Data Hiding in Audio File by Modulating Amplitude

Dr. Loay. A. Jorj*, Dr. Hilal H. Saleh & Dr. Nidaa F.Hassan****

Received on: 2/ 6 /2008
Accepted on: 2/ 4 /2009

Abstract

In this paper, two methods of a steganography are introduced for hiding secret data in audio media file (.WAV). Hiding in audio becomes a challenging discipline, since the Human Auditory System is extremely sensitive. The first proposed method is used to embed binary sequence with high data rate by modulating the amplitude of WAVE file. The embedding process utilizes the amplitude modulation of the cover signal; the manipulation of the sample depends on its previous sample and next sample. By using this hiding method, good hiding rate is achieved, but it is noticed that the secret data produced by this method does not resist the modifications produced compression.The second suggested hiding methods are oriented to embed the secret data such that it is capable of surviving against modifications produced by compression. This method exploits some of the features of speech signal, more especially the features of the Voiced-Unvoiced blocks. The second proposed hiding method is used to embed secret data by modulating the amplitude of the voiced blocks of cover audio data. Hiding rate is not high as first method since it hiddes only in voiced segments,so it could survive against compression.

Keywords: Data Hiding, Speech signal, Audio compression.

أخفاء البيانات في الملف الصوتي عن طريق تعديل سعة العينات

الخلاصة

في هذا البحث جرى عرض اثنان من الطرائق الجديدة لأخفاء البيانات السرية في ملفات صوتية ذات الأستطالة (.Wav) . أن الأخفاء في الصوت هو بالغ الدقة ,لأن النظام السمعي للأنسان جدا حساس. الطريقة المقترحة الأولى تم أخفاء البيانات السرية عن طريق تعديل سعة عينات الملف الصوتي (WAVE) , ويعتمد التعديل لأية عينة على العينتين السابقة و اللاحقة , وقد حققت هذة الطريقة نسبة جيدة من الاخفاء . لكن لوحظ بأن البيانات السرية المخفية لاتستطيع الصمود امام التغيرات التي تحصل بواسطة الضغط. اما الطريقه الثانية فقد جرى تصميمها لأخفاء بيانات سرية لها القدرة على الصمود أمام التغيرات التي يمكن ان يتعرض له الصوت بواسطة الضغط . حيث تم أستغلال بعض خصائص أشارة الكلام (بدقة اكثر المقاطع الصوتية واللاصوتية) لغرض أخفاء نسب محددة من البيانات السرية. وفي الطريقة الثالثة تم أخفاء البيانات السرية بأستخدام تقنية تعديل سعة عينات المقاطع الصوتية للملف الصوتي (الغطاء) . وقد كان عدد البتات التي يمكن أخفائها ليس عاليا كما في الطريق الاولى لان الاخفاء يتم في المناطق الصوتيه فقط ليصبح لها القدره على الصمود امام الضغط.

* Information Technology, University of Baghdad/Baghdad
** Computer Science Department, University of Technology/Baghdad

https://doi.org/10.30684/etj.28.5.7
2412-0758/University of Technology-Iraq, Baghdad, Iraq

1. Introduction

Steganography is the art and science of hiding the fact that communication is taking place.

Using Steganography, you can embed a secret message inside a piece of unsuspicious information and send it without anyone knowing

Of the existence of the secret message [1].The onset of computer technology and the Internet have given new life to Steganography and many creative methods with it are employed. Computer-based steganographic technologies introduce change to digital carries to embed foreign information to the native carries. Carries of such messages may resemble innocents sound, text, disks, network traffic and protocols, the way software or circuits are arranged, audio, images, video, or any other digitally code or transmission [2].

Data hiding is the general process by which a discrete information stream is merged within media content by imposing imperceptible changes on the original host signal. Compression is one of the most common operations on digital files; therefore, we must take into account the effect of compression when designing information hiding. There must be an appropriate compromise between data hiding in audio and perceptual coding. Traditionally, data hiding and compression have had contradictory goals. The former adds perceptually irrelevant information in order to embed data, while the latter removes this irrelevancy and redundancy to reduce storage requirements [3]. So, one of the main obstacles within the data hiding in audio is to develop a scheme which is robust to perceptual coding standard (MP3). MP3 is one of many methods to compress audio in digital form trying to consume as little space as possible but keep audio quality as good as possible. MP3 is one of the best achievements in this area [4].

In this paper, two methods of a steganography are introduced for hiding secret data in audio media file (.WAV). The first proposed method is used to embed binary sequence with high data rate by modulating the amplitude of WAVE file. By using this hiding method, good hiding rate is achieved, but it is noticed that the secret data produced by this method does not resist the modifications produced by (MP3) compression. In the second method, the process of secret data embedding will be performed before passing through the MP3 compression encoder, to determine whether secret data will be able to survive against this lossy compression scheme. The first stage in designing of this method, statistical and analytical investigation are performed to assess the difference which may occur in the WAVE audio when it is subjected to different levels of MP3 compression (128 kbps , 96 kbps, and 64 kbps). The results lead to determine that the cover file should at level 128 kbps, in such a case the similarity between the original and compressed file will be very enough to get robust hiding case. The binary sequence is embedded by modulating the amplitude of voiced blocks of WAVE file.

These methods are blind methods, they do not need original audio file in the extraction stage; and they work in the time domain exploiting some features of speech signal.

2. Audio Signal Classification (ASC)

Audio signal classification (ASC) consists of extracting physical and perceptual features from a sound, and of using these features to identify into which set of classes the sound is most likely to fit. The feature extraction and classification algorithms used can be quite diverse

depending on the classification domain of the application [5].
The first step in any classification problem is to identify the features that will be used to classify the data. Feature extraction is a form of data reduction. It is sometimes left to the brute power of an algorithm to decide which features are the most important [6].
The features typically used in ASC can be divided into physical and perceptual categories.

2.1 Physical and Perceptual Features

Physical features are properties that correspond to physical quantities, such as fundamental frequency (F0), Energy (EN), Zero-crossing rate (ZCR) and Modulation rate. In this research Energy (EN) and Zero-Crossing rate (ZCR) were utilized, these two features are used to make a simply discriminate between Voiced / Unvoiced (V/UV) audio segments [6].

A. Energy Measurement

One of the simplest representations of a signal is its energy. In the case of a real discrete-time signal x (n), the energy is defined in general in equation (1):

$$EN(n)= \sum_{n=-\infty}^{\infty} x(n)^2 \quad \ldots. (1)$$

For nonstationary signals such as speech, it is often more appropriate to consider a time-varying energy calculation such as the following equation:

$$EN(n)= \sum_{n=0}^{N-1} x(n)^2 \, W(n) \quad \ldots.. (2)$$

Where N is the number of samples, n is the nth sample in the frame.
The difficulty involves is the choice of window size. The purpose of the window is to attach lower weight speech samples, which occurred further back in time. If N is too small, EN will fluctuate very rapidly depending on exact details of the waveform. If N is too large, EN will have very little variation, and will not reflect the changing properties of the speech signal. [7]
The major significance of Energy Measure (EN) is that it provides a good measurement for separating voiced speech segments from unvoiced speech segments. EN for unvoiced segments is much smaller than for voiced segments [8].

B. Zero-Crossing Measurement

Another very simple time-domain analysis method is based on the zero crossing measurements. In the context of a digital implementation, a zero crossing can be said to occur between sampling instants n and n-1 as equation (3):

$$\text{Sign}[x(n)] \neq \text{Sign}[x(n-1)] \ldots (3)$$

Zero crossing rate measures how often the sound signal crosses from positive to negative or vice-versa. Zero crossing measurements (along with energy information) are often used in making a decision about whether a particular segment of speech is voiced or unvoiced. If the zero crossing rate (ZCR) is high, the implication is unvoiced; if the zero crossing rate is low, the segment is most likely to be voiced. Speech signals are broadband signals and the interpretation of average zero-crossing rate is therefore much less precise [8].

4. Hiding in Audio

Data hiding in audio signal is especially challenging because Human Auditory System (HAS) perceives over a range of power greater than one billion to one and range of frequencies greater than one

thousand to one. In addition, the auditory system is very sensitive to additive random noise. Any disturbance in a sound file can be detected as low as one part in ten million (80db below ambient level) [9].

When performing data hiding on audio, one must exploit the weaknesses of the HAS, while at the same time being aware of the extreme sensitivity of the human auditory system. While (HAS) has a large dynamic range, it has a small differential range, and as result, loud sound tends to mask quiet ones. Additionally, the HAS does not perceive absolute phase, but only relative phase. Finally, there are some environmental distortions so common as to be ignored by the listener in most cases. These "holes" can be exploited by data hiding techniques [10].

5. The Proposed Data Hiding Methods:

5.1 Amplitude Modulation (Not Robust To MP3 Compression)

This method could be used for hiding high data rate, but the embedded secret data cannot survive against modifications produced by MP3 compression standard. This hiding method operates in the time-domain. Processing audio signal in time-domain has the advantages of simplicity, quick calculation and easy physical interpretation. The proposed hiding algorithm consists of two-modules:

1. Embedding Module
2. Extracting Module.

5.1.1. Embedding Module

The embedding process utilizes the amplitude modulation of the cover signal. In this method, the manipulation of current sample depends on its previous sample and next sample. The current sample is set to be the mean of previous sample and the next sample plus delta, and this delta will be divided into segments depending on step value. The suggested embedding process can be summarized as follows:

For each three successive host samples (S1, S2, and S3), determine:

a. Mean = (S1+S3)/2
b. Δ = S2- Mean
c. Q= (Round (Δ/Step) and $FFFE)
d. Δ = $\begin{cases} (Q+1)\times \text{Step if embedded bit} = 1 \\ Q\times \text{Step if embedded bit} = 0 \end{cases}$
e. S2= Mean+Δ

5.1.2. Extracting Module

The extraction process can be summarized as follows:

For each three successive stego samples (S1, S2, and S3), determine:

a. Mean = (S1+S3)/2
b. Δ = S2- Mean
c. Q= Δ/Step
d. Secret Bit = $\begin{cases} 1 \text{ if (Q and 1)}=1 \\ 0 \text{ if (Q and 1)}=0 \end{cases}$

5.2 Amplitude Modulation (Robust to MP3 Compression)

The third method is introduced for hiding secret data in the cover audio data by modulating the amplitude of the samples in this cover. This method also utilizes time-domain modification of audio, by modifying the amplitude of voiced blocks in the cover; information can be undetectably encoded in it. Modifying amplitude of current sample will depend on the previous and the next samples of the current one. This embedding method

will be robust against MP3 compression. Amplitude modulation hiding secret data in voiced blocks consists of the two phases, and they are:

1. Embedding Module.
2. Extracting Module.

5.2.1 Embedding Module

This module is concerned with embedding secret message in the voiced blocks of WAVE audio file, and it consists of the following stages:

1. Computation of short average energy.
2. Energy shifting.
3. Merge unvoiced blocks.
4. Embedding process.
5.

1. Computation of short average energy

In this stage, the average energy is computed for each segment of audio data cover. The vector (Wav) represents data of the audio cover, it is segmented into the blocks (w), and the average energy for each block is estimated. The decision of whether a block is voiced or unvoiced is evaluated by following equation:

Step 1: Determine

AE (i)

$$= \sqrt{\frac{1}{w} \sum_{j=i}^{j=i+w} (Wav(j) - 128)^2} \quad \ldots (4)$$

Step 2: If AE (i) < Threshold then the block is unvoiced

Else the block is voiced.

Once the average energy of each block was estimated, Voiced / Unvoiced criteria that are based on comparison between the average energy and threshold is applied.

All blocks with average energy greater than threshold value are labeled as voiced segments, and those blocks with average energy equal or less than threshold value are labeled as unvoiced segments. When the threshold value is increased, then the number of unvoiced blocks increases and vice-versa. An illustrative example of voiced / unvoiced segmentations is shown in Figure (1). The original audio signal is shown in (a); only the voiced segments are shown in (b), the unvoiced segments shown in (c), and finally the original audio signal with voiced and unvoiced segments together are shown in (d). The detection of this example is applied to audio sample, the block size=100, threshold value =3 resulting, voiced blocks= 81, unvoiced blocks=203.

2. Energy shifting

In the previous stage, the determination of whether a block (win) is voiced or unvoiced is accomplished. However, errors occur when this file is passed through MP3 compression. These errors may cause a decrease or increase in the energy of some blocks, which in turn will cause an error in the detection of (V/UV) blocks, i.e., block is detected as voiced before compression but after compression it is detected as unvoiced and vice-versa. The errors occur in blocks whose average energy values are very close to the threshold. To avoid the occurrence of overlapped margin caused by compression, we have to shift all the samples of the block whose energy is close to the threshold far away toward the voiced regions.

3. Merge unvoiced blocks

After allocating the value of threshold, the merging of voiced / unvoiced blocks (V=0/UV=1) will be accomplished. Any sequence of similar detected blocks is established by combining the successive similar (V/UV) blocks; the produced

segments will have variable length, this stage leads to pack the number of (V/UV) blocks, but without making any overlapping between them.

4. Embedding process

The embedding process is the same process which has been used in the first amplitude modulation method, but the difference here is the secret bit is embedded in the successive voiced samples (the candidate samples for hiding secret bit lie in the voiced blocks).

Some parameters must be considered during the embedding process, they are:

1. The step value.
2. The number of host samples (within a voiced block).
3. The number of embedding repetition.

Different step values have been tested in order to reach near optimality of bit reconstruction.

The number of host samples (within a voiced block) has significant effect, since the embedding of binary sequence in voiced block from the first sample to the last one may cause an audible noise in stego cover (voiced block entirely hosted). To get imperceptible hiding in voiced blocks, the voiced block could be a partial host.

After embedding the secret message, some error is noticed in the extracted bits from stego file. To overcome this problem the secret message is hidden for several times in order to improve the probability of correct extraction of secret data.

5.2.2 Extracting Module

This module is used for extracting the secret message from WAVE file (stego), and it consists of the following stages:

1. Skip over part of data in stego file.
2. Computation of short average energy.
3. Voiced / Unvoiced blocks detection.
4. Merging unvoiced blocks.
5. Extracting bits.
6. Matching the extracted vectors of secret bits.

The audio data is skipped by 1200 byte, this is due to the results (differences) found when the original audio data is compared with corresponding construct data after it is compressed by using MP3 compression standard.

The (V/UV) blocks will be classified by applying the same sequence of the embedding module; each block of stego file is tested to find out whether it is voiced or unvoiced block. In the extraction process, the voiced blocks are scanned in order to extract the binary bits of the secret message.

The extraction process will produce vectors of binary sequence; taking in consideration that the secret data is hidden several times. Each vector represents secret message in binary form. The repeated secret binary vectors produced in the extraction process will be matched bit by bit in order to extract the correct secret message without an error. The matching process has been applied to characters (i.e., matching character by character) but it did not lead to good results like in the case of bit-by-bit matching process.

Finally, the string of binary bits produced in the matching process is converted into bytes to give the secret message.

6. Experimental Results

Distortion measures are used to measure the amount of error in the stego audio, in other words, they are useful measures to compare between

the stego audio and cover audio, and they offer a simple convenient tool for evaluating the information loss.

Two classes of distortion measures are applied on original audio cover and stego audio file, they are:

A. Objective Measures

The objective measures are often used in audio coding researches because they are easy to determine, but these measures are not necessarily correlated with our perception of an audio. These measures are all based on the differences between the original, undistorted audio file (cover) and the modified, distorted audio file (stego).

Table (4.14) lists the most commonly used distortion measures which are adopted for testing the performance of the suggested systems, and they are MAE, MSE, PSNR and SNR. These measures define the overall error between (N) samples of the original file (Wav), and the corresponding samples of the reconstructed file (RecWav).

B. Subjective Measure

This kind of measures requires a definition of a qualitative scale to assess audio quality. This scale can then be used by some human expertise to subjectively determine the fidelity. These tests imply, listening to the original and the modified audio sequence and reporting dissimilarities between the two signals, using a 5- point impairment scale: Imperceptible, Perceptible (not annoying), Slightly annoying, Annoying, Very annoying [11].

7. Test Samples

Audio files, music, speech and song were used as test samples to assess the performance of the proposed hiding methods. The sizes of test samples are listed in table (3).

All these test samples are audio files of WAVE type, PCM format, one channel (mono), and 8-bit sample size.

Test 1

In this test, the secret message is hidden in the WAVE audio files by the amplitude modification method.

Table(4) illustrates the objective and subjective results after hiding by using amplitude modulation method. Figure (2) shows the audio signal before hiding and after hiding using amplitude modulation method.

Test 2

In this test, the hiding of secret message in the audio cover is achieved by applying amplitude modulation to the voiced regions of the cover audio signal; the following parameters are used as control parameters:

1. **Step Value:** represents the quantization step value.
2. **Hiding Repetitions:** represent the number of times that the secret message is hidden in the cover audio signal.

Three audio data set are used as test samples, Speech_1, Song_2 and Music_2. Table (5) shows the objective and subjective test results for the cases of steganography by using amplitude modulation of voiced regions (for three test samples); the size of secret data is taken to be (1322 bit).Figure (3) shows the audio signal before hiding, after hiding using amplitude modulation of the voiced regions.

7. Conclusions

In this paper, the secret data is embedded by modulating the amplitude of the cover audio data. In the first method hiding is successful in hiding high data rate of bits (about 0.33 bps),but secret data hidden into audio data cover by using this method cannot resist against the

deformations formed by using the lossy compression standard (like MP3).In second method, the secret data is embedded by modulating the amplitude of the voiced regions (amplitude modulation) of the cover, so hidden secret data is capable of surviving against modifications produced by MP3 compression standard, at compression level 128 kbps, in this method if the quantization step is small, some of hidden bits cannot resist against the deformation formed by MP3 attacker, it will cause an error in extracted secret data, but if the quantization step value is large then the hidden bits could survived against MP3 attack. Repeating the message embedding along the audio cover will improve the probability of correct extraction of message, in comparison with the case of using just one embedding. Although determined SNR values of stego file are low, it has been perceived to be very good, since it depends on music, songs audio files (with loud pitch) as covers for hiding secret data.

8. References

[1] J.R. Krenn, "Steganography and Steganalysis", January 2004.URL: http://www.tifr.res.in/~sanyal/papers/Soumyendu_Steganography_Steganalysis_different_approaches.pdf

[2] Johnson N.F., Duricn Z., Jajodia S., "Information Hiding: Steganography and Watermarking Attack and Countermeasurments",Kluwer Academic Publishers, USA, 2001.

[3] Lan T., Mansour M.F.,Tewfik A.H., "Robust High Capacity Data Embedding", IEEE, Vol.4314, ppp.329-335, 2000, URL: www.spie.org/web/abstructs/4300/4314.html

[4] Supurovic P., "MPEG Audio Compression Basics", URL: http://www.chested.chalmers.se\~kf96svgu\, 1998.

[5] Cassidy S., "COMP449: Speech Recognition", Department of Computing, Macquarie University, Australia, 2002, URL: www.comp.mq.edu.au/~cassidy/comp449/html/comp449.html.

[6] Gerhard D., "Audio Signal Classification: An Overview", School of Computing Science, Simon Fraser University, Burnaby, BC, V5A 1S6, 2002, URL: http://www.cs/unr.edu/~jli/cs691.htm

[7] Gruhl D., Lu A., Bender W., "Echo Hiding", Proc. First Workshop on Information Hiding, Lecture Notes in Computer Science, Vol.1174, pp. 295-315,Springer 1996.

[8] Bouman C. A., "EE438-Laboratory 9: Speech Processing", Purdue University: EE438- Digital Signal Processing with Application, October 16, 1998.

[9] Bender W., Gruhl D., Morimoto N., Lu A., "Techniques for Data Hiding ", IBM System Journal, Vol. 35, No. 3&4,1996,URL: http://isj.www.media.mit.edu/isj/SectionA/313.pdf.

[10] Sellars D., "An Introduction to Steganography", University of Camberge,2003,URL: http://www.cs.uct.ac.za/courses/CS400W/NIS/papers99/dsellars/stego.html.

[11] Cvejic N., Seppanen T., "A wavelet Domain LSB Insertion Algorithm for High Capacity Audio Steganography", IEEE 0-7803-8116, 2002, URL: http://www.mediateam.oulu.fi/public/pdf/360.pdf

Table (1) Distortion measures.

Mean Absolute Error	$MAE = \frac{1}{N}\sum_{i=0}^{N}\left\|Wav(i) - \mathrm{Re}\,cWav(i)\right\|$
Mean Square Error	$MSE = \frac{1}{N}\sum_{i=0}^{N}\left(Wav(i) - \mathrm{Re}\,cWav(i)\right)^2$
Signal to Noise Ratio	$SNR = 10\log_{10}\left(\frac{\frac{1}{N}\sum_{i=0}^{N}Wav(i)^2}{MSE}\right)$
Peak Signal to Noise Ratio	$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right)$

Table (2) Impairment scales [9].

Scale	Impairment Scale	Quality
1	Imperceptible	Excellent
2	Perceptible, not annoying	Good
3	Slightly annoying	Fair
4	Annoying	Poor
5	Very annoying	Bad

Table (3) Tested audio files.

File Name	WAVE Format	MP3 Format
	Size (Byte)	Size (Byte)
Speech_1	69,876	27,483
Song_1	8,103,218	2,941,913
Song_2	5,902,466	2,143,611
Music_1	763,876	279,095
Music_2	2,871,836	1,043,960

Table (4) Objective and subjective test results for some stegocover audio samples produced by using amplitude modification method.

Sample	Length of secret message	Hiding Rate (bps)	MAE	MSE	PSNR	SNR	Impairment Scale	Quality
Speech_1	21265	0.304	0.39	0.83	48.95	43.09	Imperceptible	Excellent
Song_1	21265	0.003	2.28E-03	4.57E-03	71.53	65.90	Imperceptible	Excellent
Music_1	21265	0.032	3.19E-02	6.36E-02	60.09	54.13	Imperceptible	Excellent

Table (5) Objective and subjective test results of audio hiding method based on amplitude modulation method of voiced regions.

Audio Name	Length of secret message	Hiding Rate (bps)	MAE	MSE	PSNR	SNR	Impairment Scale	Quality
Speech_1	1322	0.05	18.74	1047.48	17.93	12.07	Imperceptible	Excellent
Song_2	1322	0.0006	27.41	1376.02	16.74	10.93	Imperceptible	Excellent
Music_2	1322	0.001	24.18	1855.44	15.45	9.688	Slightly annoying	Fair

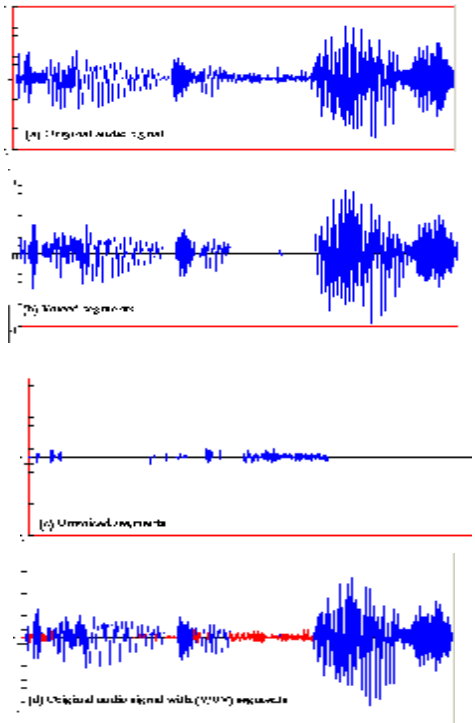
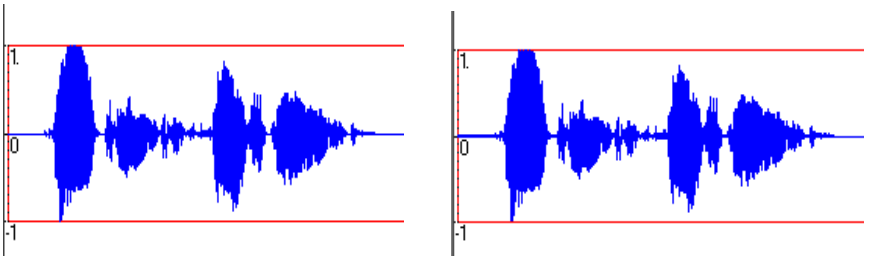
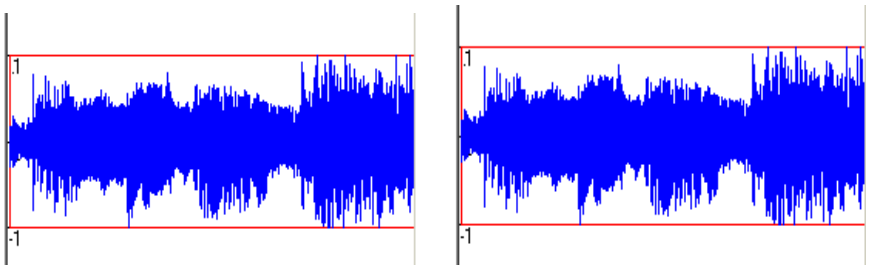


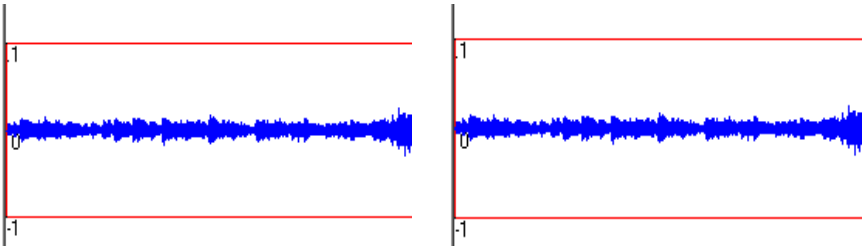
Figure (1) Voiced / Unvoiced segmentations.



(a) Speech_1



(b) Song_1



(c) Music_1

Figure (2) The test samples Speech_1, Song_1 and Music_1 signal before and after hiding by using amplitude modulation method.

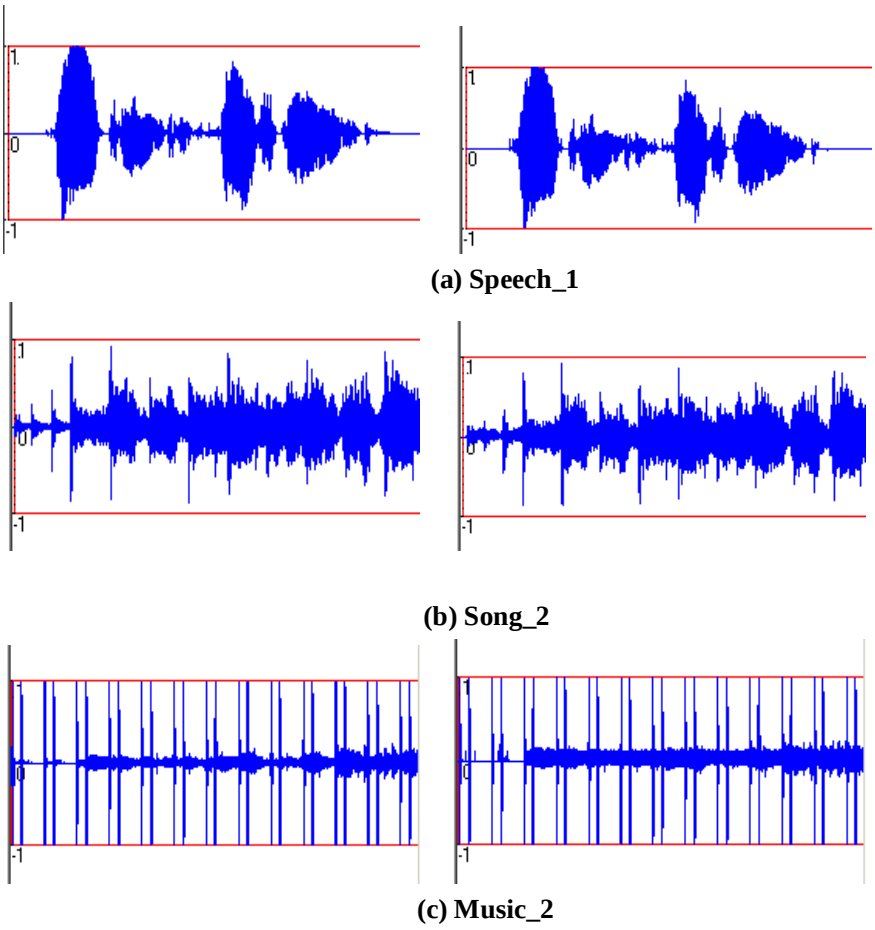


Figure (3) The test samples Speech_1, Song_2 and Music_2 signal before and after hiding by using amplitude modulation of voiced regions.