An information hiding algorithm based on voice

Yanling Zhang

*School of Electronic Information Engineering,
Xi'an Technological University, Xi'an, 710021, China*

Zhengqiang Yang

*Student Affairs Office, Xi'an Technological University,
Xi'an, 710021, China*

Defa Hu*

*School of Computer and Information Engineering, Hunan University of Commerce,
Changsha 410205, Hunan, China*

Yunfeng Wang

*Xi 'an Xu Tai Electrical Technology co.,
LTD, Shaanxi Xi 'an 710000, China*

**Corresponding author (E-mail: hdf666@163.com)*

Abstract

Based on the characteristics of speech signal, this paper develops a new digital information hiding algorithm. The basic idea: according to the characteristics of human ear's insensitiveness to speech signal absolute phases, the paper makes full use of the short-time energy of speech signal to find the silence sections of the signal, then makes the absolute silence sections, and takes advantage of the absolute silence sections to create a number of "independent sections". By changing the absolute phases of the "independent sections", all sampling points of the "independent sections" are inverted, the positive proportion of sampling points of these sections are changed accordingly, and thus those secret digital pieces of information are carried. The experiments show that the proposed algorithm has excellent properties in robustness and invisibility.
Key words: ABSOLUTE PHASE, SILENCE SECTION, INFORMATION HIDING

1. Introduction

The rapid development of technology of Internet and information science, not only provides a great convenience for the digital information transmission and processing, but also brings great challenges to the digital information security problem [1, 2]. Therefore, more attention has been paid to the digital information security research. Digital information hiding technology means that the secret digital information is embedded in the carrier digital information, and then transmits secret digital information through the transmission of public information, meanwhile the carrier information containing secret data can be access and use normally. The purpose of digital information hiding technology is to prevent the hidden information interceptor's attention and doubt, thereby the chances of being attacked are reduced [3, 4].

Because human's sense of hearing is more sensitive than vision, the information hiding in the sound is more difficult than the information hiding in the image [5]. Since the audio carrier hiding technology has been proposed for many years, especially embedded in the frequency domain, the method of embedding technology has gained great improvement, the hiding technology has achieved a certain degree of practical use, but there are a lot of inadequacies in some hiding technology put forward, mainly reflected in hiding imperception decline with the strength of robustness of the hiding technology, or for strong anti-attack capability by increasing the embedding energy [6,7].

The paper makes a deeply research on the auditory properties of human ears, analyzes all kinds of audio recordings, sums up the digital voice signal having more silence segment, and human ear insensitive to the absolute phase of the speech signal, puts forward a kind of digital information hiding algorithm. The algorithm reduces the noise at the boundary from ordinary segmental introduction, has better imperceptibility, and very strong robustness. Meanwhile, the speech signal carries the information of specific auditory discernible, the user hides his accounts, passwords and other personal digital information in his own voice to transmit, which is more trustworthy than images, digital or sequences anyway.

2. Preprocessing of carrier speech information

The feeling of different acoustic absolute phase is virtually identical to the human ear auditory organs [8]. The ratio of the positive number of sample points and total sample points has a certain proportion in each section of the speech signal, defining as a positive proportion of each section of the speech signal. If the absolute phase of a speech signal is changed, that is all of the amplitude of the sampling points in a speech signal are inversion, then the positive proportion of the speech signal will be changed, but the human ear cannot perceive the change of the speech signal after the inversion. According to the characteristics of human auditory organ, secret digital information can be carried by inversion of a speech signal sampling points.

Because there is more than one the digital information to be hided, several "independent sections" must be constructed in the speech signal. The digital information is embedded bit by bit through piecewise inversion of the sample point (that is, change the absolute phase). So before embedding the secret information, the carrier speech signal must be preprocessed.

The main carrier speech signal preprocessing is that the carrier speech signal is divided into a number of "independent sections". Firstly the carrier speech signal is segmented, then the silence sections of speech signal are detected and identified, and all the sampling point value is set to 0 in the silence sections. The carrier speech signal between the silence sections is a speech signal relatively "independent".

2.1. The segmentation of carrier speech signal

The carrier speech signal is expressed as:

$$Z = \{z(i) \ \ 1 \le i \le L\} \qquad (1)$$

In the formula: Z is the carrier of speech signal; L is the length of the carrier speech signal. The carrier speech signal is divided into short length of equal length. Taking into account the temporal masking phenomenon of human ear, the length of each segment is selected less than 20ms. If the length is elected as M, the segmented carrier speech signal can be expressed as:

$$S = \{s(i, j) \ \ 1 \le i \le K, 1 \le j \le M\} \qquad (2)$$

In the formula: S is the segmented carrier speech signal; K is the number of segment; M is the number of sampling of each segment.

2.2. Silence sections detection

The short-time zero rate, short-time energy difference, linear prediction parameters and short-time cepstrum difference are normally used to silence detection[9,10] , but the zero rate, spectrum parameters is difficult to determine between sounds and background noise, by weighing the compatibility effect and good judgment, this paper chooses short-time energy as the test parameters. The definition of short-time energy is expressed as:

$$E_i = \left\{ \sum_{j=1}^{M} \left[s(i,j) w(j) \right]^2 \quad 1 \le i \le K, 1 \le j \le M \right\} \tag{3}$$

In the formula: $w(j)$ is a window function.

In reality, the window function mainly selects rectangle window and Hamming window [11]. Because the rectangular window has the minimum main lobe width and the high spectral resolution, rectangular window is used in the short-term energy calculation [12].

$$w(n) = \begin{cases} 1, & 0 \le n < N \\ 0, & other \end{cases} \tag{4}$$

Short-time energy can be simplified as:

$$E_i = \left\{ \sum_{j=1}^{M} s^2(i,j) \quad 1 \le i \le K, 1 \le j \le M \right\} \tag{5}$$

Defines the time-domain energy threshold value μ as:

$$\mu = \left\{ 10 \times \min\left(E_i\right) \quad 1 \le i \le K \right\} \tag{6}$$

Short-time energy of each carrier speech signal is calculated, when the short-time energy is lower than the silence energy threshold, the silence section is detected [13]. That is: if

$$E_m < \mu \quad m \in \{1,...,K\} \tag{7}$$

$$E_{m+n} < \mu \quad m+n \in \{1,...,K\}, n > 0 \tag{8}$$

$$E_{m'} > \mu \quad m < m' < m+n \tag{9}$$

Section m and section $m+n$ are the detected silence sections, set 0 method is used to mark the silence sections. Namely, the sampling values of the silence sections of the carrier speech signal are set to 0, and the short-time energy of the corresponding section is set to 0 too.

$$s(m,j) = 0 \quad 1 \le m \le K, 1 \le j \le M \tag{10}$$

$$e(m) = 0 \quad 1 \le m \le K \tag{11}$$

$$s(m+n, j) = 0 \quad 1 \le m+n \le K, 1 \le j \le M, n > 0 \tag{12}$$

$$e(m+n) = 0 \quad 1 \le m+n \le K, n > 0 \tag{13}$$

In this way, the carrier speech signal between the silence sections is artificially created "independent section". Namely the carrier speech signal between the $m+1$ section and the $m+n-1$ section is an "independent section".

2.3. The preprocessing of digital information

Digital information, including individual accounts, passwords, and other digital information, is the decimal digits or letters. In order to better hide the digital information, it must be preprocessed before embedding.

The preprocessing of digital information is mainly to convert the stream of non-binary digital information to binary code. Assume one's password is a decimal number: 065136, each decimal number is converted into four binary digits: The information flow of the preprocessed secret digital information is:0000011001010001 00110110.

3. Digital information embedding

The realization of embedding of digital information is mainly based on full inversion of sampling points in each of the "independent sections" and human ear insensitiveness to speech signal absolute phases. The specific steps are as follows:

Step1: Selecting the initial embedding section;

Step2:Starting from the corresponding section of short-time energy and find back the first zero value, and continuing to find the second zero value, then the carrier speech signal between the first zero value section and the second zero value forms an “independent section”;

Step3:The number of positive sampling points of the “independent section” is obtained, then divides by the total of sampling points of the “independent section”, thus the positive proportion value of the section is gotten;

Step4: When the positive proportion value of the section is 0, the sampling points of this section is totally negative, part of the negative, the other part is 0, or totally 0; When the positive proportion value of the section is 1, the sampling points of this section are all positive. When the positive proportion value of the section is 0.5, the positive sampling points of this section are as much as the negative; All three of these conditions after inversion will lead to the misjudgment of the secret digital information, so when the positive proportion value is 0, 1, or 0.5, the "independent section" is invalid, return to Step2; when the positive proportion value is not 0, 1, or 0.5, go to Step5;

Step5:When the positive proportional value is greater than 0.5 and the secret digital information is 0 or the proportional value is less than 0.5 and the secret digital information is 1, all the sample points of the "independent section" are inverted; Conversely, when the positive proportional value is less than 0.5 and the secret digital information is 0 or the positive proportional value is greater than 0.5 and the secret digital information is 1, all the sample points of the "independent section" remain unchanged;

Step6: If all binary digits are embedded, then the program ends, otherwise returns to Step2.

Through the above six steps, the carrier speech signal has carried secret digital information, becoming a mixed speech signal, without being perceived by the human ears.

4. The extraction of secret digital information

The secret digital information is embed in the carrier speech signal by using this algorithm, before extracting the secret digital information, the mixed speech signal is preprocessed too. Namely, the mixed speech signal is segmented, and the short-time energy value of each segment is calculated, the silent section is detected and then some "independent sections" are obtained. The length of each section and the silence energy threshold are consistent with the embedding.

The steps of extracting the secret digital information are basically the same as the embedding steps, except Step5:

When the positive proportion value of the section is greater than 0.5, then the extractive secret digital information is 1; when the positive proportion value of the section is less than 0.5, then the extractive secret digital information is 0.

Finally, the extractive binary digital information is converted into original information form. Namely, the secure transmission of digital information has been achieved.

Obviously, the secret digital information is embedded in the carrier speech signal by using this algorithm, its extraction method is very simple, and the secret digital information can be extracted without the original carrier speech signal, the blind extraction of the secret digital information can be realized.

5. Experimental results and analysis

5.1 The basic experiment

This experiment uses a single channel speech signal as the carrier speech signal, and its length is 1.78s, sampling frequency is 8 kHz, the resolution is 16 bit, digital information is 065136, its segment length is 8, the initial embedding section is 400. Figure1 - Figure 4 are the results of the experiment.

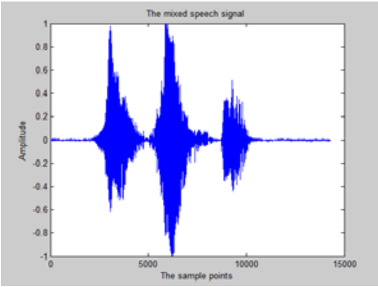

Figure 1. The carrier speech signal

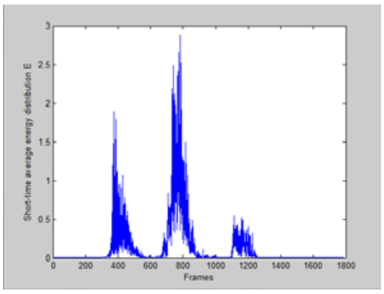

Figure 2. Short-time average energy distribution

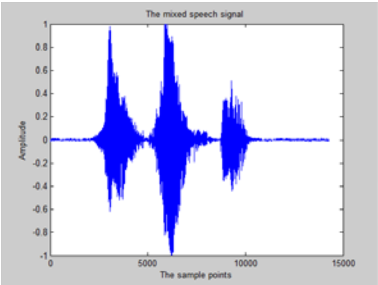

Figure 3. The mixed speech signal

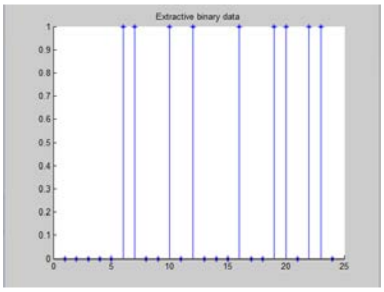

Figure 4. Extractive binary data

From the graph, we can see that the algorithm can recover the secret digital information well. Through the audition, the mixed speech signal and the original carrier speech signal have almost no difference, The algorithm reduces the noise at the boundary from ordinary segmental introduction, as the result, the requirement of the transparency of the hidden information is reached.

5.2 Robustness experiments

In this paper, the evaluation index of the design algorithm is error rate P_e :

$$P_e = \frac{T_e}{T_a} \tag{14}$$

In the formula: T_e represents the number of digital information that transmitting error, and T_a represents a total number of digital information.

In order to test the robustness of the algorithm, the attack strategy is the filtering attack, the noise attack, the sample attack, and the quantitative attack [14, 15].

1) Filtering attack. Making the mixed voice through the 2.5 kHz low-pass filter.

2) Noise attack. White gaussian noise is added to the voice mixture, its central value is 0, and its variance is 0.02.

3) Sample attack. The mixed speech is sampled down 2 times and then sampled to the original signal frequency.

4) Quantitative attack. The mixed speech signal is first quantized into 8 bit and then quantized into 16 bit.

Table 1. Error rate caused by different kinds of attack

Attack type	error rate P_e
filtering attack	0
noise attack	0
sample attack	0
quantitative attack	0

Test results show that under the various attacks, the proposed algorithm can guarantee the zero error rate, also has excellent performance in robustness, resistant to a variety of common speech signal attack means.

6.Conclusions

In this paper, according to characteristics of the human ear's insensitiveness to speech signal absolute phase, the digital information hiding algorithm based on speech signal is designed. The algorithm makes full use of the silent section of the speech signal and artificially creates several "independent sections". The secret digital information skillfully is embedded in the "independent sections" of speech signal, as the result, the application of digital information embedded in the speech signal is realized. Experimental results show that the proposed algorithm has excellent performance in robustness and invisibility. The proposed algorithm can be applied to the user's personal secret digital information transmitted in his voice. The research of proposed algorithm focuses on the robustness and invisibility, but the capacity of the embedded data is not enough. To improve the embedding capacity of the algorithm, we need to carry out more in-depth researches.

Acknowledgements

This work was supported by Science and Technology Planning Project of Shaanxi Province, China (Grant No: 2014K05-19), Science and Technology Planning Project of Xi'an, China (Grant No: CXY1442(1)), Scientific Research Program Funded by Shaanxi Provincial Education Department, China (Grant No:14JK1355), and National Natural Science Foundation of China (Grant No: 61202464).

References

1. Xie Bin, Ren Keqiang, Zhongwen Tao. (2009) An auditory masking audio watermarking based operator Method. *Communications technology*, 42(3), p.p.183-185.
2. Xi Ji, Liangrui Yu, Zhao, etal. (2013) Robust voice hidden data extraction algorithm. *Southeast University (Natural Science)*, 43(3), p.p.463-466.
3. Qiu Zhihong. (2011) Based on the quantization index modulation watermarking robust design. *Guangxi University for Nationalities: Natural Sciences*, 17(3), p.p.33-36.
4. Chen Baoyuan, Zhu Yiqiang, Tian Leilei, etal. (2013) Robust audio blind watermarking algorithm based on hear transform. *Journal of Harbin Institute of Technology*, 20(3), p.p.104-108.
5. Tang Guangming, Wang Yadi. (2008) Safety of information hiding[J]. *Computer Engineering*, 34(16), p.p.183-185
6. Yukawa M, de Lamare R C, Yamada I. (2009) Robust re-duced-rank adaptive algorithm based on parallel subgra-dient projection and Krylov subspace. *IEEE Trans-actions on Signal Processing*, 57(12), p.p.4660 -4674.
7. Liu Xin, Niu Shaozhang. (2005) Review of Steganalysis. *Journal of Beijing Electronic Science and Technology Institute*, 13(04), p.p.90-94.
8. Hanshan Li. (2015) Limited magnitude calculation method and optics detection performance in a photoelectric tracking system. *Applied Optics*, 54(7), p.p.1612-1617.
9. Hanshan Li, Yao Li. (2015) Multiburst Location Optical Testing Method and Target Damage Efficiency Assessment Arithmetic. *IEEE sensors journal*, 15(1), p.p.240-247.
10. Bender W, Butera W. (2000) Applications for Data Hiding. *IBM System Journal*, 39(3), p.p.547-568.
11. Tong Ming, Bian Zhengzhong. (2003) Study on Phase Perception in Speech. *Journal of Electronics*, 20(5), p.p.387-392.
12. F.Petitcolas. (2000) Watermarking scheme sevaluation. *IEEE Signal Proeessing Magazine*, 17(5), p.p.58- 64.
13. Shijun Xiang, Jiwu Huang. (2006) Analysis of D/A and A/D conversions in quantization-based audio watermarking.

Intemational Joumal of Network Security, 3(3), p.p.230-238.

14. Darko Kirovski, HenriqueMalva. (2003) Robust Covert Communication over a Publie Audio Channel Using Spread Spectrum. *IEEE Journal on Select Ares in Communications*, 16(4), p.p.561-572.
15. Tong Ming, Ji Hong-bing, Liu Xiao-jun. (2008) Audio Information Hiding Method Based on Phase Coding. *Computer Engineering*, 34(9), p.p.7-9.