TELECOMMUNICATIONS AND RADIO ENGINEERING

O. Yu. Lavrynenko

A METHOD FOR BIOMETRIC CODING OF SPEECH SIGNALS BASED ON ADAPTIVE EMPIRICAL WAVELET TRANSFORM

Department of Telecommunication and Radio Electronic Systems, Faculty of Air Navigation,
Electronics and Telecommunications, State University "Kyiv Aviation Institute", Kyiv, Ukraine
E-mail: oleksandrlavrynenko@gmail.com   ORCID 0000-0002-7738-161X

Abstract—In this research, a biometric speech coding method is developed where empirical wavelet transform is used to extract biometric features of speech signals for voice identification of the speaker. This method differs from existing methods because it uses a set of adaptive bandpass Meyer wavelet filters and Hilbert spectral analysis to determine the instantaneous amplitudes and frequencies of internal empirical modes. This makes it possible to use multiscale wavelet analysis for biometric coding of speech signals based on an adaptive empirical wavelet transform, which increases the efficiency of spectral analysis by 1.2 times or 14 % by separating high-frequency speech oscillations into their low-frequency components, namely internal empirical modes. Also, a biometric method for encoding speech signals based on mel-frequency cepstral coefficients has been improved, which uses the basic principles of adaptive spectral analysis using an empirical wavelet transform, which also significantly improves the separation of the Fourier spectrum into adaptive bands of the corresponding formant frequencies of the speech signal.

Index Terms—Speech signal; biometric coding; speaker identification; information protection; voice authentication; wavelet transform; bandpass wavelet filters; mel-frequency cepstral coefficients.

I. INTRODUCTION

The development of new methods and means of ensuring information security is intended primarily to prevent threats of access to information resources by unauthorized persons. To solve this problem, it is necessary to have identifiers and create identification procedures for all users. Modern identification and authentication include various systems and methods of biometric identification [1]. One of the most common biometric characteristics of a person is his or her voice, which has a set of individual characteristics that are relatively easy to measure (e.g., the frequency spectrum of the voice signal). The advantages of voice identification also include ease of application and use, and the fairly low cost of devices used for identification (e.g., microphones).

Voice identification capabilities cover a very wide range of tasks, which distinguishes them from other biometric systems. First of all, voice identification has been widely used for a long time in various systems for differentiating access to physical objects and information resources. Its new application in systems based on telecommunication channels seems promising. For example, in mobile communications, voice can be used to manage services, and the introduction of voice identification helps protect against fraud [2].

Voice identification also plays an important role in solving such an important task as protecting speech information. This identification is used to create new technical means and software and hardware devices for protecting speech information, in particular, from leakage through acoustic, vibroacoustic and other channels [3].

Voice identification is of particular importance in the investigation of crimes, in particular in the field of computer information, and in the formation of the evidence base for such an investigation. In these cases, it is often necessary to identify an unknown voice recording. Voice identification is an important practical task when searching for a suspect based on a voice recording in telecommunication channels. Determining such characteristics of the speaker's voice as gender, age, nationality, dialect, and emotional coloring of speech are also important in the field of forensics and anti-terrorism. The identification results are important in conducting phonoscopic examinations, in carrying out expert forensic research based on the theory of forensic identification [4].

Thus, the development of new methods of voice identification is a promising and relevant scientific and technical task in providing biometric authentication in information and telecommunication systems.

II. LITERATURE REVIEW AND PROBLEM STATEMENT

The paper investigates a well-known method of biometric coding of speech signals based on mel-frequency cepstral coefficients (MFCC) [5], which consists in finding the average values of the coefficients of the discrete cosine transform (DCT)

$$c[n] = \sum_{m=0}^{N_f - 1} E[m] \cos\left(\frac{\pi n \left(m + \frac{1}{2} \right)}{N_f} \right),$$

$$n = 0, \dots, N_f - 1,$$

prologarithmized energy of the spectrum [6]

$$E[m] = \ln\left(\sum_{k=0}^{N-1} |X[k]|^2 H_m[k] \right), \quad m = 0, \dots, N_f - 1,$$

discrete Fourier transform (DFT)

$$X[k] = \sum_{n=0}^{N-1} x[n] w[n] e^{\frac{-2\pi j}{N} kn}, \quad k = 0, \dots, N-1,$$

processed with a triangular filter [7]

$$H_m[k] = \begin{cases} 0, & k < f[m-1] \\ \dfrac{(k - f[m-1])}{(f[m] - f[m-1])}, & f[m-1] \le k < f[m] \\ \dfrac{(f[m+1] - k)}{(f[m+1] - f[m])}, & f[m] \le k \le f[m+1] \\ 0, & k > f[m+1] \end{cases}$$

where

$$f[m] = \left(\frac{N_f}{Fs} \right) M^{-1} \left(M(F_{\min}) + m \frac{M(F_{\max} - F_{\min})}{N_f + 1} \right)$$

in mel scale $M = 1127.01048 \ln(1 + F/700)$.

The problem is that the presented method of biometric encoding of speech signals based on MFCC does not meet the condition of adaptability [8]

$$\bigcup_{n=1}^{N} \Lambda_n = [0, \pi],$$

where $\Lambda_n = [\omega_{n-1}, \omega_n]$ are the segments of the Fourier spectrum $[0, \pi]$ of the speech signal under study, which is divided into N adjacent segments with boundaries ω_n (where $\omega_0 = 0$ and $\omega_N = \pi$), which leads to suboptimal extraction of biometric features of speech signals and to a decrease in the probability of recognizing the voice features of a person [9].

Therefore, it is necessary to develop a new method of biometric coding of speech signals based on empirical wavelet transform (EWT). This method should differ from existing approaches by constructing a system of adaptive bandpass Meyer wavelet filters, followed by the use of Hilbert spectral analysis to determine the instantaneous amplitudes and frequencies of the functions of internal empirical modes. The application of this method will reveal the biometric characteristics of speech signals and increase the efficiency of their coding.

III. PROPOSED METHOD

The developed method includes the following steps (Fig. 1). The speech signal, whose frequency range is from 300 to 3400 Hz, is divided into K frames of 20 ms in length by N counts, which intersect at $1/2$ frame length to ensure the stationarity of the process (Fig. 2).

The sequence of counts of the K frame is submitted to the DFT.

$$X[k] = \sum_{n=0}^{N-1} x[n] w[n] e^{\frac{-2\pi j}{N} kn}, \quad k = 0, \dots, N-1,$$

where the Hamming window is used as a weighting function:

$$w[n] = 0.53836 - 0.46164 \cdot \cos\left(2\pi \frac{n}{N-1} \right),$$

$$n = 0, \dots, N-1.$$

The values of k indexes correspond to frequencies:

$$f[k] = \frac{F_s}{N} k, \quad k = 0, \dots, N/2,$$

where, F_s is the sampling rate of the speech signal.

The normalized Fourier spectrum in terms of frequency $[0, \pi]$ and amplitude $[0,1]$ is divided into N segments $\Lambda_n = [\omega_{n-1}, \omega_n]$, where $\omega_n = (\Omega_n + \Omega_{n+1})/2$ are the segment boundaries ($\omega_0 = 0$ and $\omega_N = \pi$), and Ω_n are local maxima in the frequency spectrum characterizing the biometric features of speech signals, then it is obvious that $\bigcup_{n=1}^{N} \Lambda_n = [0, \pi]$ (Fig. 3). Each boundary (filter cutoff frequencies) ω_n, has a transient phase of width $2\tau_n$, where τ_n is chosen in proportion to ω_n: $\tau_n = \gamma \omega_n$, and the parameter γ must meet the condition [10]:

$$\gamma < \min_n \left(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right), \quad 0 < \gamma < 1,$$

which guarantees the absence of overlap between the transition regions $2\tau_n$ and ensures the orthogonality of the basis of the bandpass Meyer wavelet filters $\left\{\varphi_1(t), \left\{\psi_n(t)\right\}_{n=1}^{N}\right\}$.

Weight function

Speech signal → Splitting the signal into frames → Fourier transform → Segmentation of the Fourier spectrum

Empirical wavelet transform ← Construction of Meyer bandpass filters

Hilbert spectral analysis

Threshold processing

Biometric features of speech signals
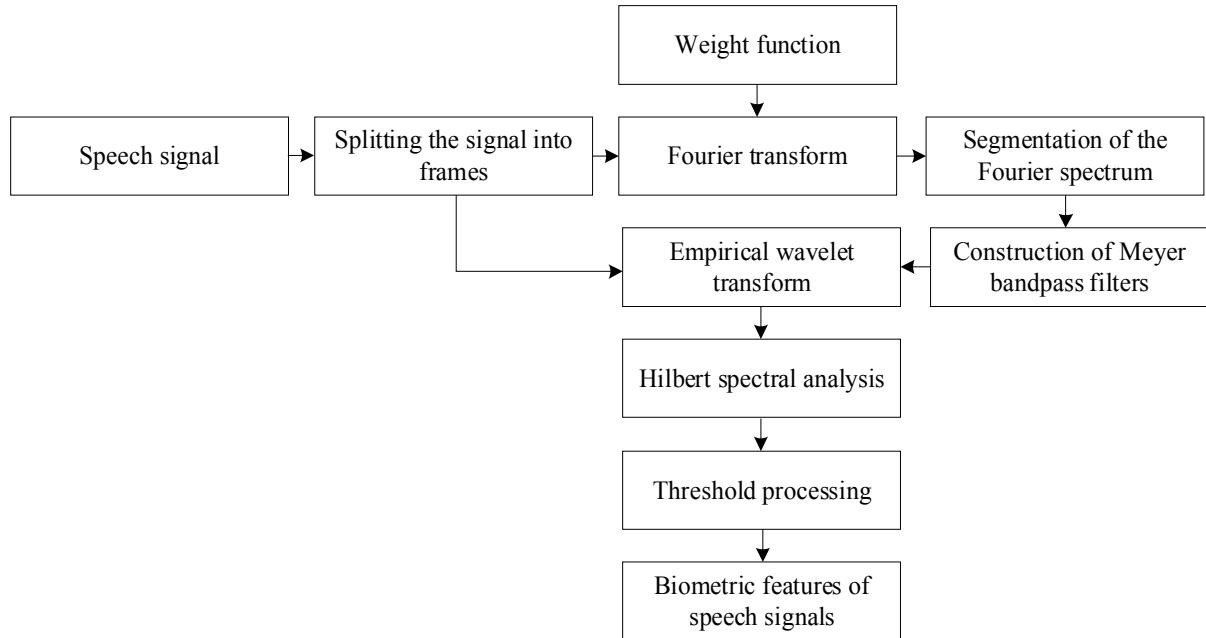
Fig. 1. Method of biometric coding of speech signals based on EWT
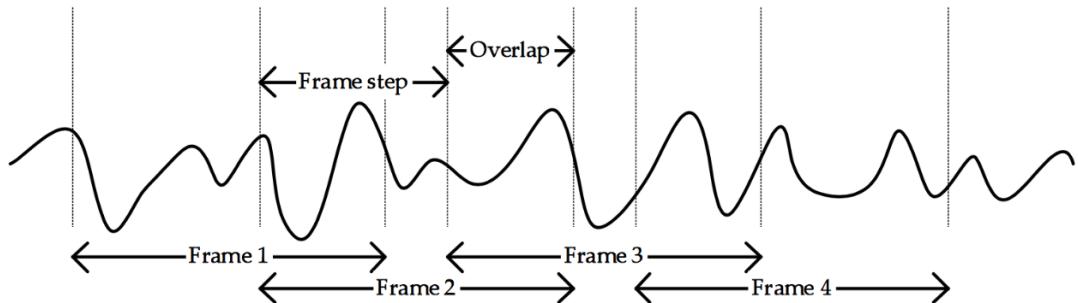

Fig. 2. Splitting the speech signal into frames
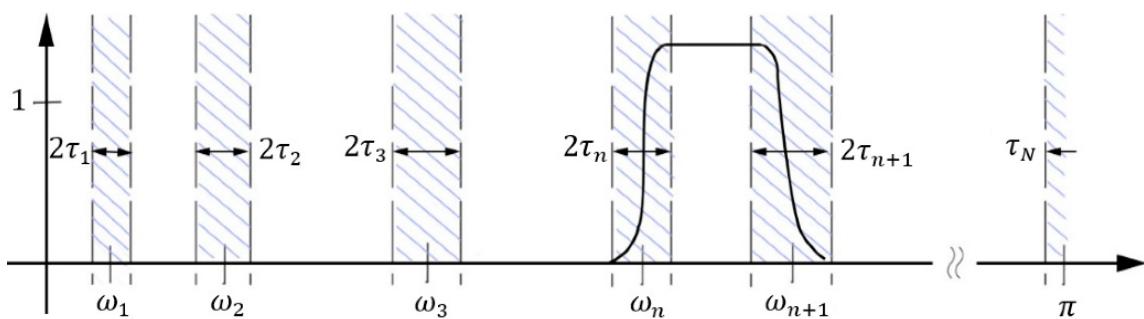
Fig. 3. Fourier spectrum separation by adaptive low-pass $\phi_1(\omega)$ and bandpass $\psi_n(\omega)$ Meyer filters

Then $\forall n > 0$, the adaptive basis $\left\{\varphi_1(t), \left\{\psi_n(t)\right\}_{n=1}^{N}\right\}$ is set by the scaling function $\hat{\phi}_n(\omega)$ and wavelet functions $\hat{\psi}_n(\omega)$, which corresponds to the low-pass filter and $N-1$ bandpass Meyer filters for each spectrum segment Λ_n

$$
\hat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq (1-\gamma)\omega_n \\ \cos\left[\dfrac{\pi}{2}\beta\left(\dfrac{1}{2\gamma\omega_n}\left(|\omega| - (1-\gamma)\omega_n\right)\right)\right] \\ & \text{if } (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases}
$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } (1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_{n+1}}\left(|\omega|-(1-\gamma)\omega_{n+1}\right)\right)\right] \\ & \text{if } (1-\gamma)\omega_{n+1} \le |\omega| \le (1+\gamma)\omega_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}\left(|\omega|-(1-\gamma)\omega_n\right)\right)\right] \\ & \text{if } (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases}$$

where the function $\beta(x)$ must meet the condition

$$\beta(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 1 & \text{if } x \ge 1. \end{cases} \quad \text{and} \quad \beta(x)+\beta(1-x)=1$$
$$\forall x \in [0,1].$$

In practice, the following polynomial function is used [11]

$$\beta(x) = x^4\left(35 - 84x + 70x^2 - 20x^3\right).$$

As can be seen from the scaling function $\hat{\phi}_n(\omega)$ and wavelet functions $\hat{\psi}_n(\omega)$, adaptability is achieved by building bandpass filters centered around the frequencies ω_n, which characterize the biometrics of the speech.

Then the detail coefficients of $W_f^\varepsilon(n,t)$ are given by scalar products with empirical wavelet functions:

$$W_f^\varepsilon(n,t) = \langle f, \psi_n \rangle = \int f(\tau)\overline{\psi_n(\tau-t)}d\tau$$
$$= \left(\hat{f}(\omega)\overline{\hat{\psi}_n(\omega)}\right)^\vee,$$

and the approximation coefficients $W_f^\varepsilon(0,t)$ by a scalar product with a scaling function:

$$W_f^\varepsilon(0,t) = \langle f, \phi_1 \rangle = \int f(\tau)\overline{\phi_1(\tau-t)}d\tau$$
$$= \left(\hat{f}(\omega)\overline{\hat{\phi}_1(\omega)}\right)^\vee,$$

where $\hat{\psi}_n(\omega)$ and $\hat{\phi}_1(\omega)$ are defined by the equations of the wavelet functions $\hat{\psi}_n(\omega)$ and the scaling function $\hat{\phi}_n(\omega)$, respectively.

The reconstruction of the speech signal $f(t)$ using the wavelet coefficients of detail $W_f^\varepsilon(n,t)$ and approximation $W_f^\varepsilon(0,t)$ is given by the following expression

$$f(t) = W_f^\varepsilon(0,t)\cdot\phi_1(t) + \sum_{n=1}^{N} W_f^\varepsilon(n,t)\cdot\psi_n(t)$$
$$= \left(\hat{W}_f^\varepsilon(0,\omega)\cdot\hat{\phi}_1(\omega) + \sum_{n=1}^{N}\hat{W}_f^\varepsilon(n,\omega)\cdot\hat{\psi}_n(\omega)\right)^\vee.$$

Then the internal empirical modes of the studied signal $f(t)$ are given by the formulas

$$f_0(t) = W_f^\varepsilon(0,t)\cdot\phi_1(t),$$
$$f_k(t) = W_f^\varepsilon(k,t)\cdot\psi_k(t),$$

and the orthogonality of the expansion is proved by the fact that [12]

$$f(t) = \sum_{k=0}^{N} f_k(t).$$

To determine the instantaneous frequency and amplitude of the internal empirical modes (IEMs) of the speech signal, we will resort to Hilbert spectral analysis.

The Hilbert transform (HT) of EWT $x(t)$ is given by the following expression

$$y(t) = \frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau,$$

where P is the principal Cauchy value of the singular integral.

With the help of HT EWT $x(t)$ you can get an analytical signal

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)},$$

where $i = (-1)^{1/2}$.

Then the instantaneous amplitude and frequency of the EWT can be expressed as

$$a(t) = \sqrt{(x^2 + y^2)}, \quad \omega(t) = \frac{d\theta}{dt},$$

where the instantaneous frequency of $\omega(t)$ is determined by the rate of change of the instantaneous phase

$$\theta(t) = \arctan\frac{y}{x},$$

and the EWT $x(t)$ can be expressed as the real part of the following equation [13]

$$x(t) = \text{Re}\left\{\sum_{j=1}^{n} a_j(t)\exp\left[i\int\omega_j(t)dt\right]\right\}.$$

Then the Hilbert energy density spectrum is defined as

$$S_{i,j} = H(t_i,\omega_j) = \frac{1}{\Delta t\cdot\Delta\omega}H\left[\sum_{k=1}^{n} a_k^2(t)\right],$$

where the intervals $\Delta t\cdot\Delta\omega$ represent the values of $a^2(t)$ at a given time and frequency.

Let's set the threshold function (Fig. 4), which is described by the following expression:

$$y(x) = \begin{cases} x, & |x| \geq T \\ 0, & |x| < T \end{cases}$$

where x is the value of the coefficients before thresholding, y is the value of the coefficients after thresholding, and T is the threshold [14].
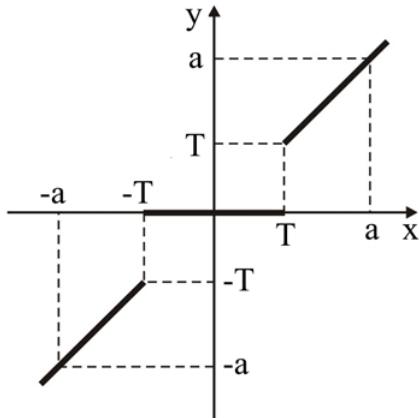

Fig. 4. Threshold function

Let's assume that the probability of recognizing P frequency and amplitude of the function of the harmonic distribution law $x(t) = A \cdot \sin(\omega t + \varphi)$ is 1, and the function of the uniform distribution law
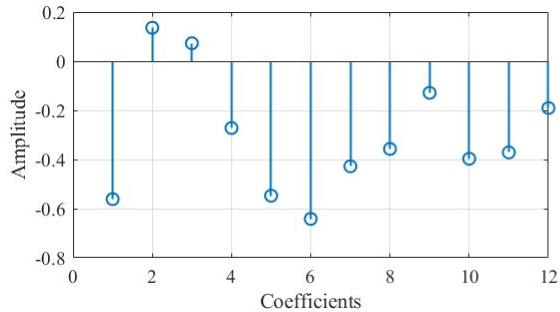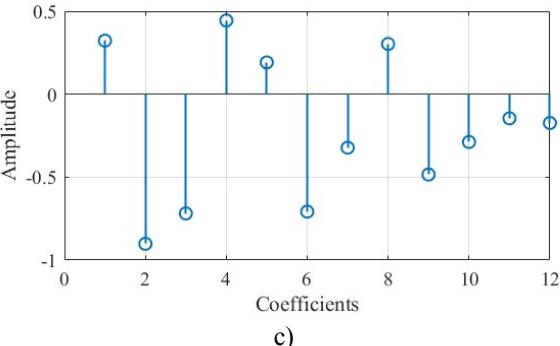
$$x(t) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$$

is $1/2$.

Then the theoretical criterion for finding the maximum possible probability of recognizing the biometric speech features of the analyzed frame is written in the following way, which is based on the balance between the energy of the biometric speech features and their number

$$P = \frac{\sqrt{\sum_{k=1}^{N} |C_{i,...,N}|^2}}{\sqrt{\sum_{k=1}^{N} |C|^2}} = \frac{N-i}{N}, \quad i = 1, ..., N,$$

where C is the Hilbert energy spectrum of length N, and $T = C_i$ [15].

IV. RESEARCH RESULTS

Figure 5 depicts calculated by the developed algorithm of experimental samples of voice commands of the control subject No. 1: "up", "down", "right", "left". In this system to evaluate the results of automatic recognition of voice control commands, a classifier built by the criterion of minimum distance is used.


a)

b)

c)
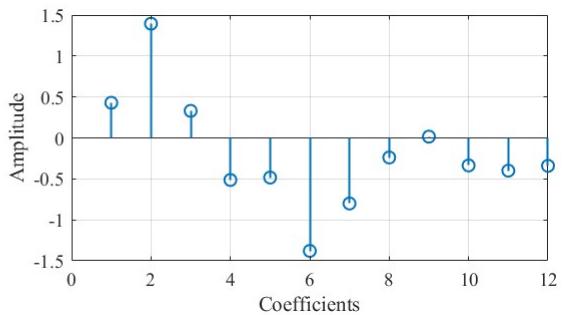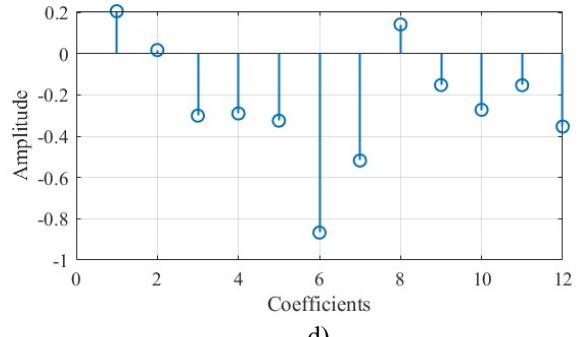
d)

Fig. 5. Recognition features based on the developed method of voice commands of control subject No. 1:
(a) "up", (b) "down", (c) "right", (d) "left"

The dispersion of the difference between the mathematical expectation of the mathematical expectation of the recognition features based on the developed method of the reference voice images

stored in the database and the mathematical expectation of the recognition features based on the developed method at the testing level of the system is used as such an indicator.

The variance of the difference of the difference of the mathematical expectations of two samples of voice control commands (recognition features based on the developed method), is written as follows:

$$D = \frac{\sum_{i=1}^{n}\left(\frac{\sum_{i=1}^{n}x_i}{n} - \frac{\sum_{i=1}^{n}\overline{x}_i}{n}\right)^2}{n},$$

where, x_i is recognition features based on the developed method stored in the base of reference voice images, $\overline{x}_i$ is recognition features based on the developed method at the system testing level, n is number of recognition features based on the developed method.

The decision on biometric identification of voice commands is made by the criterion of minimum variance $D_{\min} < \Theta$ in a given threshold of acceptable recognition $\Theta = 1 - \Delta = 0.15$, where $\Delta = 0.85$.

TABLE I.   TEST RESULTS OF THE RECOGNITION SYSTEM VOICE COMMANDS FROM CONTROL SUBJECT NO. 1

Training	Testing			
Control Subject No. 1	*Control Subject No. 1*			
Voice commands	"Up"	"Down"	"Right"	"Left"
"Up"	0.0222	0.1832	0.2781	0.1381
"Down"	0.5234	0.0559	0.7256	0.4195
"Right"	0.3166	0.4959	0.0034	0.1511
"Left"	0.1648	0.1846	0.1559	0.0023

In the first experiment (Table I), we compared the recognition features based on the developed method of voice commands of control subject No. 1: "up", "down", "right", "left", which were stored at the training level in the base of reference voice images with the recognition features based on the developed method of voice commands of the same control subject No. 1, but already in the system testing mode (the recognition features based on the developed method of spoken voice commands in the testing mode are compared with the recognition features based on the developed method of voice commands spoken earlier in the system training mode). From the obtained results (Table I) it can be seen that the recognition features based on the

developed method of the voice commands of the control subject No. 1 meet the criterion of minimum dispersion $D_{\min}$ in the given threshold of acceptable recognition $\Theta = 0.15$: "up" is $D_{\min} = 0.0222$, "down" is $D_{\min} = 0.0559$, "right" is $D_{\min} = 0.0034$, "left" is $D_{\min} = 0.0023$, based on this, the decision about positive biometric identification of the spoken voice commands is made (voice commands are recognized). In other cases (Table I) it is clearly seen that the values of $D_{\min}$ do not correspond to the selected criterion, which means that the recognition features based on the developed method of the spoken voice commands do not coincide with the recognition features based on the developed method that are stored in the database of reference voice images, i.e. the voice commands are not recognized.

TABLE II.   TEST RESULTS OF THE RECOGNITION SYSTEM VOICE COMMANDS FROM CONTROL SUBJECT NO. 2

Training	Testing			
Control Subject No. 1	*Control Subject No. 2*			
Voice commands	"Up"	"Down"	"Right"	"Left"
"Up"	0.0362	0.3150	0.3912	0.1525
"Down"	0.3682	0.0393	1.1169	0.5372
"Right"	0.1966	0.4899	0.0878	0.1733
"Left"	0.1179	0.2967	0.2675	0.0624

In the second experiment (Table II), we compared the recognition features based on the developed method of the spoken voice commands of control subject No. 2 in the testing mode with the recognition features based on the developed method of the voice commands of control subject No. 1 spoken earlier in the system training mode. From the obtained results (Table II), we can conclude that the recognition features based on the developed method of voice commands of the control subject No. 2 meet the criterion of minimum variance $D_{\min}$ in a given threshold of acceptable recognition $\Theta = 0.15$: "up" is $D_{\min} = 0.0362$, "down" is $D_{\min} = 0.0393$, "right" is $D_{\min} = 0.0878$, "left" is $D_{\min} = 0.0624$, and therefore, a decision is made about positive result of recognizing the spoken voice commands. In all other cases, voice commands are not recognized because the resulting values do not meet the specified recognition criterion.

In the third experiment (Table III), the recognition features based on the developed method of the spoken voice commands of control subject No. 3 in the testing mode were compared with the

recognition features based on the developed method of voice commands of control subject No. 1 spoken earlier in the training mode of the system, which are stored in the database of reference voice images of control commands. The obtained values of the comparison results: "up" is $D_{\min} = 0.0523$, "down" is $D_{\min} = 0.0823$, "right" is $D_{\min} = 0.0757$, "left" is $D_{\min} = 0.0696$, fully meet the criterion $D_{\min} < \Theta$, where $\Theta = 0.15$, and therefore, the decision about the positive result of recognizing the spoken voice commands is made. As for the other obtained resultant values, they do not meet the specified recognition criterion, and thus, the voice commands are not recognized.

TABLE III.   TEST RESULTS OF THE RECOGNITION SYSTEM VOICE COMMANDS FROM CONTROL SUBJECT NO. 3

Training	Testing			
Control Subject No. 1	**Control Subject No. 3**			
Voice commands	"Up"	"Down"	"Right"	"Left"
"Up"	0.0523	0.1568	0.4457	0.1854
"Down"	0.3911	0.0823	1.1771	0.3683
"Right"	0.2051	0.4432	0.0757	0.1695
"Left"	0.1633	0.1857	0.2833	0.0696

V. CONCLUSIONS

The article develops a method of biometric coding of speech signals based on empirical wavelet transform, which differs from existing methods by constructing a set of adaptive bandpass Meyer wavelet filters with the subsequent application of Hilbert spectral analysis to find instantaneous amplitudes and frequencies of functions of internal empirical modes, which will allow to determine biometric features of speech signals and increase the efficiency of their coding by 1.2 times or 14 %. From the obtained results (Table I) we can see that the recognition features based on the developed method of voice commands of the control subject No. 1 meet the criterion of minimum variance $D_{\min}$ in the given threshold of acceptable recognition $\Theta = 0.15$: "up" is $D_{\min} = 0.0222$, "down" is $D_{\min} = 0.0559$, "right" is $D_{\min} = 0.0034$, "left" is $D_{\min} = 0.0023$, based on this, the decision about positive biometric identification of the spoken voice commands is made. From the obtained results (Table II), we can conclude that the recognition features based on the developed method of voice commands of the control subject No. 2 meet the criterion of minimum variance $D_{\min}$ in a given threshold of acceptable recognition $\Theta = 0.15$: "up" is $D_{\min} = 0.0362$, "down" is $D_{\min} = 0.0393$, "right" is

$D_{\min} = 0.0878$, "left" is $D_{\min} = 0.0624$, and therefore, a decision is made about the positive result of recognizing the spoken voice commands. In the third experiment (Table III) the obtained values of the comparison results: "up" is $D_{\min} = 0.0523$, "down" is $D_{\min} = 0.0823$, "right" is $D_{\min} = 0.0757$, "left" is $D_{\min} = 0.0696$, fully meet the criterion $D_{\min} < \Theta$, where $\Theta = 0.15$, and therefore, the decision about the positive result of recognizing the spoken voice commands is made. A software complex has been developed, including means for compiling a database of reference voice images of control subjects for training and testing of the voice control system, and program that delineating the proposed methods and algorithms in the MATLAB environment.

REFERENCES

[1] Y. Zhang, C. Chen and C. Yang, "Task Extension of Robot with Voice Control Based on Dynamical Movement Primitives," *2020 International Symposium on Autonomous Systems (ISAS)*, Guangzhou, China, 2020, pp. 82–87, https://doi.org/10.1109/ISAS49493.2020.9378861.

[2] A. V. Elasarapu, P. Bevara, K. Buramsetty, H. A. Mirza, V. N. Marriwada and N. S. Murthy, "Smart BOT for Face Recognition and Voice Controls," *2024 International Conference on Computing and Data Science (ICCDS)*, Chennai, India, 2024, pp. 1–6, https://doi.org/10.1109/ICCDS60734.2024.10560389

[3] O. Lavrynenko, A. Taranenko, I. Machalin, Y. Gabrousenko, I. Terentyeva and D. Bakhtiiarov, "Protected Voice Control System of UAV," *2019 IEEE 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, Kiev, Ukraine, 2019, pp. 295–298, https://doi.org/10.1109/APUAVD47061.2019.8943926

[4] Y. Ü. Sönmez and A. Varol, "The Necessity of Emotion Recognition from Speech Signals for Natural and Effective Human-Robot Interaction in Society 5.0," *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*, Istanbul, Turkey, 2022, pp. 1–8, https://doi.org/10.1109/ISDFS55398.2022.9800837.

[5] C. -Y. Li and N. T. Vu, "Improving Speech Recognition on Noisy Speech via Speech Enhancement with Multi-Discriminators CycleGAN," *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Cartagena, Colombia, 2021, pp. 830–836, https://doi.org/10.1109/ASRU51503.2021.9688310.

[6] A. Bhattacharjee et al., "Bangla voice controlled robot for rescue operation in noisy environment," *2016 IEEE Region 10 Conference (TENCON)*, Singapore, 2016, pp. 3284–3288, https://doi.org/10.1109/TENCON.2016.7848659.

[7] O. Lavrynenko, B. Chumachenko, M. Zaliskyi, S. Chumachenko and D. Bakhtiiarov, "Method of Remote Biometric Identification of a Person by Voice based on Wavelet Packet Transform," *CEUR Workshop Proceedings*, 2024, vol. 3654, pp. 150–162.

[8] S. Wen, W. -S. Gan and D. Shi, "An Improved Selective Active Noise Control Algorithm Based on Empirical Wavelet Transform," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 1633–1637, https://doi.org/10.1109/ICASSP40776.2020.9054452.

[9] O. Lavrynenko, D. Bakhtiiarov, V. Kurushkin, S. Zavhorodnii, V. Antonov and P. Stanko, "A method for extracting the semantic features of speech signal recognition based on empirical wavelet transform," *Radioelectronic and Computer Systems*, 2023, vol. 107, no. 3, pp. 101–124. https://doi.org/10.32620/reks.2023.3.09.

[10] M. Joorabchi, S. Ghorshi and Y. Naderahmadian, "Speech Denoising Based on Wavelet Transform and Wiener Filtering," *2023 8th International Conference on Frontiers of Signal Processing (ICFSP)*, Corfu, Greece, 2023, pp. 43–46, https://doi.org/10.1109/ICFSP59764.2023.10372899.

[11] M. M. Azmy, "Gender of Fetus Identification Using Modified Mel-Frequency Cepstral Coefficients Based on Fractional Discrete Cosine Transform," in *IEEE Access*, vol. 12, pp. 48158–48164, 2024, https://doi.org/10.1109/ACCESS.2024.3373430.

[12] K. V. Veena and D. Mathew, "Speaker identification and verification of noisy speech using multitaper MFCC and Gaussian Mixture models," *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Thrissur, India, 2015, pp. 1–4, https://doi.org/10.1109/PICC.2015.7455806.

[13] O. Veselska, O. Lavrynenko, R. Odarchenko, M. Zaliskyi, D. Bakhtiiarov, M. Karpinski and S. Rajba, "A Wavelet-Based Steganographic Method for Text Hiding in an Audio Signal," *Sensors*, 2022, vol. 22, no. 15, pp. 1–25. https://doi.org/10.3390/s22155832.

[14] A. Jovanović, Z. Perić, J. Nikolić and D. Aleksić, "The Effect of Uniform Data Quantization on GMM-based Clustering by Means of EM Algorithm," *2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)*, East Sarajevo, Bosnia and Herzegovina, 2021, pp. 1–5, https://doi.org/10.1109/INFOTEH51037.2021.9400662

[15] O. Lavrynenko, R. Odarchenko, G. Konakhovych, A. Taranenko, D. Bakhtiiarov and T. Dyka, "Method of Semantic Coding of Speech Signals based on Empirical Wavelet Transform," *2021 IEEE 4th International Conference on Advanced Information and Communication Technologies (AICT)*, Lviv, Ukraine, 2021, pp. 18–22, https://doi.org/10.1109/AICT52120.2021.9628985.

Oleksandr Lavrynenko. ORCID 0000-0002-7738-161X. Candidate of Science (Engineering). Associate Professor. Department of Telecommunication and Radio Electronic Systems. Faculty of Air Navigation Electronics and Telecommunications, State University "Kyiv Aviation Institute", Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2014).
Research area: telecommunication systems and networks, speech recognition, digital signal processing, information security.
Publications: more than 40.
E-mail: oleksandrlavrynenko@gmail.com

О. Ю. Лавриненко. Метод біометричного кодування мовних сигналів на основі адаптивного емпіричного вейвлет-перетворення

Розроблено метод біометричного кодування мови, в якому емпіричне вейвлет-перетворення використовується для вилучення біометричних ознак мовних сигналів для голосової ідентифікації мовця. Цей метод відрізняється від існуючих тим, що використовує набір адаптивних смугових вейвлет-фільтрів Мейєра та спектральний аналіз Гільберта для визначення миттєвих амплітуд і частот внутрішніх емпіричних мод. Це дає можливість використовувати багатомасштабний вейвлет-аналіз для біометричного кодування мовних сигналів на основі адаптивного емпіричного вейвлет-перетворення, що підвищує ефективність спектрального аналізу в 1.2 рази або на 14% за рахунок розділення високочастотних мовних коливань на їх низькочастотні складові, а саме внутрішні емпіричні моди. Також удосконалено біометричний метод кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів, який використовує основні принципи адаптивного спектрального аналізу з використанням емпіричного вейвлет-перетворення, що також суттєво покращує розділення Фур'є-спектру на адаптивні смуги відповідних формантних частот мовного сигналу.
Ключові слова: мовний сигнал; біометричне кодування; ідентифікація диктора; захист інформації; автентифікація голосу; вейвлет-перетворення; смугові вейвлет-фільтри; мел-частотні кепстральні коефіцієнти.

Лавриненко Олександр Юрійович. ORCID 0000-0002-7738-161X. Кандидат технічних наук. Доцент.
Кафедра телекомунікаційних та радіоелектронних систем. Факультет аеронавігації, електроніки та телекомунікацій, Державний університет "Київський авіаційний інститут", Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2014).
Напрямок наукової діяльності: телекомунікаційні системи та мережі, розпізнавання мови, цифрова обробка сигналів, захист інформації.
Кількість публікацій: більше 40.
E-mail: oleksandrlavrynenko@gmail.com