Speech signal marking on the base of local magnitude and invariant segmentation

Rashkevych Yu., Peleshko D., Pelekh I., Izonin I.

Lviv Polytechnic National University
12 S. Bandera str., 79013, Lviv, Ukraine

(Received 23 December 2014)

The paper suggests a new watermarking scheme based on invariant method of segmentation and the use of local magnitude for marking speech signals. The watermark is embedded in the chosen form at peaks with the spectrum magnitude of each nonoverlapping frame of audio signal.
The suggested scheme of marking speech signals ensures the reliable protection against several types of attacks, i.e. noise, trimming, re-sampling, re-quantization, compression, and lower frequencies filtering.

Keywords: *speech signal, marking, watermark, primary segmentation, symmetric matrix of distances, pseudoinverse matrix, Moore-Penrose algorithm, singular decomposition, Euclidean, Chebyshev and cosine metrics, similarity measure*

2000 MSC: 68R99

UDC: 681.142.2; 622.02.658.284; 621.325

1. Introduction

Availability of access to information and its fast proliferation is a major factor in the development of new technologies during the past decade. Audio information has become more widely used through the Internet. With the increasing use of audio content on the Internet serious problems have emerged, namely forgery, fraud, and piracy. In fact, anyone can copy an audio file and use it for his/her own purposes, such as usage in presentations or for marketing campaigns.

Thus, the abuse of copyright thrives among media users, being a motivating factor in the development of new technologies to protect audio information, and the problems of preventing piratic copying information, copyright counterfeiting and infringement should be solved worldwide.

One of the speech signal protection technology is the use of digital watermarking. Digital watermarking methods are considered to be an effective solution to the problems of copyright. Watermarks methods embed secret information on the rightholder, using the masking effect of the human visual system (Human Visual System) or human audio system (Human Audio System) in order to establish the information copyright.

Information masking can be divided into three processes: cryptography, steganography, and marking. To a great extent, marking resembles steganography, but the difference is that the information is usually hidden in the very object. Moreover, the digital watermark can be used not only to protect information against unauthorized copying, but also for the identification of individual data.

In recent years, many methods of embedding watermarks have been developed to create reliable and invisible watermarks for audio information. In [14] Lee et al. suggested a method of embedding watermarks in the time domain of an audio signal, which is based on the use of differential amplitude in each group of audio samples to represent one bit of information. The basis of this technique is marking lower frequencies changes of amplitude for scaling amplitudes in chosen segments of the samples, so that time domain of baseband signal can be almost completely preserved.

In [23] there is suggested a scheme according to which the watermark is embedded in the time domain of the audio signal. Watermark signal is formed with a key, and embedding the watermark depends on the amplitude and frequency of the audio signal that minimizes the watermark sensitivity in signal.

In [24] a method for embedding watermark on the base of nonuniform discrete Fourier transform (NDFT) is developed, in which the frequency of spots of embedded watermark is determined using a secret key.

In [25] Zeng et al. presented a "blind" method of embedding watermarks, which embeds a watermark into coefficients of discrete cosine transform (DCT) using technique of modulation

Today there are two main directions of "blind" method. In the base of the first trend is the use of adaptive quantization against synchronization attacks [23]. Characteristics of the discrete wavelet transform (DWT) and the characteristics of discrete cosine transform (DCT) are combined in this method to improve the transparency of the watermark. Thus, the watermark is embedded in the lower frequencies components using adaptive quantization according to the human hearing system. Note that the use of wavelet transforms for marking is quite common today. For example, in [18] Puyang at al. presented a method of marking by which the watermark is embedded in the wave (wavelet) domain. Once the watermark being embedded, it is encrypted and connected with the code of synchronization, and integrated into the lower frequencies coefficients of the signal in the wave domain. A size of quantization step and embedding sustainability are determined adaptively according to the characteristics of human hearing system.

The second way is based on the use of cepstrum. For example, in [15] there is suggested "blind" watermarking scheme based on the use of cepstrum function resistant against attacks and the ability of VSN code error correction to improve the reliability and secrecy of audio watermarking. The original method of using cepstrum [13] should be mentioned as well. Its authors propose a system of embedding watermarks into cepstrum domain where the pseudo-random sequence is used as a watermark. The watermark is displayed in the cepstrum domain in accordance with the distribution of cepstral coefficients and the masking frequency properties of the human hearing system.

One of the best known methods of marking is the Cox method [9], in which the watermark is embedded in the highest n coefficients of discrete cosine transform (DCT) of the whole sound except the discrete cosine component as follows:

$$u_i' = u_i(1 + \alpha x_i),$$

where u_i is the amplitude value, where the watermark is embedded, x_i is the watermark embedded in u_i, α is the coefficients of scaling, and u_i' is the modified value of u_i. The sequence of watermarks is chosen by performing the reverse operation represented by the following equation:

$$x_i^* = \big((u_i^*)/u_i - 1\big)/\alpha.$$

Cox method is determined by high reliability behavior. However, this method cannot achieve a high level of secrecy in terms of signal to noise ratio, as far as the very watermark is embedded into DCT highest coefficients of the sound components, which sometimes affect the sound quality. Therefore, the task of developing methods of marking which possesses the optimum ratio of reliability and secrecy is very urgent.

2. Statement of the problem

The main purpose of this paper is to develop a speech signal invariant to certain types of attacks and modifications by the way of marking method. Resistance of the speech signal against transforms will be provided by the developed segmentation algorithm, and the very watermark should be embedded at peaks in the chosen form of the magnitude spectrum of each nonoverlapping frame of audio signal.

2.1. Segmentation of the speech signal on the base of the Moore-Penrose pseudorotation

It is known that one of the most common problems that arise in the analysis of speech signals in the development of modern artificial intelligence systems is to determine their temporal and frequency characteristics because any non-determined signal is a nonlinear object [3]. However, at a given sampling period, in such signal you can always choose a time sample in which the values of these characteristics vary within insignificant deviation range. Then the characteristics of the signal in this range are considered to be permanent and the procedure of obtaining the very sample is considered to be the segmentation of the speech signal [2,4].

If in the speech signal to determine qualitative characteristics, then in the sample of segmentation one can construct a quite accurate parametric model of the speech signal, which, in its turn, can be successfully used in solving a wide range of problems, including problems of marking.

Therefore, the first task in developing a method of marking is to develop a method for automatic segmentation of the speech signal. For this purpose, the use of pseudorotations of a symmetric matrix of distances and the evaluation criteria for automatic determining the threshold segmentation values are recommended.

Let the compact $\tau = [0;T]$ is given in the domain $\mathbb{R}^1$, $T \in \mathbb{R}^1$, $T > 0$, which is to be the speech signal domain $x(t)$, $t \in \tau$. Then the speech signal can be considered to be a continuous surjective mapping

$$x\colon \tau \to \mathbb{R}^1. \tag{1}$$

In the sample τ, the system of open sets T_i is defined by the rule

$$T_i = [t_{i-1};\ t_i] : \forall t_0 \in T_i,\ \exists r > 0\colon T_i^r\,(t_0) \subset T_i,\ T_i^r\,(t_0) = \{t \in T\colon \rho(t,t_0) = \|t - t_0\| < r\}\,, \tag{2}$$

with a diameter $\rho\,(T_i) = \sup\limits_{t_a,t_b \in T_i} \rho(t_a,t_b)$ (here r is the metric of the domain $\mathbb{R}^1$) can be defined by the topology $\Gamma = \{T_i\}_{i=1,2,\ldots}$. Since the sample τ is a compact (closed bounded set), then in the topology G it is always possible to distinguish a disjunctive (with trivial intersection) finite set $\chi = \{T_i | i = 1 \ldots n\}$

$$\forall i, j \in [1;n]\colon i \neq j,\ T_i, T_j \in \chi,\ T_i \neq \varnothing, T_j \neq \varnothing \to T_i \cap T_j = \varnothing, \tag{3}$$

which is a covering τ of the compact

$$\tau = \bigcup_{i=1}^{n} T_i,\ T_i \in \chi,\ n = |\chi|\,. \tag{4}$$

where n is the power of a set χ.

Further, we take the assumption that power of all elements $|T_i|$ of the covering χ is equal to r, i.e.: $\forall i \in [1,n] :\ |T_i| = l$. This determines χ as l-covering of the sample τ.

Continuous mapping $x(t)$ with covering χ generates the covering η (not necessarily disjunctive) of the domain $X \subseteq \mathbb{R}^1$ of the function $x(t)$

$$\eta = \{X_i | i = 1 \ldots n\}, \quad x\,(t) = \bigcup_{i=1}^{n} X_i,\ X_i \in \eta, \tag{5}$$

where $X_i = \{x(t) | t \in T_i\}$ is the element of covering η. The power of covering h due to the continuity $x(t)$ is equal to the power of covering χ: $|\eta| = |\chi| = n$. It is obvious that the power of the element X_i is equal to the power of the element T_i.

In practice, the compact τ is considered to be discrete and finite. Elements T_i of the covering χ of such compact will be also discrete, finite, and of similar power. This defines the mutual homeomorphism of elements T_i. Then the power of compact τ with disjunctive covering χ of the topology Γ will be defined by the Grassmann formula

$$T+1=|\tau|=\Big|\bigcup_{i=1}^{n}T_i\Big|=\sum_{i=1}^{n}|T_i|-\Big|\bigcap_{i=1}^{n}T_i\Big|=nl. \tag{6}$$

As a result, *the problem of speech signal $x(t)$ segmentation at the discrete compact τ with defined disjunctive covering χ can be formulated as a synthesis of a new representation of the signal $x(t)$ at the compact τ*

$$x(t)=\bigcup_{i=1}^{m}Y_i,\ m\leqslant n. \tag{7}$$

where m is a number of quasi-stationary sections Y_i. Quasi-stationary area (the quasi-statsionar) Y_i is some union of consistent elements of covering η

$$Y_i=\bigcup_{j=I_i}^{m_i+I_i-1}X_j, \tag{8}$$

here I_i is the initial index of conjunction (8) in the covering η, $I_1=1$; m_i is the number of elementary sections of X_i in the conjunction (8). It is obvious that $n=\bigcup\limits_{j=1}^{m}m_i$. Following (6), the power of quasi-stationary section Y_i is determined as follows: $|Y_i|=m_il$ and is dependent on the characteristics of the covering η.

The set of indices $\{I_i\}_{i=1\dots m}$ defines a new covering $\eta'=\{Y_i|i=1\dots m\}$, which, in turn, induces new covering χ' of the compact τ as a modification of the corresponding covering χ. Then the problem of construction of the quasi-stationary sections lies *in determining parameters m_i and I_i for given initial coverings χ and η*. In general, this problem can be considered to be a construction of the transform operator of coverings $f\colon\chi\to\chi'$ or $F\colon\eta\to\eta'$.

Initial covering χ *and* η in practical problems of speech signals processing are called *the signal segmentation with elementary sections.*

Following the initial segmentation of the speech signal by elementary sections, the next step of the procedure of segmentation is a construction of divergence aggregate matrix.

In contrast to the typical approach (e.g. [19]), in the course of the construction of the matrix operator, the speech signal is not normalized and is not moved to the positive domain. Amplitude values of the speech signal of the element X_i, which correspond to the elementary area of T_i without any additional transformations, are used to build some matrix operator $\nabla_i\colon T_i\to X_i$.

Operator $\nabla_i\colon T_i\to X_i$ of the transform l-dimensional vectors defined along the sample T_i, which corresponds to the element X_i of the covering η we construct as an aggregate symmetric matrix of distances [5], which in general is one of the varieties of divergence matrix:

$$\forall i\in[1;n]:\nabla_i=\begin{pmatrix}\delta_{i,(1,1)} & \dots & \delta_{i,(1,l)}\\ \dots & \dots & \dots\\ \delta_{i,(l,1)} & \dots & \delta_{i,(l,l)}\end{pmatrix},\quad \delta_{i,(z,k)}=|x_{i,k}-x_{i,z}|\,,z,k=1..l \tag{9}$$

where i is the index of the elementary area X_i, $_i$, $j=x(t_j)$, $t_j\in T_i$. The dimension of the matrix (10) is equal to: $\dim\nabla_i=l\times l$. In the case of Euclidean distance, by the matrix (9) we can obtain

$$\forall i\in[1;n]:\nabla_i=\begin{pmatrix}0 & \Delta\frac{dx_{1,1}}{dt} & \dots & (l-1)\,\Delta\frac{dx_{1,1}}{dt}\\ \Delta\frac{dx_{1,1}}{dt} & 0 & \dots & (l-2)\,\Delta\frac{dx_{1,2}}{dt}\\ \dots & \dots & \dots & \dots\\ (l-1)\,\Delta\frac{dx_{1,1}}{dt} & (l-2)\,\Delta\frac{dx_{1,2}}{dt} & \dots & 0\end{pmatrix}. \tag{10}$$

To solve the differentiation task of vectors which will be characteristics of the speech signal elementary area for each element X_i of the covering η, let us consider the equation

$$\nabla_i g_i = x_i, \tag{11}$$

where $x_i = \{x(t_p) \in X_i | p = 1 \dots l\}$ is l-dimensional vector of speech signal amplitude values $x(t)$ in the sample X_i; $g_i = (g_{i,1}, \dots, g_{i,l})$ is vector of unknows.

According to (11), the vector g_i wiil be defined as follows

$$g_i = \nabla_i^{-1} x_i. \tag{12}$$

Since the matrix ∇_i is confluent ($\det(\nabla_i) = 0$), then the inverse matrix ∇_i^{-1} does not exist. Therefore, to solve the problem (11) by Theorem of minimization of the discrepancy $\|x_i - \nabla_i g_i\|^2$ of the linear system (12), it is suggested the following method determining the vector y_i [1]

$$g_i = \nabla_i^+ x_i + \left(1 - \nabla_i^+ \nabla_i\right) r_i, \tag{13}$$

where ∇_i^+ is the generalized inverse matrix of Moore-Penrose (pseudoinverse to ∇_i matrix [1,17]); $\left(1 - \nabla_i^+ \nabla_i\right)$ is the operator of ∇_i projection on the core; r_i is the random l-dimensional vector. First term in (13) is the pseudo inverse solution, and the second is a solution of the homogeneous system $\nabla_i g_i = 0$. Being represented through (13), the way to determine the vector of characteristics of i-th element of the covering χ is possible because according to [1] the matrix $\nabla_i^+ \nabla_i$ is not degenerated.

The Moore-Penrose matrix ∇_i^+ is determined by singular decomposition of the matrix ∇_i in the following way

$$\nabla_i^+ = V_i \Sigma_i^+ U_i^T, \tag{14}$$

where U_i, V_i are unitary matrices of $l \times l$ order of the singular decomposition of ∂i; Σ_i^+ is the matrix of $l \times l$ order, which is the pseudo inverse to the diagonal matrix Σ_i of the singular decomposition [3] of matrix ∇i. Since the matrix Σ_i is also degenerated, then the matrix Σ_i^+ can be obtained from Σ_i in the way of replacement of all nonzero singular numbers $\sigma_{i,q}$ ($\sigma_{i,1} \geqslant \sigma_{i,2} \geqslant \dots \sigma_{i,l} \geqslant 0$) with the corresponding inverse to them $1/\sigma_{i,q}$.

In the iterative process for finding g_i^{j+1} (14), the random vector r_i^{j+1} is defined by the discrepancy: $r_i^{j+1} = \|x_i - \nabla_i g_i^j\|_l$, here $\|\cdot\|_l$ is l-norm [1].

3. Scheme of speech signal segmentation

Using vectors $g_i' = (g_{i,1}', \dots, g_{i,l}')$ obtained by (13), let us form a matrix G'

$$G' = \begin{pmatrix} g_1' \\ g_2' \\ \dots \\ g_n' \end{pmatrix} = \begin{pmatrix} g_{1,1}' & g_{1,2}' & \cdots & g_{1,l}' \\ g_{2,1}' & g_{2,2}' & \cdots & g_{2,l}' \\ \dots & \dots & \dots & \dots \\ g_{n,1}' & g_{n,2}' & \cdots & g_{n,l}' \end{pmatrix}, \tag{15}$$

which should be normed with respect to the maximum element: $G = G'/\max(G')$. As a result, we obtain the matrix $G = \langle g_i \rangle_{i=1\dots n}$ which elements are defined as: $g_i = g_i'/\max(G')$.

In the space of vectors $\{g_i\}$ we introduce a metrics for two elements X_i and X_j through Chebyshev distance of corresponding vectors g_i and g_j

$$\mu(X_i, X_j) = \max_{k \in [1;l]} \{|g_{i,k} - g_{j,k}|\}, \tag{16}$$

Using the metrics (16) we can hold a primary segmentation by determining the conditions of belonging of the elementary area X_j to the quasistationar Y_i'

$$X_j \in Y_i' \Leftrightarrow \forall z \in [a_i; b_i] : \mu\left(X_j, X_z\right) \leqslant \varepsilon, \tag{17}$$

where $\varepsilon \in \mathbb{R}^{1,+is}$ is the threshold value, which is the error of determining the elementary area X_j to the quasistationar Y_i'.

The term 'primary segmentation' is used in the context of the fact that in some cases the usage of some metrics (e.g. the correlation one) can need an additional procedure. The essence of this procedure is to merge neighboring segments Y_i' and Y_j' if the distance between their beginnings is equal to l. As a result, we receive a set of quasi-stationary sections $\{Y_i\}$. It is obvious that the need for additional procedure must be determined for each metrics individually. If they chose metrics such that additional union would not be needed $Y_i = Y_i'$, and additional procedure is proved to be a zero additive operator. Its use does not affect the results of segmentation and only slows down the speed of the calculation process.

The problem (17) has a solution for the threshold value ε given in advance However, in the case of determining the criterion K, the procedure of ε choice can be automated by solving some extreme problem.

Realization of this approach can be automated way to determine the value of ε, the core of which as the criterion K, is the chosen deviation optimization of Y segmentation results of our developed method from the results Y_{et} obtained by means of a 'standard' method.

$$K\colon \ \|Y - Y_{et}\| \to \operatorname{opt}(\varepsilon)\,. \tag{18}$$

During the practical implementation of the described method, a method for automated determining of threshold value ε is based on maximizing the value of the chosen coefficient of similarity (similarity measure) [6]. In general, instead of one coefficient we can choose a few ones and with their help to determine some integral parameter. For this purpose, from the general formula of measures of Syomkin similarity continuum [6]

$$K_{\tau,\iota}\left(Y, Y_{et}\right) = \left(\frac{K_{\tau,\iota}\left(Y|Y_{et}\right) + K_{\tau,\iota}\left(Y_{et}|Y\right)}{2}\right)^{\frac{1}{\iota}}; \quad -1 < \tau < \infty, \ -\infty < \iota < \infty. \tag{19}$$

for $\tau = 0$ (degree of proximity of neighboring objects according to the general formula of average values of Kolmogorov [6]) and $\iota = +\infty, 1, 0, 1, -\infty$, there is a chosen set of most used similarity measures arranged with respect to ι, in particular the measures of Kulczinsky $K_{0,1}$ [12], Ochiai $K_{0,0}$ [16], Sorensen $K_{0,-1}$ [21], Braun-Blanquet $K_{0,-\infty}$ [8], Szymkiewicz-Simpson $K_{0,+\infty}$ [20]. Among the elements of this set of measures, the average value can be defined

$$K_{\Sigma} = \frac{K_{0,+\infty} + K_{0,1} + K_{0,0} + K_{0,-1} + K_{0,-\infty}}{5}. \tag{20}$$

Then according to (18), to calculate the value ε the problem of finding maximum is to be solved:

$$\|Y - Y_{et}\| \to \max_{0 < \varepsilon \leqslant 1} K_{\Sigma}. \tag{21}$$

Software implementation of the proposed segmentation algorithm is approved by the example of separation of the word 'миша' (in Ukrainian). Characteristics of the speech signal are: the word length is 1.03 seconds, sampling rate is 11 025 Hz, the length of the elementary section $l = 120$ indications.

Note that in practical implementations, to determine the Chebyshev distance and cosine metrics have been used. As a result of the procedure of automatically determining the threshold value ε in the case of Chebyshev metrics, the value $\varepsilon = 0.15$ is obtained, and in the case of cosine metrics $\varepsilon = 0.01$.

As the evaluation criteria for automatic determining the threshold deviation ε, there are chosen results of the same signal segmentation by the algorithm DELCO [7] with a threshold 1.6 for the same dimension of the elementary area (numerical values of DELCO segmentation method results are shown in the table). With this, the extreme problem has been solved with respect to the integral index (20). Figure 1 shows the values of all indices (20). In addition to them, Yurtsev $K_{0,+\infty}$ [9] and Jaccard $K_{1,-1}$ [11] measures have been also calculated. They are not included into the integral index (2) because the Yurtsev measure is dual to measure of Braun-Blanquet, and Jaccard overlapping measure is equivalent to Sorensen one.

The segmentation results by this method are given in the table.

Table 1.

Interval (ms:ms)	**Segment by index**	**Count of elements**	**Interval (ms:ms)**	**Segment by index**	**Count of elements**
Segmentation By Matrix of Divergences			Segmentation by		
#Euclidean metric			Assymetric Convergence Matrix		
(120:480)	[1;4]	3	(480:1200)	[4;10]	6
(480:5760)	[4;48]	44	(1200:1680)	[10;14]	4
(5760:6000)	[48;50]	2	(1680:2160)	[14;18]	4
(6000:6240)	[50;52]	2	(2160:2760)	[18;23]	5
(6240:6840)	[52;57]	5	(2760:3000)	[23;25]	2
(6840:7920)	[57;66]	9	(3000:3600)	[25;30]	5
(7920:10800)	[66;90]	24	(3600:3960)	[30;33]	3
(10800:11160)	[90;93]	3	(3960:4680)	[33;39]	6
#Cosine metric			(4680:5160)	[39;43]	4
(240:1920)	[2;16]	14	(5160:5760)	[43;48]	5
(1920:5760)	[16;48]	32	(5760:6240)	[48;52]	4
(5760:6000)	[48;50]	2	(6240:6600)	[52;55]	3
(6000:6240)	[50;52]	2	(6600:7200)	[55;60]	5
(6240:6600)	[52;55]	3	(7200:7920)	[60;66]	6
(6600:6840)	[55;57]	2	(7920:8400)	[66;70]	4
(6840:7080)	[57;59]	2	(8400:9000)	[70;75]	5
(7080:7320)	[59;61]	2	(9000:9240)	[75;77]	2
(7320:7920)	[61;66]	5	(9240:9840)	[77;82]	5
(7920:11040)	[66;92]	26	(9840:10440)	[82;87]	5
(11040:11160)	[92;93]	1	(10440:11160)	[87;93]	6
Delco Segmentation					
(120:3000)	[1;25]	24	(6240:6840)	[52;57]	5
(3000:3600)	[25;30]	5	(6840:7920)	[57;66]	9
(3600:5760)	[30;48]	18	(7920:9000)	[66;75]	9
(5760:6240)	[48;52]	4	(9000:9840)	[75;82]	7
			(9840:10200)	[82;85]	3

Having analysed these results, it should be noted that the measures of Kulczinsky, Ochiai, Sorensen, and Braun-Blanquet approximately reflect to the same extent the similarity of segmentation results with respect to results obtained by DELCO method. With this, Chebyshev metrics ensures closer results to those obtained by the segmentation of DELCO method.

The values of all these coefficients demonstrate that the results of segmentation on the basis of the proposed method with Chebyshev metrics are the most similar to the results of DELCO. Value deviations, even when to take into account the value of Simpson measure, are sufficiently small. This suggests that, unlike the case of usage of convergence matrix method, the developed method based on Chebyshev distance is quite resistant against measures of proximity and allows us instead of the integral index (20), to calculate one of the factors of Kulczinsky, Ochiai, Sorensen, and Braun-Blanquet or Simpson.

In the case of cosine metrics, the values of similarity measures are less by the magnitude, but deviation is also small. Therefore, with the use of cosine metrics it also can be stated the stability of results of segmentation and automatic determining threshold deviation to the choice of measure of proximity.

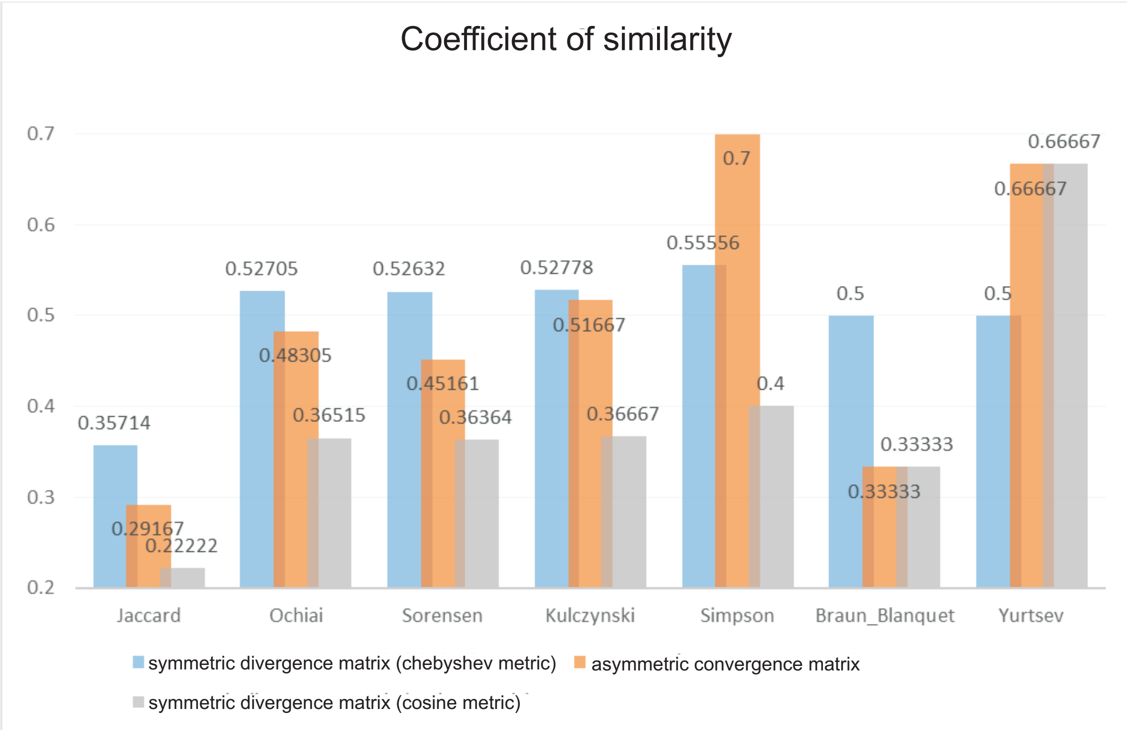

Fig. 1. Coefficients of similarity of DELCO segmentation algorithm results and results obtained by the developed method.

4. The marking algorithm based on local magnitudes and practical use

Marking method lies in embedding watermark bits into the speech signal samples. For this it is necessary to solve the problem of finding the necessary samples.

The main criteria for selection of these samples for marking are the difficulty of their identifying by unauthorized persons or devices; the possibility of their calculation after distortion/reconstruction of signal or filtering, intangibility of done changes to the human ear, minimizing data losses.

Search of signal samples is a calculation of the maximum distance between two adjacent samples:

$$d = \max(\|x_i - x_{i+1}\|), \quad x_i \in Y_i. \tag{22}$$

After calculating the samples for marking, watermark bits (text, another speech signal, image, anything that can be digitized and converted in a sequence of bits) sequentially are embedded into calculated samples as follows:

- When encoding 0-th bit, we reset (nullify) least significant bits of three samples from the left reference point; encoding the 1st bit, three least significant bits from the right are reset.
- When decoding the marker, the sum of least significant bits of three adjacent samples from the left and from the right of the marking point are compared. On the basis of comparison we get a bit of marker.

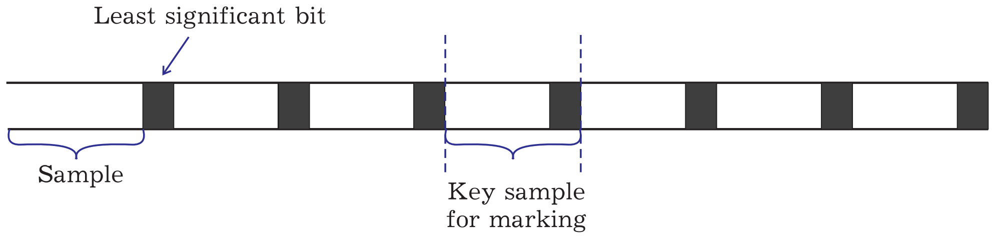

Fig. 2. Layout of encoding/decoding watermark into the speech signal.

The table shows the results (presented as a percentage) of the method. The first number represents the percentage of coincidence of encoded and decoded watermark, and the second one represents the percentage of coincidence points of the speech signal, in which encoding/decoding watermark had a place.

Table 2. Results encoding/decoding a watermark into a digital signal.

Direct segmentaion	Segmentation into quasistationars			Test word
Without FFT	With FFT	Without FFT	With FFT	
85 / 87	95 / 100	89 / 92	97 / 100	Тест
80 / 80	93 / 100	79 / 79	96 / 100	Миша

5. Conclusions

On the basis of carried out experiments, it has been established that the method of digital marking speech signals using local magnitudes and developed technique of speech signal segmentation improves the search of maximum amplitudes in the spectrum, what gives a more accurate result of encoding and decoding the watermark into the speech signal.

[1] Albert A. Regression, pseudoinverse, and recursive estimation. Moscow: Nauka (1977).

[2] Dorohin O., Starushko D., Fedorov E., Shelepov V. The segmentation of the speech signal. Artificial Intelligence. **3**, 450–458 (2000).

[3] Kaganovsky Y. Application of linear prediction model for the analysis of stochastic signals. Engineering: Tradition and Innovation: Proceedings of intern. sci. conf. Chelyabinsk: Two Komsomolets, 12–14 (2012).

[4] Markel J., Gray A. Linear prediction of speech. Edited by Prohorov Yu. and Zvezdin V. Moscow: "Svyaz" (1980).

[5] Distance Matrix. [Electron resource] http://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0_%D1%80%D0%B0%D1%81%D1%81%D1%82%D0%BE%D1%8F%D0%BD%D0%B8%D0%B.

[6] Siomkin B., Dvoychenkov V. On the equivalence of measures of similarity and difference. Systems Research. V. 1. Analysis of complex systems. Vladivostok: Far Eastern Scientific Center, Academy of Sciences of the USSR (1973).

[7] Bassia P., Pitas I., Nikolaidis N. Robust audio watermarking in the time domain. IEEE Transaction on Multimedia. **3**, n.2, 232–241 (2001).

[8] Braun-Blanquet J. Pflanzensoziologie Grundzьge der Vegetationskunde. Berlin: Verlaq von Julius springer (1928).

[9] Cox I., Killian J., Leighton F., Shamoon T. Secure Spread Spectrum Watermarking for Multimedia. IEEE Transactions on Image Processing. **6**, n.12, 1673–1687 (1997).

[10] Huang J., Wang Y., Shi Y. Q. A Blind audio watermarking algorithm with self-synchronization. Proceedings of IEEE International Symposium on Circuits and Systems, (ISCAS02), 627–630, **3** (2002).

[11] Jaccard P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines. Bull. Soc. Vaudoise sci. Natur. **37**, Bd.140, 241–272 (1901).

[12] Kulczinsky S. Zespoly ryslin w Pienach. Bull. intern. acad. polon. sci. lett. Cl. sci. math. natur. Ser. B. **2**, 57–203 (1927).

[13] Lee S. K., Ho Y. S. Digital Audio Watermarking in the Cepstrum Domain. IEEE Transactions on Consumer Electronics. **46**, n.3, 744–750(2000).

[14] Lie W. N., Chang L. C. Robust and High-Quality Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modification, IEEE Transaction on Multimedia. **8**, n.1, 46–59 (2006).

[15] Liu S. C., Lin S. D. BCH Code Based Robust Audio Watermarking in the Cepstrum Domain. Journal of Information Science and Engineering. **22**, 535–543 (2006).

[16] Ochiai A. Zoogeographical studies on the soleoid fishes found Japan and its neighboring regions. II. Bull. Jap. Soc. sci. Fish. **22**, n.9, 526—530 (1957).

[17] Penrose R. A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society. **51**, 406–413 (1955).

[18] Pooyan M., Delforouzi A. Adaptive and Robust Audio watermarking in Wavelet Domain. Proceedings of International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007). **2**, 287–290 (2007).

[19] Rashkevych Y., Peleshko D., Kovalchuk A., Kupchak M., Pelekh Y. Speech signal pseudoinvariants. Computer Science and Information Technologies: Materials of the Vith International Scientific and Technical Conference CSIT 2011. Lviv: Publishig House Vezha&Co. 21–22 (2011).

[20] Simpson G. G. Holarctic mammalian faunas and continental relationship during the Cenozoic. Bull. Geol. Sci. America. **58**, 613–688 (1947).

[21] Sųrensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. Kongelige Danske Videnskabernes Selskab. Biol. krifter. **4**, 1–34 (1948).

[22] Szymkiewicz D. Une contribution statistique a la gňographie floristique. Acta Soc. Bot. Polon. **34**, n.3, 249—265 (1934).

[23] Wang X. Y., Zhao H. A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT. IEEE Transaction on Signal Processing. **54**, n.12, 4835–4840 (December 2006).

[24] Xie L., Zhang J., He H. Robust Audio Watermarking Scheme Based on Nonuniform Discrete Fourier Transform. Proceedings of IEEE International Conference on Engineering of Intelligent System, pp. 1–5, 2006.

[25] Zeng G., Qiu Z. Audio Watermarking in DCT: Embedding Strategy and Algorithm. Proceedings of 9th International Conference on Signal Processing (ICSP-09). 2193–2196 (2008).

Маркерування мовного сигналу на основі локальних магнітуд та інваріантної сегментації

Рашкевич Ю. М., Пелешко Д. Д., Пелех Ю. М., Ізонін І. В.

Національний університет «Львівська політехніка»
вул. С. Бандери, 12, 79013, Львів, Україна

У роботі пропонується нова схема водяних знаків на основі інваріантного методу сегментації та використанні локальних магнітуд для маркерування мовних сигналів. Водяний знак вбудовується в обраному виді у піки з магнітудою спектра кожного неперекриваючого кадру аудіосигналу.
Запропонована схема маркерування мовних сигналів забезпечує надійність від декількох видів атак, таких, як шуму, обрізки, повторної дискретизації, повторного квантування, стиснення і фільтрації нижніх частот.

Ключові слова: *мовний сигнал, маркерування, водяний знак, первинна сегментація, симетрична матрична відстаней, псевдо обернена матриця, алгоритм Мура-Пенроуза, сингулярний розклад, евклідова, чебишевська та косинусна метрики, міра подібності*

2000 MSC: 68R99

УДК: 681.142.2; 622.02.658.284; 621.325