Text Data Fusion to Speech Signal Using Hash Function Algorithm

Rajinder kaur, Gurwinder kaur

Abstract—Proposed paper work concentrates on the secure transmission of text data when digital contents are commonly distributed by Internet and more security attacks are appeared from eavesdropping, data exposure and data tampering. In this paper we concentrate on hiding some secret text into a real time speech signal considering PSNR (Peak-Signal-to-Noise Ratio) and duration of speech signal as key parameters for hiding data. Text message is intended to appear as unwanted noise in speech signal to the unintended recipients and this process leads to a way direct to covert communication. A function called hash function is applied on text data before embedding it into speech signal which includes text data division into small chunks and interchanging their places in a specified manner to enhance the security level and intended recipients are supposed to extract the secret text with predetermined hash key. In this way this paper purposes a technique which can be called a cousin of steganography where a secret information is transferred by hiding it into another unsuspicious cover media such as image, audio or video signal to conceal actual existence of secret message.

Index Terms - Data security, Hash function, Speech signal, Steganography

I. INTRODUCTION

The data hiding technique like digital watermarking etc. has been required and hence developed for the purpose of annotation, authentication and copyright protection. In concern to the data hiding performance, imperceptibility of embedding data and its robustness against common signal processing or security attacks is required. Data hiding in audio signals exploits imperfections of human auditory system (HAS). The existing techniques for hiding data within audio signals includes Spread Spectrum (SS), Least Significant Bit (LSB), Echo Hiding (EH), Phase Encoding (PE), Spectrum Transform (ST) etc. [4].

Acoustic data transmission needs a fast and capable system for hiding information of any kind into any type of arbitrary signal. There are many applications using these techniques like digital rights management, authentication of secret voice message in military and intelligence operations, proof for authorship of musical works, hiding executables for access control, covert communication, annotation etc. [5].

Lossless or reversible data hiding method has been receiving much attention in recent few years. Unlike traditional watermarking methods which make sacrifice of an imperceptible amount of host data, lossless data hiding algorithm has an extra advantage that, after extraction of watermark, the host data can be recovered exactly [7].

Steganography is an art and science used for secret communication. A steganographic scheme embeds secret data in cover data e.g., digital images such that unintended recipients cannot even notice it. On the contrary, steganalysis is used to detect that whether any secret data is present in a given medium. Furthermore, steganalysis serves the way to judge the security level of a steganography technique. [8].

II. LITERATURE SURVEY

Rashid Ansari et al., [200]. This paper has proposed a data hiding technique where digital audio exploits the low sensitivity of the human auditory system to phase distortion. Inaudible but controlled phase changes were introduced in the host audio using a set of all pass filters (APFs) with distinct parameters of all pass filters, i.e., pole-zero locations. The APF parameters were chosen to encode the embedding information. During the detection phase, the power spectrum of the audio data was estimated in the -plane away from the unit circle. The power spectrum was used to estimate APF pole locations, for information decoding. Experimental results show that the proposed data hiding scheme could effectively withstand standard data manipulation attacks. Moreover, the proposed scheme was shown to embed 5–8 times more data than the existing audio data hiding schemes while providing comparable perceptual performance and robustness [1].

Deng Lixin, [2004]. In this paper, an approach for data hiding within speech signals based on Hash and Hilbert Transform (HT) has been proposed. The secret data was firstly pre-processed by Hash to enhance its security. Then, exploiting the orthogonality of HT and insensitivity of human perception to the phase of speech, they embedded processed secret data into a speech signal by using the HT. In this paper both blind and non-blind methods to extract embedded data along with experimental results demonstrated transparent, secure and robust data hiding performances [2].
Mohamed F. Mansour et al., [2005]. They proposed a new framework for data embedding in audio in this paper. The basic idea of this algorithm was to change the length of the intervals between salient points of the audio signal to embed data. The intervals were quantized and the data was embedded in the Quantization indices. In this particular implementation, they used the wavelet extreme of the signal envelope as the salient points. They proposed novel ideas for practical implementation that can be used by other data embedding schemes as well. The proposed algorithm was robust to common audio processing operations, e.g.: MP3 lossy compression, low pass filtering, sampling rate conversion, and time-scale modification (TSM) [3].

Jing Liu et al., [2008]. Considering chaotic synchronization, the TCP protocol was usually adopted to transport encrypted data on networks. A block encryption algorithm was proposed to digital speech codes in this paper. The proposed method used the UDP protocol to cipher text. It partially solved the problem of decryption for receiver when some data packages were lost during transportation. It encrypted message with chaotic sequences which randomly came from chaos model database, so the randomness of chaotic sequence was enhanced greatly. Furthermore, it could overcome the disadvantages of short period when data amount was great. The algorithm was testified with testing program written by the authors [6].

Marina Ponomar, [2009]. This paper was focused on crypto security for data hiding in electro acoustic speech signals, and an example of data hiding in a speech signal based on segment fundamental frequency modification was described. In the process of embedding of data into speech artificial distortions could be perceivable for human ear. The author set a task to determine acceptable limits of modification to maintain optimal security level and natural speech communication [9].

Fatiha Djebbar et al., [2011]. Steganography has been proposed as a new alternative technique to enforce data security. Lately, novel and versatile audio steganographic methods have been proposed. A perfect audio Steganographic technique aim at embedding data in an imperceptible, robust and secure way and then extracting it by authorized people. Hence, up to date the main challenge in digital audio steganography is to obtain robust high capacity steganographic systems. Leaning towards designing a system that ensures high capacity or robustness and security of embedded data has led to great diversity in the existing steganographic techniques [10].

III. REAL TIME SPEECH SIGNAL FEATURES

In this paper real time speech signal is used as an input signal that is used as a cover media for text data signal to be fused and transmitted for maintaining its security so that attackers or unintended users could not detect existence of the text data in speech signal. Here we use exploitation of human auditory system (HAS) which is not able to detect minor changes in speech signal and consider them as noise and ignores them.

1. Sampling Frequency

Here for the particular purpose we take sampling frequency of speech signal as 8khz because signals sampled at this frequency are best when input is given through microphone because for standard applications in telecommunication ‘Bandwidth’ of ‘0.3……….3.4 kHz’ digital signal is represented at a sampling frequency 8 kHz. According to ‘Sampling Theorem’, ‘A band limited signal can be reconstructed exactly if it is sampled at a rate at least twice the maximum frequency component in it.’

2. Signal Time

Signal time depends upon the size of text data to be fused. More long will the speech signal, more data it will be able to carry without any complication.

3. PSNR (Peak Signal-to-Noise Ratio)

It is easily defined with the help of the mean squared error (MSE). Typical values of PSNR in lossy image and video compression remains between 30 and 50 dB, but also higher value of PSNR is preferably better. Acceptable values of PSNR for wireless transmission quality loss are considered to be between 20 dB and 25 dB approximately. Zero value of MSE leads to undefined PSNR refers to the fact that the two images are identical.

IV. HASH ALGORITHM FOR TEXT FUSION

It will proceed in multiple small steps from taking real time speech input to extract that speech signal. These steps can be explained as follows,

- First of all we take real time speech signal. Input is taken through microphone at sampling frequency 8 kHz.
- Reading text data to be hide; It is any text file placed anywhere in our computer to be fused in speech signal. File should be in ‘.txt’ format only. We will combine all lines of this file to form a string and convert it into ASCII format.
Now compute length of text data i.e. number of bytes we want to hide and it will be displayed on results window.

In next step computation of threshold is done. Threshold value is the limit in speech signal to which text data will be fused. Its value lies between ‘127+a and 127-a’ where value of a will be increased from 1 depending upon size of text data.

Hiding data in speech with Blind Source Separation; here first we divide total text data (in ASCII format) into small data chunks each of size 5 bytes. Let us name each chunk as DC.

Reconstruction of original signal; Intended receiver will receive and reconstruct the signal with the help of predetermined HASH key.

V. RESULTS AND DISCUSSIONS

With increase in data size, MSE increases which in turn decreases PSNR Value. We take three tables showing values of data size, threshold value, MSE (Mean Square Error) and PSNR (Peak Signal-to-Noise Ratio) for signal time 1 second, 2 second and 3 second. Comparison of these three tables will help us to know that how signal time affects the PSNR (Peak Signal-to-Noise Ratio) values. One can notice that longer is the speech signal’s duration, more is PSNR value, make more easy reconstruction of text data from speech signal.

Table 1 Table of signal parameters for increasing data size for time = 1 second

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Data Size (in Bytes)</th>
<th>Threshold Value</th>
<th>MSE(Mean Square Error)</th>
<th>PSNR(Peak Signal-to-Noise Ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84</td>
<td>2</td>
<td>18.5547</td>
<td>35.4509</td>
</tr>
<tr>
<td>2</td>
<td>641</td>
<td>2</td>
<td>135.1806</td>
<td>26.8217</td>
</tr>
<tr>
<td>3</td>
<td>1522</td>
<td>3</td>
<td>333.3288</td>
<td>22.9021</td>
</tr>
<tr>
<td>4</td>
<td>3603</td>
<td>4</td>
<td>854.6031</td>
<td>18.8132</td>
</tr>
<tr>
<td>5</td>
<td>5819</td>
<td>6</td>
<td>1.5741e+00</td>
<td>16.1605</td>
</tr>
</tbody>
</table>

Table 2 Table of signal parameters for increasing data size for time = 2 second

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Data Size (in Bytes)</th>
<th>Threshold Value</th>
<th>MSE(Mean Square Error)</th>
<th>PSNR(Peak Signal-to-Noise Ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84</td>
<td>2</td>
<td>9.2674</td>
<td>38.4612</td>
</tr>
<tr>
<td>2</td>
<td>641</td>
<td>2</td>
<td>67.5903</td>
<td>29.8320</td>
</tr>
<tr>
<td>3</td>
<td>1522</td>
<td>2</td>
<td>165.3469</td>
<td>25.9468</td>
</tr>
<tr>
<td>4</td>
<td>3603</td>
<td>3</td>
<td>421.9138</td>
<td>21.8786</td>
</tr>
<tr>
<td>5</td>
<td>5819</td>
<td>4</td>
<td>776.1112</td>
<td>19.2316</td>
</tr>
</tbody>
</table>
Table 3 Table of signal parameters for increasing data size for time = 3 second

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Data Size (in Bytes)</th>
<th>Threshold Value</th>
<th>MSE (Mean Square Error)</th>
<th>PSNR (Peak Signal-to-Noise Ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84</td>
<td>2</td>
<td>6.1783</td>
<td>40.2221</td>
</tr>
<tr>
<td>2</td>
<td>641</td>
<td>2</td>
<td>45.0602</td>
<td>31.5929</td>
</tr>
<tr>
<td>3</td>
<td>1522</td>
<td>2</td>
<td>110.2313</td>
<td>27.7078</td>
</tr>
<tr>
<td>4</td>
<td>3603</td>
<td>3</td>
<td>281.8956</td>
<td>23.6299</td>
</tr>
<tr>
<td>5</td>
<td>5819</td>
<td>3</td>
<td>515.9761</td>
<td>21.0045</td>
</tr>
</tbody>
</table>

Now for MSE (Mean Square Error) and PSNR (Peak Signal-to-Noise Ratio) we plot a graph where we can see that with increase in data size, MSE (Mean Square Error) increases and PSNR (Peak Signal-to-Noise Ratio) decreases. In MSE graph, data size is taken in bytes on x-axis and MSE (mean Square Error) is taken on y-axis and same as for PSNR (Peak Signal-to-Noise Ratio) graph, data bytes are shown on x-axis whereas PSNR (Peak Signal-to-Noise Ratio) is shown on y-axis.

Figure 3 MSE (Mean Square Error) increases with data size, signal time 1 second

Figure 4 PSNR (Peak Signal-to-Noise Ratio) graph, time 1 second

Figure 5 MSE (Mean Square Error) increases with data size, signal time 2 seconds

Figure 6 PSNR (Peak Signal-to-Noise Ratio) graph, time 2 seconds
Figure 7 MSE (Mean Square Error) increases with data size, signal time 3 seconds

Figure 8 PSNR (Peak Signal-to-Noise Ratio) graph, time 3 seconds

Now we show digital representation of speech signal for minimum and maximum value of text data and see the physical changes in speech signal after embedding text message into it. More is the size of fused text data more is the noise like signal appearing in recorded speech. We take these digital representations of stego or cover speech for two values of data size i.e. 84 bytes and 5819 bytes for two values of time i.e. 1 second and 3 second respectively to show difference between them with change in size of secret information and time duration of recorded speech signal used as cover media.

Figure 9 Recorded and encoded signals with data size 84 bytes at time 1 second

Figure 10 Recorded and encoded signals with data size 5819 bytes at time 1 second
Figure 11 Recorded and encoded speech signals of data size 84 bytes at time 3 seconds

Figure 12 Recorded and encoded speech signals of data size 5819 bytes at time 3 seconds
VI. CONCLUSION AND FUTURE SCOPE

The main objective of this paper i.e. secure transmission of text data in the environment where more and more data is being transmitted and shared via internet and data is more prone to security attacks, is achieved which gives very good results for secure data transmission. Text data is fused into a real time speech signal which acted as cover signal for embedded text hence text signal is not transparent to unintended users and it appears to them like noise in speech signal. To maintain high security hash function is applied to text prior to its embedding into speech signal and receiver is supposed to extract signal with predetermined hash key. Program is designed and run in ‘MATLAB’ for recording speech signal for various time durations and it is noticed that more log is time duration of recorded speech signal, more it can accommodate text data in non areas of interest which further makes easy reconstruction of signal due to higher value of PSNR (Peak Signal-to-Noise Ratio). But for larger data volume at same speech duration, PSNR decreases.

The main feature of this technique is that it is economic, less time consuming and developed in real time. Also program is capable of being modified because of its flexibility if need arises. Hash function complexity can be increased for higher security purpose. It can be implemented practically for any specific application. We can make use of it in ‘Artificial Intelligence’ by optimizing text data limits with the help of ‘Particle Swarm Optimization’ and ‘Ant Colony Optimization’.

REFERENCES