Massachusetts Institute of Technology) database as host sample. TIMIT is noise-free database. 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 tow tones f0 (set at 1875Hz) and f1 (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 they compute frame power (FE) and embedding only one bit into frame. If the embedding bit is zero the power of the f0 is set to 25% of the FE and the power of f1 is set to 0.001 of f0 and vice versa if the embedding bit is one the power of f1 set to 25% of the FE and f0 is set to 0.001 of f1 figure 1 illustrate the technique. For recovery the every fame power is computed more over the power p0 and p1 for f0 and f1 respectively is computed. Then calculate the ratio of the power fe/p0 and fe/p1 if the first term greater than the second one then the embedded bit was 0 otherwise the embedded bit was 1.

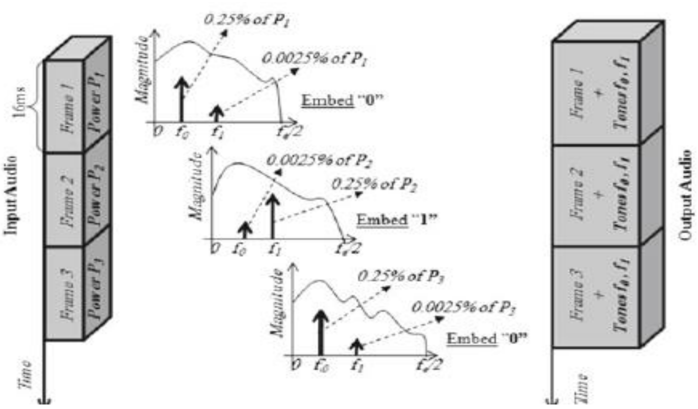


**Figure 2.2: Tone insertion method using two frequencies**

The second experiment is applied firstly in TIMIT utterances then into noise environment. 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 was 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 massage 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 fe the other frequencies are set to negligible power. To embedding set of (0, 0) set the power of f0 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.

They prove the flexibility in terms of imperceptible embedding, high data rate and secure recovery of the proposed method by applied the technique into deferent environments (noisy and free-noisy). The low power levels ensure that the tones imperceptibility in hearing and into spectrogram of the stego signal also conceals the existence of embedded information. These tow feature save the stego-signal from obviously and escrow detection. set the power of the tone frequency to determine value concealing the stego and keep the recovery easy. They test the robustness of the technique against malicious attack. Malicious attack on the stego may render the host message noisy while still carrying the coded covert audio message in a perceivable form. The authors also test the proposed technique against additive noise (Gaussian noise) and against Cropping by zeroing at sample. The author claimed that any attempt to use 8 frequencies to embedded 3 bit per frame lead to audible and/or visible artifacts unless the selected tones are absent in the host audio. Also, noise – intentional or unintentional – may cause high bit errors at high capacity.

KaliappanGopalan, Stanley Wenndt, Andrew Noga, Darren Haddad, and Scott Adams [12] proposed paper in tone insertion steganography called "Covert speech communication via cover speech by tone insertion". It is not deferent from the first paper they use the same frequency and the same utterances. The authors recommend the future work by Choice of Tone Frequencies carefully. It is preferable to select the tones in a mid band. Any attack on the stego in an attempt to destroy the embedded information cannot succeed without destroying the ‘cover’ audio as well. Beside advice to use retrieval technique to reconstruct the cover data if it is distort.

David Wheeler, Daryl Johnson, Bo Yuan, Peter Lutz proposed paper called “Audio Steganography Using High Frequency Noise Introduction". They use different technique to embedded data into injected tone[13]. They relay into the frequency band that cannot perceive by the human heart. The human auditory system can perceive the frequency range from 20Hz to 20 KHz. They claimed that their proposed method is imperceptible to HAS and resistant to preventative software techniques. Authors firstly mention the weakness and strong’s of the following method (LSB, echo hiding, phase coding, and spread spectrum). They relay on the characters of the wave file that rarely include frequency over 20 KHz beside the imperceptible of these frequency to the HAS. The method of encoding binary information in this manner is relatively simple. First, an audio file must be broken into discrete chunks known as frames. The character data to be hidden in the audio is then converted into binary information and mixed into the audio frames. If the embedded bit is 1then add HFT (high frequency tone ) the frame buffer otherwise don’t do so. This process is come after computing frame buffer. For recovery the receiver know the HFT they decode the stego- audio and fetch the HFT into the streaming frame then if found the embedded data was 1 otherwise was 0 then reconstruct the original secret character from these stego.

They implemented the proposed technique in a java program called AudioStego which allowed for encoding of character data into wav-format audio files. The program made use of a WavFile java class for basic input and output operations of wave files and was otherwise implemented with standard java libraries. A multitude of different songs and audio information were tested. Each audio clip featured different dynamic and spectral ranges. All audio was sampled at 44.1 KHz with 16 bits of depth and the length of various audio files ranged from ten seconds up to five minutes in length. They make many test after including HFT every time changing the HFT power and test the stego with many people and record their notices. Finally they recommended using the power size to the HFT into mid band or low band.

The proposed technique is very simple beside they coded it. Also the size of the buffer doesn't change and it is resistant against Hausdorff distances stego- analysis. But the weakness is that this method is not resistant to high pass filter for very high frequency, beside noise and the payload capacity is low.

| **Paper name** | **Most important features** | **Strength** | **weakness** |
|---|---|---|---|
| Audio steganography for covert data transmission by imperceptible tone insertion | *Deferent environments (noisy and free-noisy).<br>*Test the robustness of the technique against malicious attack(Gaussian noise).<br>*Cleared and detailed | *Increase capacity<br>Robustness against cropping and noise<br>*Can hide audio into audio | *PSNR and MSE test.<br>*Still low capacity.<br>*Effects of audio segment into stego |
| Covert speech communication via cover speech by tone insertion | *How to Choice of Tone Frequencies.<br>*Advice to use retrieval technique to reconstruct the cover data if it is distort. | *Robustness against cropping and noise<br>*Recommend to select tone frequency | *PSNR and MSE test.<br>*Still low capacity.<br>*Effect of audio segment into stego |
| Audio Steganography Using High Frequency Noise Introduction | *Inserted tone at HF(high frequency).<br>*Implemented the proposed technique in a java.<br>*Human air test | *Simple and easy to codec | *Collapses in HPF<br>The HTF can traced |

**Table 2.1:Related works comparison**

# Chapter 3

## PROPOSED METHOD AND TOOLS

## 3.1 Overview

From the previous chapters it is clear that the tone insertion method relay on the HAS (chapter one). The related works shows how we can embed data into an audio file using tone insertion method. we proposed new algorithm to increase capacity of the embedded data to more than two bits at single frame. Without increase the number of inserted frequencies to eight frequencies but using just two frequencies.

If we try to increase the inserted tone to eight frequencies in order to embedded three bits in the frame this covert media will be unsecure and it is easy to detect the modification in it by HAS[11].

In this thesis we increase the levels of the power of the inserted tones to three levels rather than two levels into the related work [11][25%,0.001], and decrease the carrier tone to two tones rather than four tones. By using this way we can embed two bits in the frame. Moreover we use third inserted tone as controlled tone (convoy frequency CF) to increase the capacity more than two bits into the frame.

## 3.2 Proposed method

Here we use an English text to conceal it into an audio file. So to find the best way to use convoy frequency CF we study 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). Firstly we decide to use the CF for 0000 and 1111 pattern but we found it is more useful to use it for 0110 pattern. In the next chapter we will make a random set of short text from people and study the feasibility of using the pattern 0110 rather than 0000 and 1111. Moreover we will determine the benefits of CF for that random set in percentage. Note that 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 rather than 0000 and 1111.

We use the stego-table (S.T " SUHAIP.TALAAT") that shows the frequencies and the power of the frequencies related with the embedded bits. table 3.1 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 power.

| **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 |

**Table 3.1 the S.T table**

In case of (0,1) and (1,0) when the power of F1 or F2 is raised to 25% of fe(frame energy)the other frequency is set to level three power. Table 3.2 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.

| Convoy frequency CF | Embedded bits |
|---|---|
| F3 10% fe | The following bit is 0110 |

**Table 3.2 the CF table**

In the embed process after convert text data into binary check the 0110 pattern into the binary if it is found raised the CF to 10% of fe. In the recovery process firstly check the CF if it is 10% of fe extract the 0110 pattern else check the f1p and f2p according to S.T table see figure 3.3 for more information.

## 3.3 Embedding algorithm

From [11] it is clearly that to increase capacity to 2 bit per frame. I have to use four frequency. So I design a new algorithm using differentsetgo- table (mention above).

This algorithm can embed 2 bit per frame using two frequency. Moreover, I use CF table to increase capacity.

1. Convert text file into binary file according to ANSII code.
2. Divide the audio file into frame.
3. Compute 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.

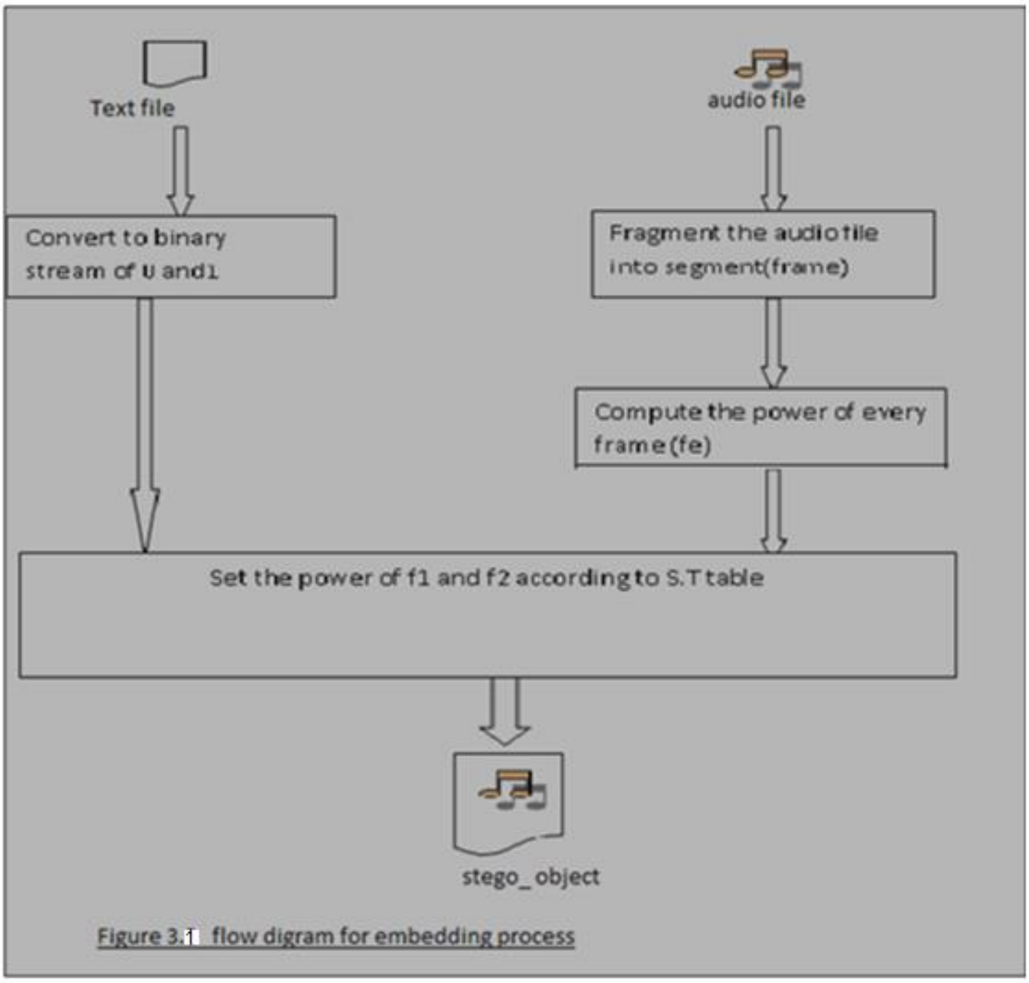

Text file
audio file
Convert to binary stream of 0 and1
Fragment the audio file into segment(frame)
Compute the power of every frame (fe)
Set the power of f1 and f2 according to S.T table
stego_ object
Figure 3.1 flow digram for embedding process

**Figure 3.1: Diagram of embedding algorithm**

# 3.4 Extraction algorithm

1. Divide the stego-object into frame.
2. Compute the power of every frame fe.
3. Check the power of f1 (f1p)

If (f1p=fe/4)
The hidden data is (0,1)
Else if (f1p=15%fe)
Then hidden data is (1,1)
Else if (f1p= fe/100) and (f2p=15% fe)
Then the hidden data is (1,o)
Else if(f1p & f2p= fe/100)
Then the hidden data is (0,0).

4. Go to step 1 for second frame until the end of the text.

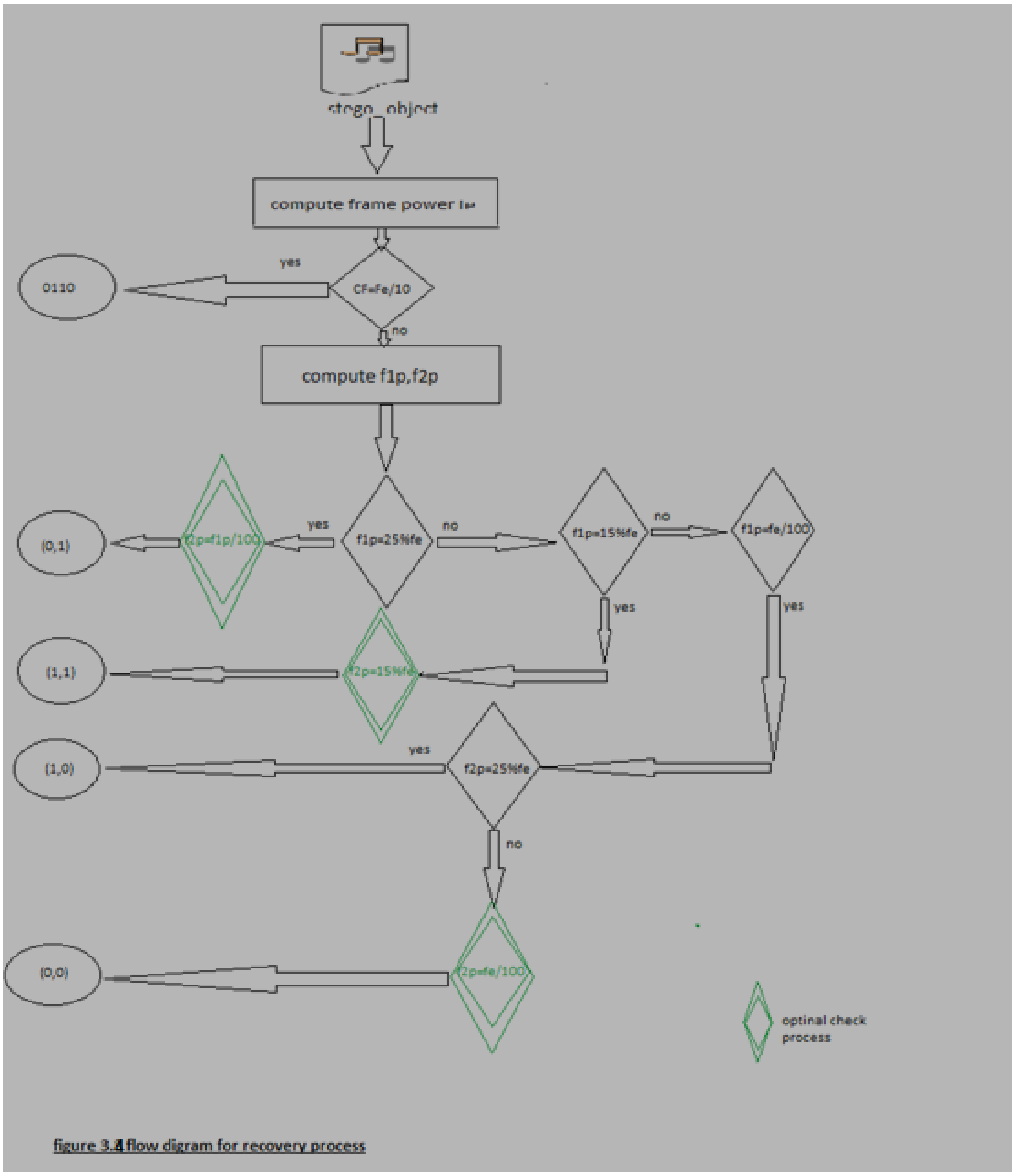


**Figure3.2: Diagram of extraction algorithm**

# 3.5 Application and tools

Matlab is high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Matlab features a family of add-on application-specific solutions called toolboxes. Toolboxes can use in signal processing, control systems, neural networks, fuzzy logic, simulations, and many others fields.

Matlab is well designed to handle digital signal especially the long strings of audio samples. Built-in functions allows most common manipulation to be performed easily.audio recording and playback are equally possible, and the visualization and plotting tools are excellent. In this thesis I decide to use matlabversion 7.8.0.347 on December 2009. I think it is enough to do my job so no more need for the oldest version 2012 which is so big and include many functions and buckets I doesn't need it. Octave environment is also useful to do my job but matlab is more common and Octave lacks in plotting process.

What is the suitable segment length of audio??

There is no direct answer for this question because there are many metrics that affects on the suitable length(example sampling rate and sample bit). Generally, we must know two concepts when we focus on signal processing best segmentation. This concepts are signalstationaryand the second is time-frequency resolution. Most signals requiring analysis are continually changing. A single sustained note playedon a musical instrument is stationary, but quite clearly when one note is replaced by the next one, the signal characteristics have changed in some way (at least in frequency, but possibly also in amplitude, tone, timbre, and so on).For an application analyzing recorded music to determine which note is currently being played, it would make sense to segment the recording roughly into analysis windows of length equal to the duration of a single note, or less.

In order to achieve ahigher frequency resolution, we need to collect a longer duration of samples. However for rapidly changing signals, collecting more of them means we might end up missing some time-domain features. Single FFT can trade off between higher frequency resolution (more samples) or higher time resolution (fewer samples) but cannot do both simultaneously.

# Chapter 4

## RESULT AND DESCUSSION

# 4.1 Results

The main problem has been divided into many sub problems to solve it. These sub problems are:

- How to convert the text data which I need to embed it into audio binary file.
- How to divide the binary file into pairs of possible value(00 01 10 11) at the same time looking for the pattern (0110).
- How to read audio and what is the suitable segment length I can use.
- How to embed data from [2] into the segment according to s.t table chapter three.
- How to write back the stego-audio.
- How can I test the stegofeasibility.
- How can I extract the hidden data from stego and write it back to another file.

Step by step we go through this points and finally we finished the job. Here in this chapter we focus on some point from the above also we make statistical test for random set of document to arrange the feasibility of using the pattern (0110) rather than (1111) or(0000).firstly we focus on this.

To determine the feasibility of using the pattern (0110) .we take random set from different persons( twelve text ). These text has different size. (we use text documents). We design function to calculate the pattern benefit ratio.After that we send these text files one by one to the function and we get this results in table below:

| File name | Size of s.t file | No.of.pattern | Benefit ratio | notes |
|---|---|---|---|---|
| File 1 | 7432 | 502 | 15.6192% | |
| File 2 | 16720 | 964 | 5.7656% | Ascii.table |
| File 3 | 10679 | 1725 | 16.1532% | Mypropose |
| 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 | All typed letter and sample |
| File 12 | 87 | 17 | 19.5402 | Small letters |

**Table 4.1: The feasibility of using the pattern(0110)**

It is clearly from the result on the above table that the feasibility of using the pattern 0110 is approximately 15% .although the pattern 2 benefits is 5.7656% but the numbers files are rarely use conceder able with text one.

I had three different experiments into different host and text all the experiments are use into single channel audio. All the result of stego-audiois the same size of it is original audio.

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

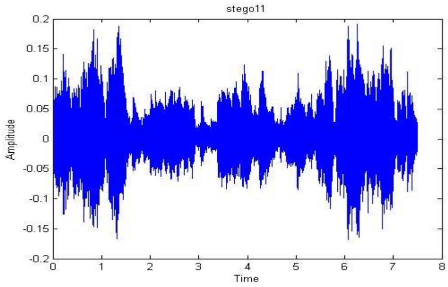
The second experiment we use 14 sec (clean) available into 32000 b/s sample rate at 8 sample as host file. Divide this audio into 256 sample length and use the same text above to hide it into this audio. Also we do it with segment length 128 sample (4 m.sec) and 64 (2 m.sec). the following table shows all experiments. We specify f1,f2,f3 values to be 1750,2250,2750 respectively.

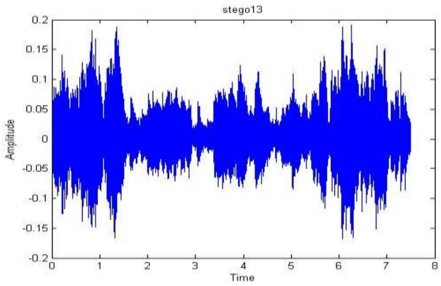
# 4.2 Discussion

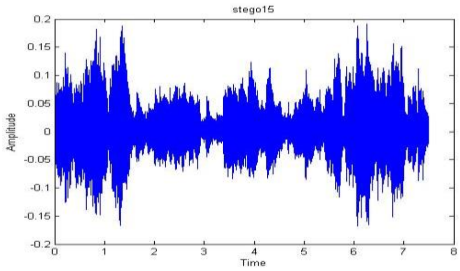
we use three ways to ensure the security of this method these methods are:

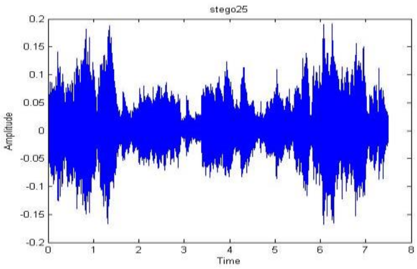
**Spectrogram** of the audio and stego file. After every experiment we plot the original audio and the stego-audio. The figure of the original audio and stego-audio has been checked carefully. Our observations show  no visual differences between the two figure for all experiments.

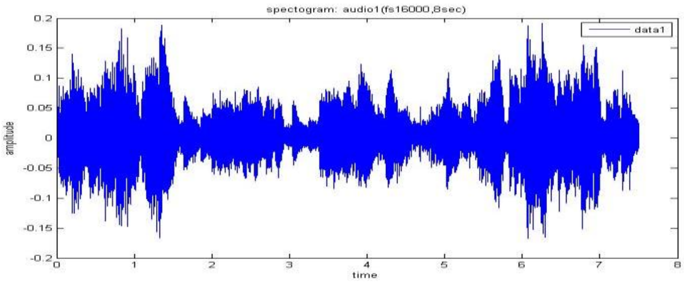
There are seven spectrogram bellow the first one is the original audio and the other are the stego of it according to experiments.



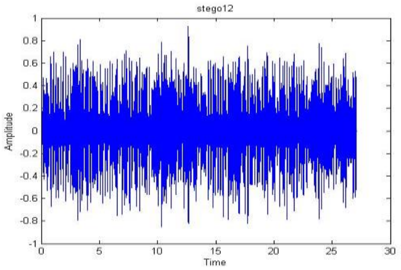


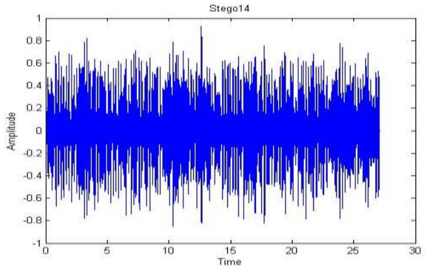


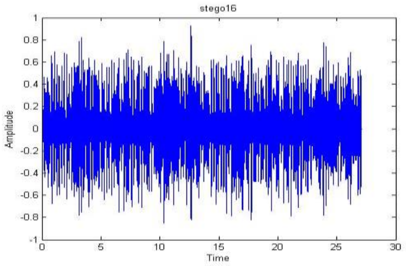


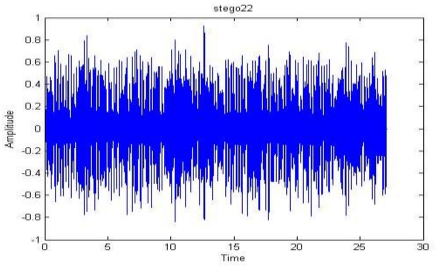


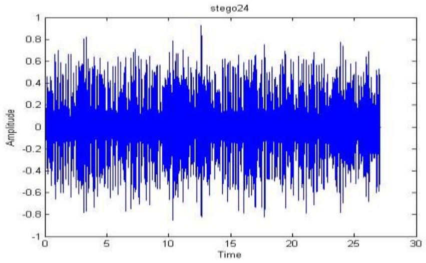
There are another seven spectrogram for the second audio.

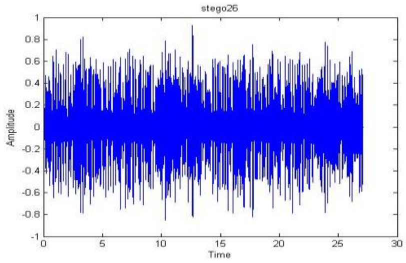


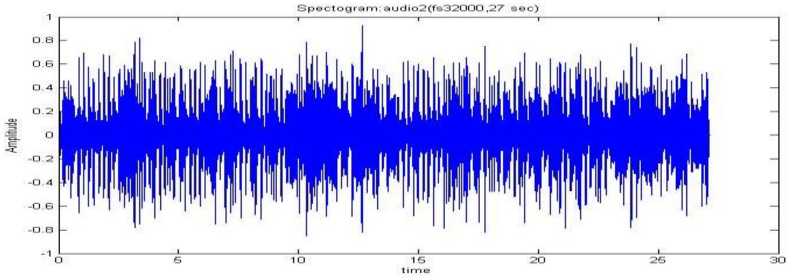












**PSNR**function and**MSE** function for both file below:

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. Because many signals have a very wide dynamic range, (ratio between the largest and smallest possible values of a changeable quantity) the PSNR is usually expressed in terms of the logarithmic decibel scale.

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. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.

| **Audio, stego name** | **Stego details** | **Psnr value** | **MSE value** |
|---|---|---|---|
| Audio1,stego11 | **Stego11(audio1,130byte)** | 121.6298 | 0.000003 |
| Audio1,stego13 | **Stego13(audio1,130byte)** | 125.7655 | 0.000003 |
| Audio1,stego15 | **Stego15(audio1,130byte)** | 125.9546 | 0.000003 |
| Audio1,stego21 | **Stego21(audio1,518byte)** | cannot | |
| Audio1,stego23 | **Stego23(audio1,518byte)** | cannot | |
| Audio1,stego25 | **Stego25(audio1,518byte)** | 122.0325 | 0.000003 |
| Audio2,stego12 | **Stego12(audio2,130byte)** | 103.8790 | 0.000003 |
| Audio2,stego14 | **Stego14(audio2,130byte)** | 111.1388 | 0.000001 |
| Audio2,stego16 | **Stego16(audio2,130byte)** | 114.8495 | 0.0000001 |
| Audio2,stego22 | **Stego22(audio2,518byte)** | 98.2985 | 0.000010 |
| Audio2,stego24 | **Stego24(audio2,518byte)** | 102.4410 | 0.000004 |
| Audio2,stego26 | **Stego26(audio2,518byte)** | 104.9162 | 0.000002 |

**Table 4.2 PSNR and MSE value**

From the PSNR table above it is clear that the PSNR value increase when the segment length decrease and the available capacity increase. And this result make sense because of the frame power is more accuracy. The frame power is usually calculates into median value. The audio signal isn't regular signal. So the small segment gives more accuracy.

**Table 4.3 The experiments detail**

This table shows the experiments detailsand the effects of the segmentation length

# Chapter 5

## CONCLUSION AND RECOMMENDATION

# 5.1 Recommendation

- From the experiments above and others not mentioned in this thesis I had experiences. *So I recommend of*:
- The host audio proprieties must be well-known to you and avoid to use audio's that begin with nothing (0 value).
- If you use speech as host audio. Your generated tone must be in range of speech(1KHz to 3KHz) using music as host you can use greater range of tone frequency.(using median band is recommended).
- Learn your host audio carefully before use it (use FFT or any other transformation) that makes you select the suitable segment. So you can divide host audio into small segments in order to conceal more data but that will effect to the accuracy of the stego.
- If the hidden data is very sensitive and short usestego-key to hide the data into different frame of the audio.
- Don't use generated tone frequencies that are closely near to each other.

# 5.2 Future works

- You can use another CF table to another pattern, or pattern size, or the same pattern for all position, concatenate pattern with space pattern, control characters.
- Use frequency hopping with known key to deny the frequency tracking.
- Use frequency for special letters(enter, space, null).
- Encrypted the covert data with good encryption algorithm.
- Use four frequency to embed 3 bit per frame with 3 levels of power.
- Use the same technique in to another format of audio(mp3).
- Developed this technique to multilevel steganography with packet steganography or other types in order to increase capacity.
- Use tone insertion steganography into VOIP or mobile telephone into real time.

# 5.3 Conclusion

Audio steganography using tone insertion method is very interested failed. Use the generated tone must be at the suitable band (mid band) to strength the stego-object. Also the frequency of tone must chose carefully.

You can use any numbers of level in order to increase capacity but the power of inserted tone must be less than 25% of the frame power. Make segments to process signal. The PSNR value increase when the segment length increase and the available capacity increase.

The stego-object using the proposed method is not observed in spectrogram and the value of PSNR and MSE is good.

# REFERENCES

1. Alothmani, A. (2012, January). IJCS. *A survey on steganography technigues in real time audio signals and evaluation* .

2. B.S.Patil. (2013). *Optimized and secure audio steganography for hidding secret information.*

3. CURRIE D, L. (Oct-1996). Surmounting the effects of lossy compression on steganography. *The 19th national information systems security.* Maryland.

4. E.Zwicher. (1990). *Psychoastics.* Berlin.

5. F.Rocha. (May 2006). *France adaptive audio equalization of rooms based of a technigue of transparent.* Paris.

6. Fatiha, D. (2012). Comparative study of digital audio steganography technigues. *EURASIP Journal on audio,speech and music* .

7. Goplan, K. (2003). Cover speech comunication via cover speech by tone insertion. *Hammond 120th convention.* Hammond.

8. Goplan, K. (2003, March). IEEE. *Covert speech comunication via cover speech by tone insertion* .

9. Goplan, K. (2004, July). WOC. *Audio steganography for cover data transmition by imperceptible tone insertion* .

10. J.R.Krenn. (January 2004). *steganography and stegoanalysis.*

11. Kraetzer. (2008). *Pros and Cons of mel-cepstrum based on audio steganalsis using SVM clasification.*

12. Mitcheel. (1998). Multimedia data-embedding and watermaking technology .

13. Wheeler, D. (2012). *Audio steganographyusing high frequency noise introduction.*