

UDK 004.415.24

DOI:10.18413/2518-1092-2016-1-4-21-24

Lykholob P.G.
Medvedeva A.A.
Likhogodina E.C.
Mishina O.O.

**RESEARCH OF SENSITIVITY OF SOME MEASURES OF QUALITY
ASSESSMENT OF HIDDEN INFORMATION IN THE AUDIO CONTENT**

Belgorod State National Research University, 85 Pobedy St., Belgorod, 308015, Russia
e-mail: likholob@bsu.edu.ru, medvedeva_aa@bsu.edu.ru, elza9313@gmail.ru, olgamishina96@gmail.com

Abstract

The paper presents a comparison of some measures of difference between the original signal and the result of the introduction of additional information. The comparison was based on the analysis of the implementation of the results-based steganographic method of spectrum spreading. The paper presents the results of the comparison of some measures of difference based on the analysis of speech signals in their division into segments of equal length.

Keywords: speech signals; steganography; measures of differences; correlation coefficient; mean square error; signal-to-noise ratio; Itakura-Saito measure of distance.

УДК 004.415.24

Лихолоб П.Г.
Медведева А.А.
Лихогодина Е.С.
Мишина О.О.

**ИССЛЕДОВАНИЕ ЧУВСТВИТЕЛЬНОСТИ НЕКОТОРЫХ МЕР КАЧЕСТВА
СКРЫТИЯ ИНФОРМАЦИИ В АУДИОДАНЫХ**

Белгородский государственный национальный исследовательский университет, ул. Победы д.85,
г. Белгород, 308015, Россия
e-mail: likholob@bsu.edu.ru, medvedeva_aa@bsu.edu.ru, elza9313@gmail.ru, olgamishina96@gmail.com

Аннотация

В статье представлено сравнение чувствительности некоторых мер различия между исходным сигналом и сигналом, полученным в результате добавления дополнительной информации. Сравнение основано на анализе результата реализации стеганографического метода расширения спектра. В данной статье рассмотрены результаты сравнения некоторых мер различия, основанные на анализе речевых сигналов при их разделении на отрезки равной длины.

Ключевые слова: речевые сигналы; стеганография; меры различия; коэффициент корреляции; среднее квадратическое отклонение; отношение сигнал/шум; мера расстояния Итакуры-Сайто.

The development of modern information and telecommunication systems is aimed at ensuring the possibility of providing natural human forms of information exchange. One of these forms, the most commonly used, which is convenient for a person, is speech. Modern information systems allow the storage and transmission of voice messages at a distance. The provision of such opportunity led to the rapid development of technology, to ensure the implementation in the audio records of additional information that will not be perceived by human senses. This can be a label date and time, label, confirming the copyright, etc. The introduction of

additional information in such a way that the fact of implementation was discovered, is used in steganography. This aspect describes the basic principle of steganography [2].

In the case of the use as an object, which will be implemented information (container), the speech signal, the result of the implementation, i.e. the stego-container are (the container along with the embedded information), "hearing" should not differ from the original container.

Obviously, the most effective methods of change detection (identifying the degree of change) are the subjective assessment. However, the increasing

demand for stego-algorithms and, as a consequence, the increase in processed speech data leads to the need for automating the process assessment of results of introduction of additional information.

This requires the use of objective methods in some numerical form to assess the degree of difference of speech signals before and after the introduction of additional information.

In addition, for methods that evaluate the quality of the attachment has the following requirements:

- the method must allow expressing the sound quality of a quantitative measure;
- the method should consider the properties of auditory perception;
- the method should not need to use experts, but it is necessary that it should provide the best correlation with subjective evaluations.
- the method should allow to determine the critical level (detection threshold) at which changes caused by the steganographic method of encoding will be noticeable to the ear;
- the method should not depend on parameters of the analyzed signal (sample rate, bit count, etc.), it should equally respond to changes in the time and frequency domains.

Currently, the most widespread use of received methods of evaluating the differences of the compared signals, is based on the analysis of segments of speech signals in the time domain. Using such estimates of the differences as the mean square error (*MSE*), relative error, the signal-to-noise (*SNR*), the correlation coefficient (*cor*), measure the distance Itakura-Saito (distance maximum likelihood, *ISD*). Each of these assessments allows us to identify the differences in the signals. However, they have different sensitivity.

In particular, the mean square error (*MSE*) measures the absolute difference between the energy of segments signals in the time domain [7, 12, 3]:

$$MSE = \sum_{n=1}^N (x_n - \tilde{x}_n)^2, \quad (1)$$

where x_n – the amplitude of the initial data segment, \tilde{x}_n – the amplitude of the segment of data containing additional information, N – the number of samples of the compared segments of the signals.

This measure allows identifying the differences in the envelopes of the amplitudes of the segments of speech signals. The fewer changes can be made when introduction additional information, the closer the value for this score to zero.

However, this estimate does not take into account the energy of the signal itself, and this means that the choice of this evaluation has a difficulty of

choosing a threshold. Therefore it is more likely to use the normalized estimate of the *MSE* to the norm of the original signal [2]:

$$MSE = \sum_{n=1}^N (x_n - \tilde{x}_n)^2 / \sum_{n=1}^N x_n^2. \quad (2)$$

The reaction of this assessment is a similar reaction of *MSE*.

Also, to consider the extent of differences between the original signal and the result of the introduction of additional information it is necessary to make use of the assessment which is sensitive to the time alignment of the compared segments of the signals [7, 12, 3]:

$$SNR = 10 \cdot \lg \frac{\sum_{n=1}^N x_n^2}{\sum_{n=1}^N (x_n - \tilde{x}_n)^2}. \quad (3)$$

The higher the *SNR* rating, the less changes were made. In case of equality of two segments (source and exposed to changes in coding) the evaluation will be equal to infinity (∞).

To assess the degree of similarity of two segments of data, they often use the mutual energy of these signals, determined by the correlation coefficient [7, 3]:

$$cor = \frac{\sum_{n=1}^N \left(x_n - \frac{1}{N} \sum_{n=1}^N x_n \right) \cdot \left(\tilde{x}_n - \frac{1}{N} \sum_{n=1}^N \tilde{x}_n \right)}{\sqrt{\sum_{n=1}^N \left(x_n - \frac{1}{N} \sum_{n=1}^N x_n \right)^2 \cdot \sum_{n=1}^N \left(\tilde{x}_n - \frac{1}{N} \sum_{n=1}^N \tilde{x}_n \right)^2}}. \quad (4)$$

The closer the correlation value to one, the higher the similarity of the segment of data containing the control information and the source.

All the above estimates calculate the extent of the differences used for comparison the values of samples in the time domain. However, along with changes in the time domain it is also necessary to account for differences in the frequency domain. To do this, we use a measure based on the distance Itakura-Saito [7, 12, 3]:

$$ISD = \int_{-\pi}^{\pi} \left(\frac{|X(v)|^2}{|X(v)|^2} + \ln \frac{|X(v)|^2}{|X(v)|^2} - 1 \right) dv / 2\pi. \quad (5)$$

It is known that the energy of the segment of the signal can be expressed as follows [2,11]:

$$\|x\|^2 = \sum_{n=1}^N x_n^2 = \sum_{r=1}^R \left(\int_{v \in \Omega} |X(v)|^2 dv / 2\pi \right) = \sum_{r=1}^R P_r, \quad (6)$$

where P_r – the value of the energy of the frequency components of the segment signal.

Then measure based on the distance Itakura-Saito can be represented as:

$$ISJD = \sum_{r=1}^R \Delta\omega_r \cdot \left(\frac{\tilde{P}_r}{P_r} + \ln \frac{P_r}{\tilde{P}_r} - 1 \right) / \pi, \quad (7)$$

where \tilde{P}_r – the value of the energy of the frequency components of the initial data segment, P_r – the value of the energy of the frequency components of a segment of data that contains additional information.

Measure is a sense of distance between spectra of the two signals, and estimates the discrepancy between the energy changed and the source of the segment data. In case of equality of the segments of data, the measure becomes zero.

The comparison of the sensitivity estimates was based on the use of one of the most common steganographic methods [12], taking into account the frequency characteristics of the voice signal – the spread-spectrum method.

The method involves adding to the segment of the original speech a signal pseudorandom sequence (SRP) in accordance with expression [4, 8]:

$$\tilde{x} = x + \alpha_m \cdot e_m \cdot \bar{u}, \quad (8)$$

where \tilde{x} – the original segment of the data, \bar{u} – interval corresponding pseudo-random sequence, α_m – the weighting factor, e_m – a code mapping binary bits of the control information determined by the equation:

$$e_m = 2e_m - 1, \quad m = 1, \dots, M, \quad (9)$$

where e_m – bits of the control information in the binary system, $e_m \in \{0, 1\}$, M – the amount of secretly encoded control information, e_m – a code mapping binary bits of the control information, $e_m \in \{-1, 1\}$, m – the sequence number of bits of control information.

Weight coefficient α_m determines the secrecy of the system. In [10] it is proposed to choose is equal to:

$$\alpha_m = \frac{\langle x, \bar{u} \rangle}{\|\bar{u}\|^2}. \quad (10)$$

It should be noted that the use of non-mutual energy with the data \tilde{x} as a noise signal design \bar{u} allows to increase the noise immunity steganographically encoded control information e_m , and the use of projection ratio α_m increases the stealth of the control information.

Decoding bits of the control information from data is performed by determining the sign of the scalar product of the segment data and the pseudo-random sequence:

$$z_m = \text{sign}(\langle \tilde{x}, \bar{u} \rangle), \quad (11)$$

where $\text{sign}(\)$ – the allocation operation of the sign.

Table 1 presents the results of the evaluation of the considered measures of differences for all sounds of Russian speech. For the analysis there was used the segments of speech signals recorded with a sampling frequency of 8 kHz and bit depth 16 bit. To implement spread-spectrum speech signals were divided into segments of the same duration, $T=32\text{ms}$. It is also important to note that the study of these measures were carried out during the implementation of the overlay of noise on the signal in the absence of cross-correlation and using a weight:

$$\alpha_m = \frac{K_m}{\|\bar{u}\|^2}. \quad (12)$$

The parameter K_m was varied in the range from 0.0001 to of 0.2000.

From the above data, it is seen that the values of all the measures, except for measures based on distance Itakura-Saito, depend only on the coefficient K_m . In turn, the value of a measure based on distance Itakura-Saito depends on the coefficient K_m and the type of sound. So for the voiced sounds of Russian speech the addition of broadband noise causes more significant increasing measures, based on distance Itakura-Saito, than when adding the same fragment of the noise to hissing sounds. Thus, the measure based on distance Itakura-Saito takes into account the features of the energy distribution of the Russian speech sounds.

As shown by research, for evaluating speech quality it is necessary to use measures that take into account the distribution of the speech signal in the frequency band.

This is due to the perception of the speech signal by person, regardless of the language of communication. The methods that use psychoacoustics model [1] and the methods of prediction do not always provide ease of playing, because it have many settings [1, 2]. Thus, the use of measures based on distance Itakura-Saito, it is advisable to evaluate measures of the quality of hiding information in speech signals.

Table

Evaluation of differences of the original signal and implementation results using steganographic technique spreading ($T=32ms$)

Type of sound	K_m	SD	NSD	SNR	cor	ISD
А	0,0001	0,0001	0,0001	80,0000	1,0000	0,0021
	0,0002	0,0002	0,0002	73,9794	0,9999	0,0045
	0,0100	0,0100	0,0100	40,0000	0,9950	0,4529
	0,1000	0,1000	0,1000	20,0000	0,9524	6,3492
	0,2000	0,2000	0,2000	13,9794	0,9091	13,3037
Ч	0,0001	0,0001	0,0001	80,0000	1,0000	0,0002
	0,0002	0,0002	0,0002	73,9794	0,9999	0,0005
	0,0100	0,0100	0,0100	40,0000	0,9950	0,0182
	0,1000	0,1000	0,1000	20,0000	0,9524	0,3009
	0,2000	0,2000	0,2000	13,9794	0,9091	0,8142
Ш	0,0001	0,0001	0,0001	80,0000	1,0000	0,0007
	0,0002	0,0002	0,0002	73,9794	0,9999	0,0014
	0,0100	0,0100	0,0100	40,0000	0,9950	0,0523
	0,1000	0,1000	0,1000	20,0000	0,9524	0,6402
	0,2000	0,2000	0,2000	13,9794	0,9091	1,5429

References

1. Iser B., Schmidt G., Minker W. Bandwidth extension of speech signals. NY: Springer Science & Business Media, 2008. 190 p.

2. Zhilyakov E. G. Optimal sub-band methods for analysis and synthesis of finite-duration signals // Automation and Remote Control. 2015. P. 76, № 4. Pp. 589-602.

3. Fridrich, J. Steganography in digital media: Principles, algorithms, and applications, Steganography in Digital Media. 2012. Pp. 1-441.

4. Furui, Sadaoki. Digital speech processing, synthesis, and recognition. 2nd ed., rev. and expanded. New-York, USA: Marcel Dekker inc, 2000. 477 p.

5. Nedeljko Cvejic, Tapio Seppanen. Spread spectrum audio watermarking using frequency hopping and attack characterization// Signal Processing, 2004. №84. Pp. 207-213.

6. Steganalysis of audio based on audio quality metrics /Ozer H., Avcibas, I., Sankur, B., Memon, N.D.// The International Society for Optical Engineering 5020. 2003. Pp. 55-66.

7. Stanković, S., Orović, I., Sejdić, E. Multimedia signals and systems. Springer, 2012. 373 p.

8. Thierry Dutoit, Ferran Marques. 2009. Applied Signal Processing. A MATLAB TM-Based Proof of Concept. Springer, 2009. 456 p.

9. Vercoe B.L. Csound: A Manual for the Audio-Processing System, MIT Media Lab, Cambridge, 1995.

10. Zhilyakov E.G. Optimal subband methods of analysis and synthesis of signals of finite duration / Automation and Remote Control. M.: Academic Scientific Publishing, Production and Publishing and Bookselling Center of the Russian Academy of Science "Publishing House" Science "№ 4, 2015. Pp. 51-66.

11. Hicsonmez S., Uzun E., Sencar H. T. Methods for identifying traces of compression in audio. Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on – IEEE, 2013. Pp. 1-6.

Lykholob Peter Georgievich, Senior Lecturer, Department of Information and Telecommunication Systems and Technologies

Medvedeva Alexandra Alexandrovna, Associate Professor, Department of Information and Telecommunication Systems and Technologies, Candidate of Engineering Sciences

Likhogodina Elizaveta Sergeevna, Student, Department of Information and Telecommunication Systems and Technologies

Mishina Olga Olegovna, Student, Department of Information and Telecommunication Systems and Technologies

Лихолоб Петр Георгиевич, старший преподаватель кафедры информационно-телекоммуникационных систем и технологий

Медведева Александра Александровна, Доцент кафедры Информационно-телекоммуникационных систем и технологий, Кандидат технических наук

Лихогодина Елизавета Сергеевна, Студент кафедры информационно-телекоммуникационных систем и технологий

Мишина Ольга Олеговна, Студент кафедры информационно-телекоммуникационных систем и технологий