UDK 681.142.2; 622.02.658.284; 621.325

Y. Pelech, D. Peleshko
National University "Lviv Polytechnic",
Department of Information Technology Publishing

MARKING SPEECH SIGNAL BASED ON FACTOR HOLDER SMOOTHNESS AND FAST FOURIER TRANSFORM

За допомогою використання показника гладкості Гельдера на спектрі мовного сигналу запропоновано покращений метод маркурування мовних сигналів. Особливістю методу є висока точність декодування водяного знаку та інваріантність від оригінальних даних, сигналу та водяного знаку.

Ключові слова: мовний сигна, цифрови водяний знак, показник Гельдера, квазістаціонарна ділянка, гладкість функції, спектр сигналу, швидке перетворення Фур'є.

A new improved method of speech signals marking is presented. The method is based on employing of Hölder condition to speech signal spectrum. The key features of this method are its high accuracy during watermark decoding and independence from original data like speech signal and watermark.

Keywords: speech signal, digital watermark, Hölder condition, quasistationary, smooth function, signal spectrum, fast Fourier transform.

Introduction and analysis of the literature

Access to information and its rapid spread is a major factor in the development of new technologies in the past decade. Audio information has become more widely used over the Internet. With the increasing use of audio content on the Internet has serious problems, such as forgery, fraud and piracy. Virtually any user of modern electronic devices (gadgets) can copy the audio file and use it for their own purposes, such as use in presentations or for marketing campaigns, etc.

Thus, abuse of copyright is widespread among users of multimedia content. This, in turn, was the motivating factor in the development of new technologies protect audio information. Therefore the problem of piracy prevention information, counterfeiting and copyright infringement is urgent and requires the development of new and effective methods.

Digital markeruvannya speech signals has several options for application [9]. The most common are withholding information, copy protection and identification of personal data.

One of the technologies to protect speech signals is the use of digital watermarking. Digital watermarking methods are considered to be an effective solution to the problems of copyright. In order to confirm the copyright of the digital content, watermarking methods embed the secret information about the holder, using the masking effect of Human Visual System or Human Audio System.

Digital watermarks have proven to be one of the best ways to protect intellectual property from illegal copying. A digital watermark is a signal that is embedded in the digital data (audio, image, video and text) and to verify the identity of the data to be extracted only the specialized.

Given the relatively recent spread of technology, with the object of using watermarks can be divided into two classes: the watermark image and sound signals. The main object of research is in the speech signal. In this respect the method of implementation of algorithms technology in audio watermarking techniques can be divided into two groups: methods based on processing time series (rows) and methods based on execution sequences of transformations.

Usually the basic principle markeruvannya digital watermark is a hidden need replacement components watermarked audio signal. With technology Perceptual Audio Coding (MPEG Layer-On (IGC), MPEG-2 and MPEG-4) it is known that even slight changes in the speech signal may lead to distortion [8]. Therefore, camouflage replacement should be characterized by stability, transparency and reliability. Here, transparency refers to a perception of the speech signal, which is characterized by the fact that users can not feel the difference with the audio signal from the original watermarked signal.

Transparency perception is one of the major requirements for watermarking technology. In the process of embedding prohibited the introduction of significant tangible artifacts. That watermark should not affect the quality perception of the original audio signal. On the other hand, to ensure the stability and reliability, the power of the watermark is to be maximum which is in some contradiction to transparency. It defines the problem of finding some balance between transparency on the one hand, and stability and reliability of the other.

It should be noted that the stability of the watermark determines the protection against unauthorized access and means that the watermark can be read only by an authorized person. With this condition the technology watermark imposed another condition, which is the need to ensure sufficient capacity is integrated in such a number of the alarm bits that would ensure sustainable security.

This paper proposes a new watermarking scheme based on fast Fourier transform (FFT) for markeruvannya speech signals. The watermark is embedded in a given form of peaks with a magnitude spectrum of each frame is overlapping audio.

Once it should be noted that the proposed labeling scheme of voice signals ensures the stability and reliability of several types of attacks, such as noise, cropping, oversampling, re-quantization, compression and lowpass filtering.

Statement of the problem

The aim is to improve the method of "blind" labeling speech signal [1], which is based on calculating the Hölder exponent for quasi-stationary areas. The method of constructing quasi areas based singular decomposition based on digital signals [2, 4, 5].

1. Presentation signal

By its nature, the speech signals are continuous. However, in this paper we consider only discrete signals. Below is a representation of the signal and its division into quasi-stationary areas that do not intersect.

Suppose we have a signal $x(t)$, де $t = 0, 1, 2, \ldots, N$ – splitting functions. After constructing quasi areas [4] we obtain:

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

where Y_i – quasi-stationary area serving associations successive elementary areas x_i.

$$Y_i \leftrightarrow \bigcup_{j=I_i}^{m_i+I_i-1} x_j \ . \tag{2}$$

From the formula (2), m – number of quasi areas Y_i, I_i – initial quasi-stationary area code, m_i – number of elementary areas x_i in association (1). Thus, it follows that $m \le t$.

To solve problem (2) the method proposed in [2]. The application of this method we obtain a new representation of the speech signal as a series of quasi-stationary areas Y_i, which do not intersect.

After initial segmentation labeling problem is to find the corresponding signal samples for encoding / decoding a watermark on the site. It is necessary to find the most "favorable" signal sections which receive minimal distortion in the reconstruction. Search areas offered himself done through calculating Holder, which in its physical sense would indicate a maximum amplitude at intervals (kvazistatsionarah). To improve the stability of the method to noise, filtering and other undesirable distortion is proposed to calculate the Hölder exponent spectrum of the input signal, which is built using fast Fourier transform.

2. Algorithm and embedding scheme

2.1. Calculating Holder exponent

Suppose that the speech signal $x(t)$ in the interval Y_i is a smooth manner k. Then, by [6], the function $x(t)$ at the point of l, $l \in Y_i$ satisfies the Holder exponent α, $0 \le \alpha \le 1$, if there is a constant $A \in \mathbf{R}$ and for all points t with neighborhood l performed condition

$$\left|x(t)-x(l)\right| \le A\left|t-l\right|^{\alpha}. \quad (3)$$

According to [6] it is known that the function $x(t)$ at the point of l, $l \in Y_i$ satisfies the Holder exponent α, $\alpha \ge 1$, if and only if there exists a constant $A \in \mathbf{R}$ and polynomial $p_l(t)$ degree $m = \lfloor \alpha \rfloor$ such that for all points t with neighborhood l performed condition

$$\left|x(t)-p_l(t)\right| \le A\left|t-l\right|^{\alpha}. \quad (4)$$

According to [6] function $x(t)$ in $[a, b]$ satisfies the Holder exponent α, if $x(t)$ satisfies the Hölder exponent α for all $l \in Y_i$, with a constant $A \in \mathbf{R}$, which is independent of l.

At each interval Y_i polynomial $p_l(t)$ determined only way. If $x(t)$ is continuously differentiable $k = \lfloor \alpha \rfloor$ once a interval Y_i, then $p_l(t)$ is Taylor scheduled $x(t)$ at the point of l. If $x(t)$ satisfies a uniform Holder exponent condition $\alpha > m$ in the interval Y_i, then, as shown in [6], the function $x(t)$ necessarily k-times continuously differentiable on this interval. Limited, but bursting at the point l function satisfies Holder exponent 0 to l. If smoothness is Hölder exponent is $\alpha < 1$ at the point of l, then $x(t)$ not differentiated in l, and α determines the type of rupture [6]. Uniform Holder exponent smooth functions $x(t)$ in Y_i associated with asymptotic decline in its Fourier transform [6].

2.2. Discrete Fourier transform and discrete spectrum signals

It is known that any complex periodic signal can be fed via a Fourier series as a sum of simple harmonic oscillation. The combination of simple harmonic oscillation, which can be decomposed complex periodic signal is called its spectrum.

The distribution of the amplitudes of harmonics in frequency is called the amplitude-frequency spectrum or amplitude spectrum, and the distribution of their initial phases of the frequency - phase-frequency spectrum and phase spectrum.

Lines discrete spectrum with the dimension of the signal amplitude. Continuous spectrum indicates the distribution of the amplitudes across the spectrum and has the dimension of density signal amplitudes.

Fourier transforms - integral transforms one complex-significant functions of a real variable to another, closely related to the Laplace transform and equivalent decomposition in Fourier series for nonperiodic functions. Under this transformation function can be represented in the form of fluctuating additive operator functions.

The Fourier transform of the speech signal functions $x(t)$ mathematically defined as a complex function $F(\omega)$, given by the integral

$$F(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt. \quad (5)$$

The inverse Fourier transform given by

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{-i\omega t}d\omega = f(t). \quad (6)$$

It is known that periodic signals $x(t) = x(t + mT_0)$ represented by Fourier series as a sum of harmonic components, as in this case, the analyzed signal is only the fundamental harmonic $\omega_0 = 2\pi/ T_0$ and its multiple components k - ω_0. In the exponential form of Fourier series has the form

$$x(t) = \ldots + \xi_{-2}e^{-j2\omega_0 t} + \xi_{-1}e^{-j\omega_0 t} + \xi_0 + \xi_1 e^{j\omega_0 t} + \xi_2 e^{j2\omega_0 t} + \ldots = \sum_{k=-\infty}^{\infty} \xi_k e^{jk\omega_0 t}. \quad (7)$$

Coefficients of the series ξ_k are complex variables and are determined from the ratio

$$\xi_k = C_k e^{j\psi k} = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)e^{jk\omega_0 t}dt = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)\cos(k\omega_0 t)dt - j\frac{1}{T}\int_{-T_0/2}^{T_0/2} x(t)\sin(k\omega_0 t)dt. \quad (8)$$

The set of coefficients of the spectrum is a periodic signal. The range of amplitudes C_k and phase spectrum ψ_k uniquely determine the signal $x(t)$ determine the part each frequency harmonic components in the composition of the resulting vibrations. Often, only limited consideration C_k which determines the energy properties of the signal, and ψ_k relates only to the waveform.

Discrete signals $x(n)$ formed by sampling (taking samples) continuous signal $x(t)$ in time- interval T_S. By theorem of Kotelnikov-Shannon sampling frequency F_S = 1/ T_S must exceed at least twice the maximum frequency in the spectrum of sampled signal. For discrete signals pair of integral Fourier transforms (5) and (6) takes the form:

$$\chi(\Omega)=\sum_{n=-\infty}^{+\infty} x(n)-e^{-j\Omega}; \qquad x(n)=\frac{1}{2\pi}\int_{-\pi}^{+\pi}\chi(\Omega)-e^{j\Omega}d\Omega. \tag{9}$$

where $\Omega=\omega-T_S=2\pi\frac{f}{F_S}$ - sampling rate normalized angular frequency.

Given the periodic spectrum $\chi(\Omega)$ discrete signal of period F_S (for normalized circular frequency 2π), in expression (10) limits of integration narrowed to one period, including zero frequency.

In practice, the spectral analysis of real signals is considered continuous signal $x(t)$, and a finite sequence of values of samples $x(n)$ in a given format. However, the use of the expression (9) to calculate the spectral function $\chi(\Omega)$ impossible through endless boundary value summation. That is why the technique of digital signal processing for spectral analysis used a discrete Fourier transform.

Discrete Fourier transform (DFT) is a pair of mutually single-valued transformation that connects the sampling discrete periodic signal $x(n)$ with complex coefficients $\chi(k)$ its discrete spectrum:

$$\chi(k)=\sum_{n=0}^{N-1} x(n)-\mathrm{e}^{-j\frac{2\pi nk}{N}}, k=0,...,N-1, \qquad x(n)=\frac{1}{N}\sum_{k=0}^{N-1}\chi(k)-\mathrm{e}^{j\frac{2\pi nk}{N}}, n=0,...,N-1, \tag{10}$$

where N - number of samples in the period of the sampled signal.

The first transformation (10) is straightforward, and the second conversion - feedback DFT (FDPF). DFT is performed on finite periodic sequence of signal samples $x(n)$, in which the period consisting of N discrete values, $x(n) = x(n + N)$. Sequence $\chi(k)$ represents the spectrum of the sampled signal $x(n)$, which is also a periodic $\chi(k) = \chi(k + N)$.

Comparison of expressions (9) and (10) shows that the DFT represents a discrete sample $\chi(k)$ continuous spectral function $\chi(\Omega)$ discrete signal corresponding frequencies $\omega_k=2\pi\frac{k}{NT_S}$:

$$\chi(k)=\chi(\omega_k)=\chi\left(2\pi\frac{k}{NT_S}\right). \tag{13}$$

Thus, the value of discrete frequencies ψ_k in the spectrum depends on the sampling period T_S, identification and resolution in frequency (spectral selectivity) is inversely proportional to the time the signal surveillance

$$T_0=NT_S. \tag{14}$$

From relation (13) follows an important conclusion: if the current number of samples N to add a number of zeros, such as N, then the spectral function of a discrete signal changes, and DFT will double the number of spectral samples that correspond to frequencies, closely located in the range from zero to the sampling frequency.

3. Results and conclusions

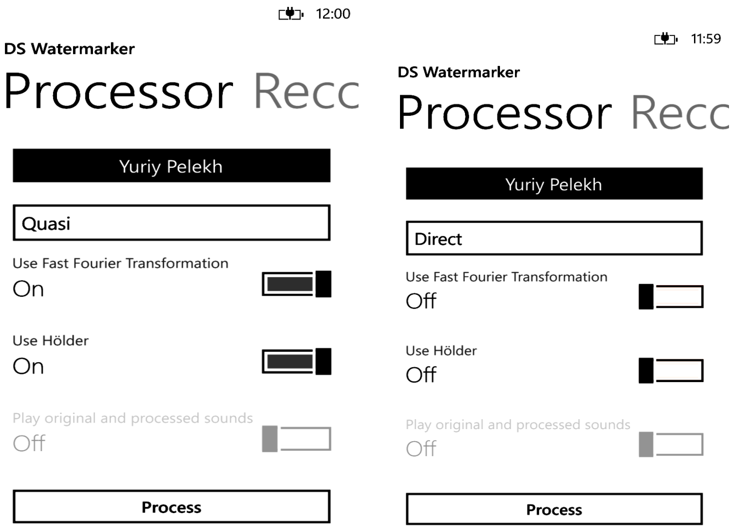
The results of this theoretical study have been verified experimentally. On the basis of the proposed method is implemented algorithm markeruvannya speech signal, which consists of the following steps:

- First, the speech signal is divided into uniform serial sections;
- Further, by combining, according to [2] built kvazistatsionary (primary segmentation of the speech signal);

- Using Fourier transform [7] obtained the spectra of quasi-stationary areas;
- Each section is calculated quasi Holder exponent;
- Using Holder exponent determined by counting signal which is encoded / decoded [5] watermark.

Software implementation of the algorithm was tested on the example label mono recordings words “test” and “mouse”. Characteristics of speech signals are as follows: length of signal 954 milliseconds and 1.09 seconds respectively. Sampling frequency 16 kHz. The length of the elementary section 512 samples. For the watermark is taken into “Yuriy Pelekh”. The test was carried out ten times for five different entries for each of the test words.

Note that this software implementation (Pict. 1) allows the user to choose the para-meters encoding / decoding watermarks. For example, to build FFT spectrum can be used or not used. Then search for labeling samples will not run on spektri signal and on the signal. You can also split the signal at kvazistatsionary or ele¬mentarni uniform areas. Alternatively, the Hölder exponent can be calculated magnitude between two adjacent reference.



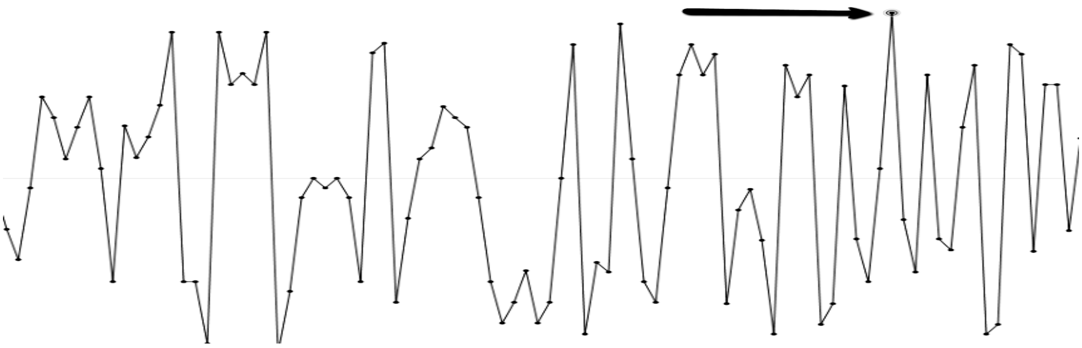
Pict. 1. Software implementation (choice of parameters labeling)

The table shows the results (presented in percentages) of the method of using the Hölder exponent and without. The first number represents the percentage of resemblance encoded and decoded watermark, and the second - the percentage of coincidence points of the speech signal, which happened encoding / decoding watermark.

Results of encoding / decoding a watermark in a digital signal

	Direct decomposition	Breakdown by quasi-stationary			Test word
	Without FFT	With FFT	Without FFT	With FFT	
Using Holder exponent	76 / 79	93 / 100	84 / 87	94 / 99	Test
	74 / 52	93 / 100	61 / 46	89 / 100	Mouse
Without the Holder exponent (calculation of magnitude between adjacent samples)	85 / 87	95 / 100	89 / 92	97 / 100	Test
	80 / 80	93 / 100	79 / 79	96 / 100	Mouse

In Pict. 2 shows the results point at which will be labeled signal with the help mo¬hoyu-Hölder exponent.



Pict. 2. Calculation smooth signal with Holder exponent

On the basis of the experiments proved that the improved method marke¬ru¬vannya digital voice signals using Hölder exponent and splitting the signal at quasi-stationary area, improves search on the maximum amplitude of the spectrum, which gives a more accurate result in the encoding and decoding of the watermark in the speech signal.

1. Peleshko D., Pelekh Y., Izonin I., Peleshko M. Digital Watermarking of Speech Signals. CSIT'2012, p.108, November 2012. 2. Пелешко Д.Д. Виділення псевдоінваріантів та квазістаціонарних ділянок мовних сигналів на основі сингулярних розкладів/ Д.Д.Пелешко, А.М.Ковальчук, Ю.М. Пелех, В.І.Киричук// Вісник НУ "Львівська політехніка" Комп'ютерні науки та інформаційні технології. – №732. – 2012. – С.58-66. 2. Пелешко Д. Д., Пелех Ю. М. Модифікація методу маркування мовних сигналів на основі моментів Зерніке. ISDMCI'2012, ст.392-393, травень 2012. 3. Peleshko D., Pelekh Y., Izonin I. Digital Watermarking of Speech Signals Based on Quasistationary Areas. CADSM'2013, p.283, February 2013. 4. Peleshko D.D., Pelekh Y.M., Izonin I.V., Constructing Of Pseudoinvariants And Digital Watermarking Of Speech Signals Based On A Singular Value Decomposition, Journal of Global Research in Computer Science. Volume 4, No. 2, February 2013. pp. 56-59. 5. Ручай А.Н. Текстозависимая верификация диктора: модель, статистические исследования, комплекс програм. Автореферат диссертациия на соискание ученой степени к.ф.-м.н. Челябинский государственний университет, 2012. 6. Білинський Й.Й., Огородник К.В., Юкиш М.Й. Електронні системи. [Електронний ресурс]. Режим доступу: http://posibnyky.vntu.edu.ua/e_s/3.htm. 7. Хуан Пенцзюнь, Рєзнік Ю. Масштабоване кодування мови та аудіо з використанням комбінаторного кодування MDCT-спектра. Патент України. №95185. Опубліковано: 11.07.2011. [Електронний ресурс]. Режим доступу: http://uapatents.com/26-95185-masshtabovane-koduvannya-movi-ta-audio-z-vikoristannyam-kombinatornogo-koduvannya-mdct-spektra.html#formula. 8. Стасюк В. І., Тимощук А. О., Старіков В. Д., Зайченко М. М. Спосіб захисту від несанкціонованого прослуховування та маскувальник мови для його здійснення Патент України. №: 44665. Опубліковано: 15.02.2002. [Електронний ресурс]. Режим доступу: http://uapatents.com/3-44665-sposib-zakhistu-vid-nesankcionovanogo-proslukhovuvannya-ta-maskuvalnik-movi-dlya-jjogo-zdijjsnennya.html.