A Method for Biometric Coding of Speech Signals based on Adaptive Empirical Wavelet Transform*

Oleksandr Lavrynenko[1,*,†], Maksym Zaliskyi[1,*,†], Denys Bakhtiiarov[1,†], Anatolii Taranenko[1,†] and Yevhen Gabrousenko[1,†]

[1] National Aviation University, 1 Lubomyr Huzar ave., 03058 Kyiv, Ukraine

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 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.

Keywords

speech signal, biometric coding, speaker identification, information protection, voice authentication, wavelet transform, bandpass wavelet filters, mel-frequency cepstral coefficients

1. 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, 2].

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) [3].

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 [4].

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 [5].

*CPITS 2025: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, February 28, 2025, Kyiv, Ukraine

* Corresponding author.

† These authors contributed equally.

oleksandrlavrynenko@gmail.com (O. Lavrynenko); maksym.zaliskyi@npp.nau.edu.ua (M. Zaliskyi); bakhtiiaroff@tks.nau.edu.ua (D. Bakhtiiarov); agt705@nau.edu.ua (A. Taranenko); gab58@meta.ua (Y. Gabrousenko)

0000-0002-7738-161X (O. Lavrynenko); 0000-0002-1535-4384 (M. Zaliskyi); 0000-0003-3298-4641 (D. Bakhtiiarov); 0000-0002-0846-6767 (A. Taranenko); 0000-0001-5852-0306 (Y. Gabrousenko)

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 is also important in the field of forensics and anti-terrorism. The identification results are important in conducting phonoscopic examinations, and in carrying out expert forensic research based on the theory of forensic identification [6].

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.

2. 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) [7, 8], which consists of finding the average values of the coefficients of the discrete cosine transform (DCT)

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

prologarithmized energy of the spectrum [9]

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

discrete Fourier transform (DFT)

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

processed with a triangular filter [10]

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

where

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

in mel scale $M=1127.01048\times\ln\left(1+F/700\right)$ [11].

The problem is that the presented method of biometric encoding of speech signals based on MFCC does not meet the condition of adaptability [12]

$$¿n=1¿N\,\Lambda_n=[0,\pi],$$

where $\Lambda_n=\left[\omega_{n-1},\omega_n\right]$ 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 [13–15].

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.

3. Purpose and research objectives

The developed method includes the following steps (see 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 (see Fig. 2) [16].

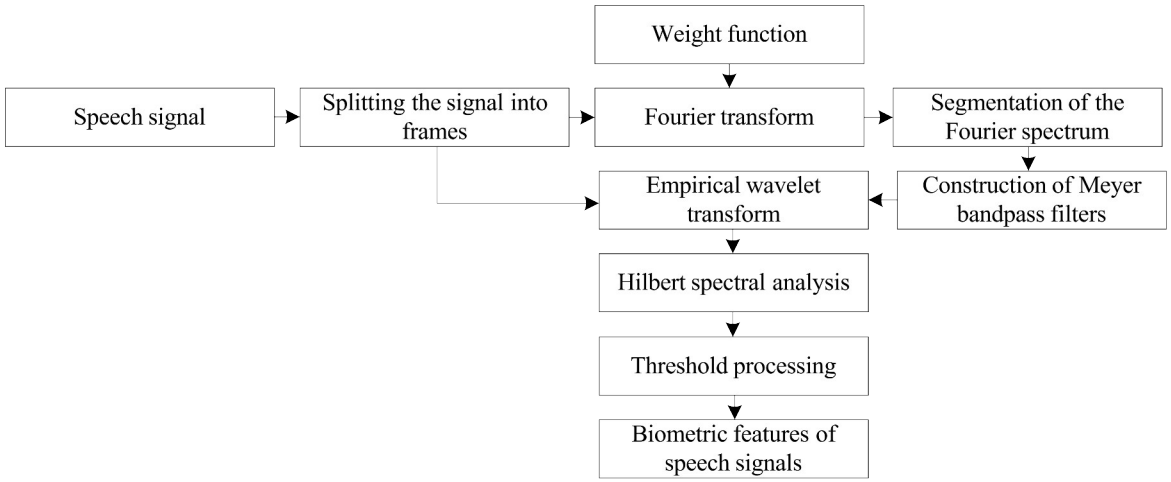

Figure 1: Method of biometric coding of speech signals based on EWT

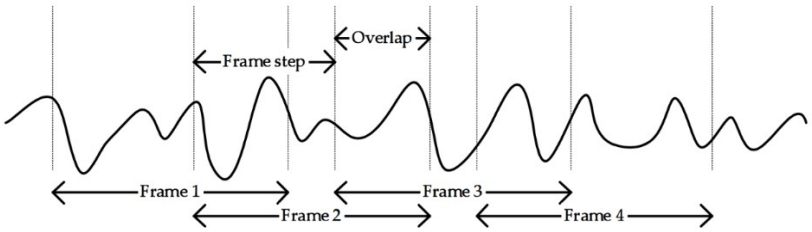

Figure 2: Splitting the speech signal into frames

The sequence of counts of the K frame is submitted to the DFT [17].

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

where the Hamming window is used as a weighting function:

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

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

The values of k indexes correspond to frequencies:

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

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 ¿$n=1$¿N $\Lambda_n=[0,\pi]$ (see Figure 3) [18, 19].

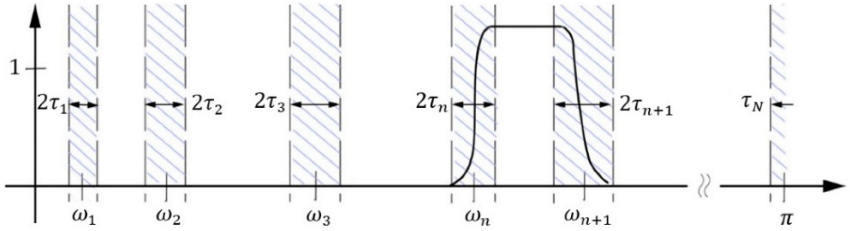

Figure 3: Fourier spectrum separation by adaptive low-pass $\phi_1(\omega)$ and bandpass $\psi_n(\omega)$ Meyer filters

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 [20]:

$$\gamma<min_n\frac{\omega_{n+1}-\omega_n}{\omega_{n+1}+\omega_n},0<\gamma<1$$

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 $\left\{\phi_1(\omega),\{\psi_n(\omega)\}_{n=1}^N\right\}$.

Then $\forall n>0$, the adaptive basis $\left\{\phi_1(\omega),\{\psi_n(\omega)\}_{n=1}^N\right\}$ 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 [21].

$$\hat{\phi}_n(\omega)=\begin{cases}1,\wedge|\omega|\le(1-\gamma)\omega_n\\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}(|\omega|-(1-\gamma)\omega_n)\right)\right],\wedge(1-\gamma)\omega_n\le|\omega|\le(1+\gamma)\omega_n\\ 0,\wedge otherwise\end{cases}$$

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

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

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

In practice, the following polynomial function is used [22]

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

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 detailed 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 [23–25].

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

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

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

$$f_0(t)=W_f^{\varepsilon}(0,t)\phi_1(t),$$
$$f_n(t)=W_f^{\varepsilon}(n,t)\psi_n(t),$$

and the orthogonality of the expansion is proved by the fact that [26–28]

$$f(t)=\sum_{n=0}^{N}f_n(t).$$

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 [29, 30].

With the help of HT EWT $x(t)$ we 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

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

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

$$\theta(t)=\arctan(y/x),$$

and the EWT $x(t)$ can be expressed as the real part of the following equation [31]

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

Then the Hilbert energy density spectrum is defined as

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

where the intervals $\Delta t\times\Delta\omega$ represent the values of $a^2(t)$ at a given time and frequency [32].

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

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

where x is the value of the coefficients before thresholding, y is the value of the coefficients after thresholding, and T is the threshold [33].

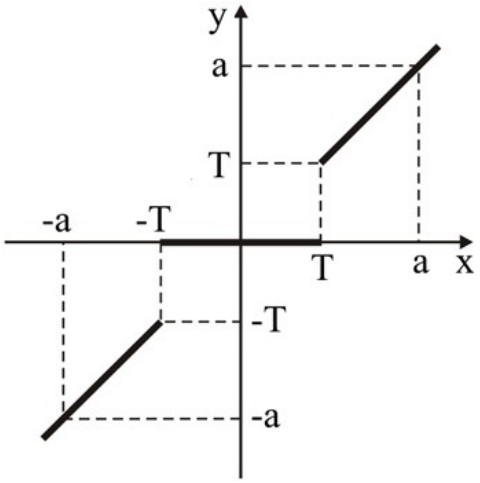

Figure 4: Threshold function

Let's assume that the probability of recognizing P frequency and amplitude of the function of the harmonic distribution law $x(t)=A\times\sin(\omega t+\varphi)$ is 1, and the function of the uniform distribution law

$$x(t)=\begin{cases} \dfrac{1}{b-a}, \wedge x \in [a,b] \\ 0, \wedge x \notin [a,b] \end{cases}$$

is 1/2 [34, 35].

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

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

where C is the Hilbert energy spectrum of length N, and $T=C_i$ [36–38].

4. Results and discussion

Figure 5 depicts calculated by the developed algorithm of experimental samples of voice commands of the control subject No. 1: "up", "down", "right", and "left".

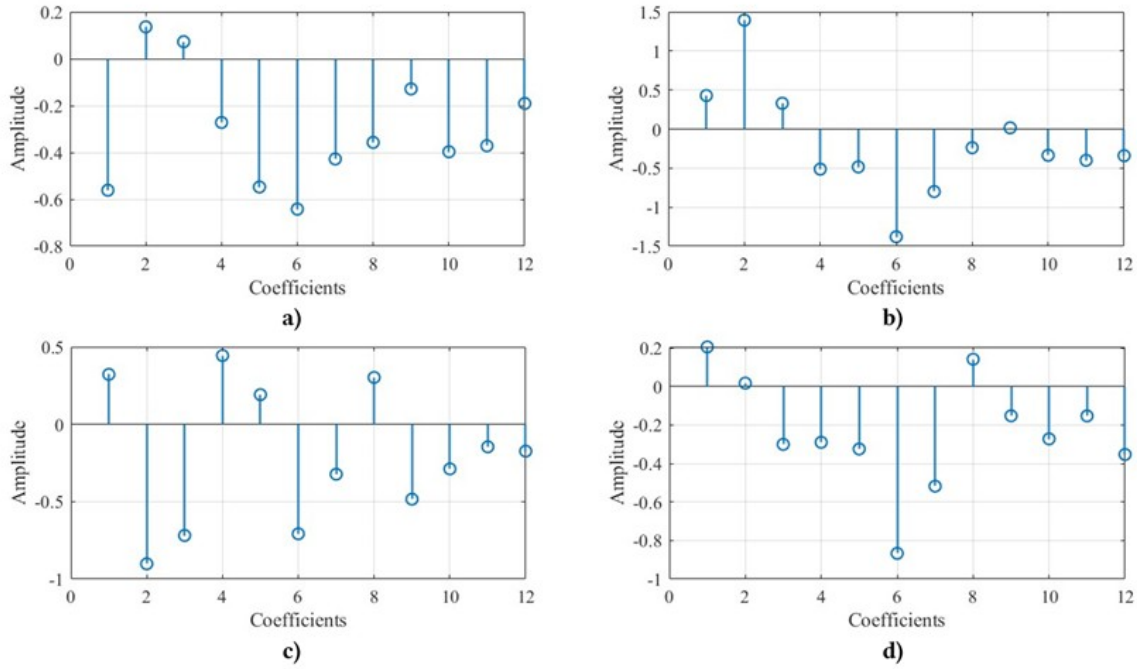

Figure 5: Recognition features based on the developed method of voice commands of control subject No. 1: (a) “up”, (b) “down”, (c) “right”, and (d) “left”.

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. 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 in 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:

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

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

The decision on biometric identification of voice commands is made according to the criterion of minimum variance, i.e., the smallest deviation of the compared recognition features based on the developed method in a certain recognition threshold which is given by the following expression:

if $D_{min} < \Theta$
identified !
else
not identified ! end

where, $D_{\min}$ is the minimum variance, $\Theta = 1 - \Delta$ is a given threshold of acceptable recognition (in practice, $\Delta = 0.80..0.90$ is usually used).

The minimum variance, which is within the specified threshold of acceptable recognition, is the best result of comparison, which means that the command is identified (recognized) is "identified!" Otherwise, the voice command fails biometric identification (is not recognized) and is "not identified!"

The paper details the obtained results of preliminary experimental research, based on which conclusions are drawn about the feasibility of further scientific and practical application of the system for recognizing voice control commands based on cepstral analysis and the developed algorithm for calculating the recognition features based on the developed method, as well as, a thorough justification of the scientific and technical significance of the conducted experimental research.

All scientific-experimental studies of the system of recognition of voice control commands set out below (Tables 1–3), were carried out taking into account the criterion of minimum distance, which is the variance of the difference between the mathematical expectations of the compared recognition features based on the developed method, depending on which varied values of the minimum variance D_{min}, thereby giving an objective assessment of the quality (reliability) of recognition of voice control commands in the testing mode of the system. The decision on biometric identification of voice commands is made by the criterion of minimum variance D_{min} in a given threshold of acceptable recognition $\Theta = 1 - \Delta = 0.15$, where $\Delta = 0.85$.

Table 1

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	0.0311	0.1921	0.2879	0.1479
down	0.5323	0.0648	0.7345	0.4284
right	0.3255	0.5048	0.0123	0.1699
left	0.1737	0.1935	0.1648	0.0112

In the first experiment (Table 1), we compared the recognition features based on the developed method of voice commands of control subject No. 1: "up", "down", "right", and "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 1) 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 $D_{min} = 0.0311$, "down" is $D_{min} = 0.0648$, "right" is $D_{min} = 0.0123$, "left" is $D_{min} = 0.0112$, based on this, the decision about positive biometric identification of the spoken voice commands is made (voice commands are recognized). In other cases (Table 1) it is 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 2
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	0.0451	0.3259	0.4090	0.1604
down	0.3770	0.0482	1.1258	0.5460
right	0.2055	0.4988	0.0967	0.1822
left	0.1268	0.3056	0.2764	0.0703

In the second experiment (Table 2), 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 2), 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 $D_{min} = 0.0451$, "down" is $D_{min} = 0.0482$, "right" is $D_{min} = 0.0967$, "left" is $D_{min} = 0.0703$, and therefore, a decision is made about the 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.

Table 3
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	0.0602	0.1657	0.4547	0.1943
down	0.4099	0.0912	1.1869	0.3772
right	0.2149	0.4521	0.0846	0.1784
left	0.1722	0.1946	0.2922	0.0785

In the third experiment (Table 3), 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 $D_{min} = 0.0602$, "down" is $D_{min} = 0.0912$, "right" is $D_{min} = 0.0846$, "left" is $D_{min} = 0.0785$, 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.

Conclusions

The paper 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.

The paper details the results of preliminary experimental studies, based on which conclusions are drawn about the feasibility of further scientific and practical application of the developed system for recognizing voice control commands based on the novelty of cepstral analysis and the algorithm for calculating the recognition features based on the developed method, as well as, justification of the scientific significance of the study.

A comparative evaluation of the calculated values obtained according to the chosen criterion of minimum distance, which is the main indicator of the quality criterion of voice command recognition, has been carried out.

In the first experiment (Table 1) we compared the Recognition features based on the developed method of voice commands of control subject No. 1: “up”, “down”, “right”, and “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.

From the obtained results (Table 1) 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 $D_{\min} = 0.0311$, “down” is $D_{\min} = 0.0648$, “right” is $D_{\min} = 0.0123$, “left” is $D_{\min} = 0.0112$, based on this, the decision about positive biometric identification of the spoken voice commands is made.

In the second experiment (Table 2), 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 2), 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 $D_{\min} = 0.0451$, “down” is $D_{\min} = 0.0482$, “right” is $D_{\min} = 0.0967$, “left” is $D_{\min} = 0.0703$, and therefore, a decision is made about the positive result of recognizing the spoken voice commands.

In the third experiment (Table 3), the recognition features based on the developed method of the spoken voice commands of the control subject No. 3 in the testing mode were compared with the recognition features based on the developed method of the voice commands of the control subject No. 1 spoken earlier in the system training mode. The obtained values of the comparison results: “up” is $D_{\min} = 0.0602$, “down” is $D_{\min} = 0.0912$, “right” is $D_{\min} = 0.0846$, “left” is $D_{\min} = 0.0785$, 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 a program mo delineating the proposed methods and algorithms for recognizing voice control commands in the MATLAB environment.

Declaration on Generative AI

While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] S. Kinkiri, S. Keates, Speaker identification: variations of a human voice, in: International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, 1–4. doi:10.1109/ICACCE49060.2020.9154998

[2] M. M. Abdulghani, W. L. Walters, K. H. Abed, Voice signature recognition for UAV pilots identity verification, in: International Conference on Computational Science and Computational Intelligence (CSCI), 2023, 125–129. doi:10.1109/CSCI62032.2023.00026

[3] A. Singhal, D. K. Sharma, Analysis of classifiers for gender identification using voice signals, in: 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, 1–4. doi:10.1109/ISCON52037.2021.9702469

[4] L. H. Palivela, V. Dharmalingam, P. Elangovan, Voice authentication system, in: International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), 2023, 1–6. doi:10.1109/ICDSAAI59313.2023.10452482

[5] M. Aliaskar, et al., Human voice identification based on the detection of fundamental harmonics, in: 7th International Energy Conference (ENERGYCON), 2022, 1–4. doi:10.1109/ENERGYCON53164.2022.9830471

[6] S. A. Jabbar, et al., Stable implementation of voice activity detector using zero-phase zero frequency resonator on FPGA, in: International Conference and Expo on Real Time Communications at IIT (RTC), 2023, 13–18. doi:10.1109/RTC58825.2023.10304243

[7] V. Kuzmin, et al., Method for correcting the mathematical model in case of empirical data asymmetry, in: Integrated Computer Technologies in Mechanical Engineering, ICTM 2022, Lecture Notes in Networks and Systems, vol. 657, 2023, 249–260. doi:10.1007/978-3-031-36201-9_21

[8] P. Jain, K. Gurugubelli, A. K. Vuppala, Study on the effect of emotional speech on language identification, in: National Conference on Communications (NCC), 2020, 1–6. doi:10.1109/NCC48643.2020.9056015

[9] A. Das, L. P. Roy, S. Kumar Das, Effectiveness of feature collaboration in speaker identification for voice biometrics, in: International Conference on Computer, Electrical & Communication Engineering (ICCECE), 2023, 1–4. doi:10.1109/ICCECE51049.2023.10085318

[10] N. J. Perdana, D. E. Herwindiati, N. H. Sarmin, Voice recognition system for user authentication using Gaussian mixture model, in: International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), 2022, 1–5. doi:10.1109/IICAIET55139.2022.9936856

[11] O. Lavrynenko, et al., A method for extracting the semantic features of speech signal recognition based on empirical wavelet transform, Radioelectron. Comput. Syst. 3(107) (2023) 101–124. doi:10.32620/reks.2023.3.09

[12] M. Barhoush, A. Hallawa, A. Schmeink, Robust automatic speaker identification system using shuffled MFCC features, in: International Conference on Machine Learning and Applied Network Technologies (ICMLANT), 2021, 1–6. doi:10.1109/ICMLANT53170.2021.9690530

[13] E. J. van Rensburg, R. A. Botha, B. Haskins, Identifying duress through voice during speaker authentication, in: International Conference on Electrical, Computer and Energy Technologies (ICECET), 2023, 1–5. doi:10.1109/ICECET58911.2023.10389204

[14] O. Lavrynenko, et al., A wavelet-based steganographic method for text hiding in an audio signal, Sensors 22(15) (2022) 5832. doi:10.3390/s22155832

[15] I. K. Alak, S. Ozaydin, Speech denoising with maximal overlap discrete wavelet transform, in: International Conference on Electrical and Computing Technologies and Applications (ICECTA), 2022, 27–30. doi:10.1109/ICECTA57148.2022.9990250
[16] Ravi, S. Taran, Emotion recognition using rational dilation wavelet transform for speech signal, in: 7th International Conference on Signal Processing and Communication (ICSC), 2021, 156–160. doi:10.1109/ICSC53193.2021.9673412
[17] G. Konakhovych, et al., Method of reliability increasing based on spare parts optimization for telecommunication equipment, in: 2nd International Workshop on Advances in Civil Aviation Systems Development, ACASD 2024, Lecture Notes in Networks and Systems, vol. 992, 2024, 296–309. doi:10.1007/978-3-031-60196-5_22
[18] S. Zhao, M. Fu, Optimization of audio signal denoising algorithm based on wavelet transform in speech communication scene, in: 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2023, 726–731. doi:10.1109/ICCASIT58768.2023.10351711
[19] O. Y. Lavrynenko, et al., Application of Daubechies wavelet analysis in problems of acoustic detection of UAVs, in: 6th Workshop for Young Scientists in Computer Science & Software Engineering, vol. 3662, 2024, 125–143.
[20] D. Pawade, et al., Voice based authentication using mel-frequency Cepstral coefficients and Gaussian Mixture Model, in: Bombay Section Signature Conference (IBSSC), 2022, 1–6. doi:10.1109/IBSSC56953.2022.10037421
[21] D. Bakhtiiarov, et al., Distribute load among concurrent servers, in: Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3826, 2024, 260–266.
[22] G. Bhatnagar, et al., System for identification of voice calls of interest in a telecom communication network, in: World Conference on Communication & Computing (WCONF), 2023, 1–6. doi:10.1109/WCONF58270.2023.10234983
[23] D. Bakhtiiarov, et al., Method of binary detection of small unmanned aerial vehicles, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654, 2024, 312–321.
[24] D. Dai, et al., A robust speech recognition algorithm based on improved PNCC and wavelet analysis, in: International Conference on Image Processing and Computer Vision (IPCV), 2023, 7–12. doi:10.1109/IPCV57033.2023.00008
[25] O. Lavrynenko, et al., Method of remote biometric identification of a person by voice based on wavelet packet transform, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654, 2024, 150–162.
[26] S. M. Kabir, et al., Vowel recognition for isolated digit using wavelet transform at decomposition level 3, in: 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), 2022, 1–4. doi:10.1109/STI56238.2022.10103316
[27] O. Holubnychyi, O. Lavrynenko, D. Bakhtiiarov, Well-adapted to bounded norms predictive model for aviation sensor systems, in: International Workshop on Advances in Civil Aviation Systems Development, ACASD 2023, Lecture Notes in Networks and Systems, vol. 736, 2023, 179–193. doi:10.1007/978-3-031-38082-2_14
[28] O. Julius, et al., Implementation of audio signals denoising for perfect speech-to-speech translation using principal component analysis, in: International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG), 2023, 1–6. doi:10.1109/SEB-SDG57117.2023.10124385
[29] O. Lavrynenko, et al., Remote voice user verification system for access to IoT services based on 5G technologies, in: 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2023, 1042–1048. doi:10.1109/IDAACS58523.2023.10348955
[30] G. Yang, Y. Song, J. Du, Speech signal denoising algorithm and simulation based on wavelet threshold, in: International Conference on Natural Language Processing (ICNLP), 2022, 304–309. doi:10.1109/ICNLP55136.2022.00055

[31] P. Warule, S. P. Mishra, S. Deb, Time-frequency analysis of speech signal using wavelet synchrosqueezing transform for automatic detection of Parkinson's disease, Sensors Letters 7(10) (2023) 1–4. doi:10.1109/LSENS.2023.3311670

[32] B. Zhao, et al., A spectrum adaptive segmentation empirical wavelet transform for noisy and nonstationary signal processing, Access 9 (2021) 106375–106386. doi:10.1109/ACCESS.2021.3099500

[33] O. Lavrynenko, et al., Method of semantic coding of speech signals based on empirical wavelet transform, in: 4th International Conference on Advanced Information and Communication Technologies (AICT), 2021, 18–22. doi:10.1109/AICT52120.2021.9628985

[34] T. Choudhary, V. Goyal, A. Bansal, WTASR: Wavelet transformer for automatic speech recognition of Indian languages, Big Data Min. Analytics 6(1) (2023) 85–91. doi:10.26599/BDMA.2022.9020017

[35] N. Holighaus, et al., Grid-based decimation for wavelet transforms with stably invertible implementation, transactions on audio, speech, and language processing 31 (2023) 789–801.

[36] R. Odarchenko, et al., Empirical wavelet transform in speech signal compression problems, in: 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), 2021, 599–602. doi:10.1109/PICST54195.2021.9772156

[37] K. Sun, et al., Wavelet denoising method based on improved threshold function, in: 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2022, 1402–1406. doi:10.1109/ITAIC54216.2022.9836698

[38] O. Lavrynenko, Method of speech signal scrambling based on matched wavelet filters, in: Cybersecurity Providing in Informationand Telecommunication Systems II, vol. 3826, 2024, 229–235.