Asian Journal of Information Technology 15 (12): 1949-1952, 2016
ISSN: 1682-3915

On the Steganography in Voice Data

[1]Evgeny G. Zhilyakov, [1]Sergey P. Belov, [1]Petr G. Likholob and [2]Vladimir P. Pashintsev
[1]Belgorod State University, Pobedy St. 85, 308015 Belgorod, Russia
[2]North-Caucasus Federal University, Pushkin Street, 355009 Stavropol, Russia

Abstract: The study proposed the method of control information covert introduction in voice data, taking into account the energy distribution of their frequency components on the frequency axis. The use of subband projections for covert coding instead of the pseudo-random sequence used in spectrum increase method allows to increase the secrecy with the same probability of error.

Key words: Speech signal fragment, voice data, energy distribution, subband analysis/synthesis, coding, steganography, control information, spectrum extension method, subband projection, subband projection method

INTRODUCTION

It seems natural for a man to exchange information using oral speech and the visual display of objects, phenomena or processes. The industry of informational, educational and entertainment content creation uses oral speech to perform the audio support of notes, movies and musical compositions. This leads to the increase of data streams containing speech. In this regard, there is the problem of automatic control provision concerning the use of speech and in particular the prevention of unauthorized actions with it.

Accordingly, in order to provide an automatic control of speech data the solution of a number of problems is necessary: the confirmation of received information identity; the identification of a person; recognition; the determination of speech integrity; the protection of speech from an unauthorized access; storage at which the control information is not detectable unless you know of its existence. From many points of view, it is reasonable to perform it in a covert mode when the information about covert coding process and the appropriate actions is available only to a specific group of persons. The measure of secrecy characterizes the ability of the information not to be detected in the process of information exchange.

In order to solve the abovementioned problems, it is possible to use the principle of steganography and in the cases of audio data one may use digital steganography when the content and control information are presented in a digital form (Alekseev and Alenin, 2010; Vercoe, 1996; Dutoit and Marques, 2009; Zhilyakov *et al*., 2014, 2010, 2012; Ivanov and Chugunkov, 2003; Arnold and Kanka, 1999; Moskowitz and Cooperman, 1999; Moulin and O'Sullivan, 2003). At the heart of not a very wide range of existing steganography algorithms different techniques of control information coding are used among which are the following ones: the use of the least significant bit (Alekseev and Alenin, 2010), the coding based on spectrum spread (Vercoe, 1996) and some others.

In order to solve this problem, the authors propose the method of adaptive covert control information coding which provides a high concealment at a set probability of an error. The method consists in the use of voice data energy properties, the mathematical basis of which is the use of eigenvectors for a subband matrix (Zhilyakov *et al*., 2014) as an orthogonal basis instead of the Pseudo-Random Sequence (PRS) which is widely used at present at a secretive coding of control information. Bayestehtashk *et al*. (2016) report results on the Aurora 4 Automatic Speech Recognition (ASR), task which contains utterances with wide range of background noise.

MATERIALS AND METHODS

Main part: Let $\vec{x}=(x_1,x_2,\ldots,x_n,\ldots,x_N)^T$ is the voice data segment which is a digital representation of oral speech fragment, recorded at discrete points in time on the microphone output (speech signal). It is known that the majority of Russian speech sounds the energy of frequency components is contained in a small fraction of a frequency band (Zhilyakov *et al*., 2012). This energetic property can be put into the foundation of speech perception model by a man. Note that the signal analysis and synthesis procedures in accordance with a certain frequency range partitioning (Fig. 1) into a plurality of intervals are called subband ones. The following formula is used as a main subband characteristics (Zhilyakov *et al*., 2014):

Corresponding Author: Evgeny G. Zhilyakov, Belgorod State University, Pobedy St. 85, 308015 Belgorod, Russia

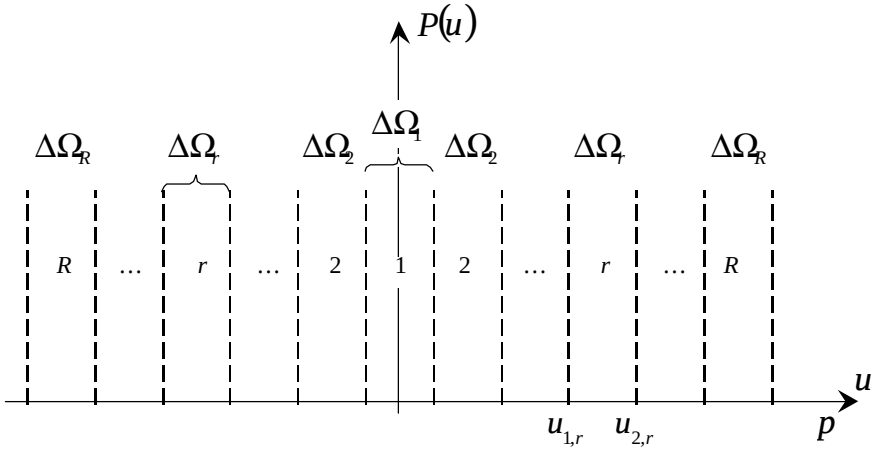

Fig. 1: Frequency band division example

$$P_r(\vec{x}) = \int_{\upsilon\in\Delta\Omega_r} \left|X(\upsilon)\right|^2 d\upsilon / 2\pi \quad (1)$$

Where:

P_r = The energy part of speech data segment attributable to a frequency band

r = Index which determines the sequence number of the frequency sub-band from possible R

$X(\upsilon)$ = Fourier transform

$$X(\upsilon) = \sum_{n=1}^{N} x_n e^{-j\cdot\upsilon\cdot(n-1)} \quad (2)$$

The j is imaginary unit ($j^2 = -1$). The substitution Eq. 1 of representation Eq. 2 provides (Zhilyakov *et al*., 2010, 2014):

$$P_r(\vec{x}) = \vec{x}^T A_r \vec{x} \quad (3)$$

where, A_r is subband matrix with the following elements:

$$A_r = \left\{a_{ik}^r\right\}\ i,k = 1,\ldots,N \quad (4)$$

$$a_{ik}^r = \frac{\sin\left(\upsilon_{2,r}(i-k)\right) - \sin\left(\upsilon_{1,r}(i-k)\right)}{\pi(i-k)},\ i \neq k \quad (5)$$

$$a_{i,k}^r = \left(\upsilon_{2,r} - \upsilon_{1,r}\right)/\pi,\ i = k \quad (6)$$

Where:

υ_{1r} = The left border of the subband $\Delta\Omega_r$

υ_{2r} = The right border of the subband

The performed studies show that for all speech sounds the overwhelming parts of energy for the segments of corresponding speech signals are concentrated in the low fraction of a frequency band determined by the half of sampling frequency. In order to place the control information it is offered to use the rest of the frequency band. In order to determine the frequency sub-band $\Delta\Omega_r$ where one may carry out covert coding a determinative rule is proposed:

$$\left|\frac{\Delta\Omega_r}{\pi}\times\left\|\vec{x}\right\|^2 - \frac{1}{2\pi}\int_{\upsilon\in\Delta\Omega_r}\left|X\ (\upsilon)\right|^2 d\upsilon\right| = \min_{\Delta\Omega_r\in[0,\,\pi]} \quad (7)$$

where, $\|\vec{x}\|^2$ the energy of speech data segment. That is the band is used which gets the energy almost equal to the average one. Covert implementation demends the use of signal segments with a high concentration of energy in narrow frequency bands. It is proposed to use the eigenvectors of a subband matrix:

$$\lambda_k^r \vec{q}_k^r = A_r \vec{q}_k^r,\ i,k = 1,\ldots,N \quad (8)$$

Eigenvalues which have the following properties:

$$\lambda_k^r = \int_{z\in\Omega_r} Q_k^r(z)|^2\ dz/2\sigma \le 1| \quad (9)$$

Where the subintegral function is determined by the spectrum of a corresponding eigenvector. Thus, an eigenvalue is equal to the energy of its own vector falling within a predetermined frequency range. Therefore, in order to implement information you should use the eigenvectors corresponding to the eigenvalues close to unity .The number of eigenvalues which are close to unity $\lambda_J \cong \ldots \cong \lambda_2 \cong \lambda_1 \approx 1$, depends on the segment length multiplied by a sub-band width Eq. 6. One may show that it is advisable to use a partition to implement the method of covert coding:

$$\Delta\Omega_r = 2\Delta\Omega_1, r = 2,3,...,R,\ \Delta\Omega_1 = \pi/\left(2(R-1)+1\right)R = (N-2)/4 \quad (10)$$

Then there is a pair of eigenvectors for a subband matrix, whose eigenvalues are close to unity. Subband eigenvectors can be made orthonormal when the following equality for scalar products is performed

$$\left\langle \vec{q}_l^r, \vec{q}_k^r \right\rangle = \begin{cases} 1, & l = k \\ 0, & l \neq k \end{cases} \quad l,k = 1,\ldots,N \quad (11)$$

This property allows to simplify the task of introduced information restoration at its encoding with the use a following subband projection:

$$\alpha_i^r = \left\langle \vec{q}_i^r, \vec{x} \right\rangle i = 1,\ldots,N \quad (12)$$

Subband projection method: Let b_m is bit value (zero or unity) in the binary representation of an introduced number. It is proposed to use the following introduction model in a speech data segment $\vec{x}$:

$$\hat{\vec{x}} = \vec{x} + \sum_{j=1}^{J} \left(\text{sign}(e_m) \times \left|\alpha_j^r\right| - \alpha_j^r\right) \times \vec{q}_j^r \quad (13)$$

$$e_m = 2b_m - 1 m \in M \quad (14)$$

Where:

Sign() = Sign revealing operation

J = The number of eigenvectors, the eigenvalues of which are close to one

The decoding of control information is carried out by projection sign determination α_i^r for the eigenvectors $\vec{q}_i^r$ of the subband matrix A_r, found for the area $r \in R_1$:

$$\hat{e}_m = \text{sign}\left(\left\langle \hat{\vec{x}}, \vec{q}_i^r \right\rangle\right) m \in M \quad (15)$$

$$\hat{b}_m = \left(\hat{e}_m + 1\right)/2 \quad (16)$$

Where:

$\hat{e}_m$ = The symbol decoded by the subband projection method

$\hat{b}_m$ = The bit decoded by subband projection method

RESULTS AND DISCUSSION

Calculation experiments: The comparative evaluation of data introduction security and sustainability on the basis of subband projection method and spectrum expansion method the computational experiments were used. At that the following data were used. The control information was generated as a binary random sequence b_m, containing 10^6 of elements with the same number of zero and one bits (Ivanov and Chugunkov, 2003). The e_m symbols were developed from obtained binary sequence Eq. 14.

Base voice segment database corresponding to the Russian language sounds was developed with the following specifications: the number of values N = 256; number of bits B = 16; sampling frequency $f_э$ = 8kГЦ; the total number speech data test pieces made Z = 6400.

Using Fibonacci generator (Ivanov and Chugunkov, 2003) ($\vec{u} \in \{0,1\}$) PSP, used in the method of spectrum extension. The sequence was divided into 10^6 of uncorrelated segments with the length of N = 256 values.

Pseudo-random white Gaussian noise was used as a noise distortion model, the energy of which is uniformly distributed in the frequency domain. The sequence is divided into 10^6 of uncorrelated with the length of N = 256 values.

In order to evaluate the control information security δ, encoded in speech data segment the following measure was used:

$$\delta = \sqrt{\frac{1}{Z}\sum_{z=1}^{Z}\left(\left\|\vec{x}\right\| - \left\|\hat{\vec{x}}\right\|\right)^2 \Big/ \left\|\vec{x}\right\|^2} \quad (17)$$

where, Z the number of analyzed speech data segments. The modeling was carried out as follows:

- According to the frequency axis division (Eq. 10) each data segment was analyzed in order to select a sub-band $\Delta\Omega_r$, satisfying the determining rule (Eq. 7)
- The covert coding in the subband Δωr, e_m symbol (Eq. 13) was carried out by subband projection method
- The energy was evaluated introduced by subband projection method at e_m symbol encoding:

$$K_m = \sqrt{(\alpha_m)^2} \quad (18)$$

- Subband projection concealment method was evaluated (Eq. 17)
- Covert coding was performed at the central frequency of the subband $\Delta\Omega_r$ of e_m symbol by spectrum expansion method (Eq. 13) with the proportionality coefficient (Eq. 18)
- The security of spectrum extension method was evaluated (Eq. 17)

Table 1: Error probability values P_{out}

Characteristic	Error probability P_{out}	
Noise/signal correlation h^{02}	Spectrum expansion method	Subband projection method
0.0010	0.370828	0.021937
0.0100	0.370966	0.067710
0.1000	0.374175	0.180317
1.0000	0.394746	0.334562

The simulation showed that the spectrum expansion method reaches 7.1×10^{-3} and for the method of subband projection this parameter does not exceed $2.69\ 10^{-16}$. In other words when you use the method of subband projections the changes in speech data energy almost equal to zero. The value of error probability at decoding P_{out} per bit (BER) was calculated according to the following Equation:

$$P_{out} = M_{out}/M \quad (19)$$

Where:

M_{out} = The number of erroneously received bits from the total volume of control information

M = The amount of control information

The error was evaluated resulting from the additive effect of noise in accordance with the following model:

$$\vec{y} = \hat{\vec{x}} + \sqrt{h_0^2}\vec{u} \quad (20)$$

Where:

h_0 = The correlation of noise/signal values in times

$\hat{\vec{x}}$ = Speech data segment, containing control information

$\vec{y}$ = Speech data segment, containing control information after noise effect

Numerical results of an error bit occurrence probability evaluation during control information decoding within the terms of noise exposure are presented in Table 1.

Table 1 data illustrate the advantages of the proposed covert method of information implementation into speech data compared with the spectrum expansion technique which is noted in literature as the most effective one.

Summary: The method of additional information introduction into speech data is proposed. The basic correlations were obtained. These correlations determine the procedures for a sub-band analysis of speech signal segments and the synthesis of segments with embedded data. Comparative computational experiments were performed concerning the study of concealment characteristics for introduced information and its decoding stability to the effects of distorting noises.

CONCLUSION

The proposed method of additional information introduction into speech data has the advantages of concealment and resistance to the effects of currently known noises.

REFERENCES

Alekseev, A.P. and A.A. Alenin, 2010. Hidden data transmission in WAV audio files. Inform. Commun. Technol., 8: 101-106, (In Russian).

Arnold, M. and S. Kanka, 1999. MP3 robust audio watermarking. Proceedings of the DFG VIIIDII Watermarking Workshop, October 5-6, 1999, Erlangen, Germany, pp: 1-5.

Bayestehtashk, A., I. Shafran and A. Babaeian, 2016. Robust speech recognition using multivariate copula models. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, March 20-25, 2016, Shanghai, China, pp: 5890-5894.

Dutoit, T. and F. Marques, 2009. Applied Signal Processing: A MATLAB™-Based Proof of Concept. Springer, USA., ISBN-13: 978-0387745343, Pages: 456.

Ivanov, M.A. and I.V. Chugunkov, 2003. Theory, Application and Evaluation of Pseudo-Random Sequence Generators. Publishing House Kudits-Obraz, Moscow, Pages: 240, (In Russian).

Moskowitz, S.A. and M. Cooperman, 1999. Method and system for digital watermarking. United States Patent No. 5,905,800, May 18, 1999.

Moulin, P. and J.A. O'Sullivan, 2003. Information-theoretic analysis of information hiding. IEEE Trans. Inform. Theory, 49: 563-593.

Vercoe, B.L., 1996. Csound: A Manual for the Audio-Processing System. MIT Media Lab, Cambridge.

Zhilyakov, E.G., S.N. Devitsyna and P.G. Likholob, 2012. Determination of possible volume of introduced information at a hidden transfer of marks in the speech data. Sci. Stat. Belgorod State Univ. Inform. Ser., 132: 222-227, (In Russian).

Zhilyakov, E.G., S.P. Belov and A.A. Chernomorets, 2010. Variational Methods of Signal Analysis based on Frequency Representations. Publishing House of OJSC, Moscow, pp: 10-26, (In Russian).

Zhilyakov, E.G., S.P. Belov and D.V. Ursol, 2014. About the best orthogonal basis for generation of the channel signals. Int. J. Applied Eng. Res., 9: 12121-12126.