UDC 621.391(045)
DOI:10.18372/1990-5548.81.19016

O. Yu. Lavrynenko

VOICE CONTROL SYSTEM FOR ROBOTICS IN A NOISY ENVIRONMENT

Department of Telecommunication and Radio Electronic Systems, Faculty of Air Navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine
E-mails: [1]oleksandrlavrynenko@gmail.com ORCID 0000-0002-7738-161X

***Abstract*—*This paper analyzes the effectiveness of the developed voice control system for robotics based on MFCC and GMM-SVM under the influence of interference in the communication channel. The system allows characterizing individual features of speech signals with their subsequent classification and making a reliable decision on the interpretation and execution of voice commands by robotic equipment. The proposed voice control system for robotics based on MFCC and GMM-SVM is implemented using the following technologies: 1) selection of active speech areas by calculating the short-term energy and the number of zero crossings between adjacent frames of the speech signal; 2) adaptive wavelet filtering of the speech signal, where it is necessary to generate threshold values, which will reduce the impact of additive noise; 3) selection of recognition features, which are used as mel-frequency cepstral coefficients; 4) classification of recognition features based on mixtures of Gaussian distributions and the support vector method using the linear Campbell kernel and the principal component method with a projection on latent structures, which will reduce errors of the 1st and 2nd kind.*

Index Terms—Speech signals; voice control; adaptive wavelet filtering; mel-frequency cepstral coefficients; mixtures of Gaussian distributions; support vector method; communication channel; nonlinear distortion coefficient.

I. INTRODUCTION

Today, robotics is one of the most promising areas of science and technology development. Robots no longer only perform simple actions but can also replace humans in complex and dangerous tasks. However, in order for robots to perform their tasks as efficiently as possible, it is necessary to provide a convenient and intuitive control interface [1].

For humans, the most natural way to interact with robotic technology is through speech, namely through voice commands, which is one of the most common interfaces for controlling robots. Their use provides quick and easy access to robot control without the need for physical contact with the control device. Currently, voice interfaces are used in various fields, including manufacturing, medicine, education, and home use [2].

As technology advances, the use of voice-activated robots will expand, improving efficiency, safety, and quality of life in many sectors. Thus, it is expected that the future of robotics will become more promising with the advent of human interaction with robotic technology through speech signals [3].

Although the achievements in the field of voice-activated robotics are impressive, a number of scientific and technical challenges still need to be addressed, namely, to develop a voice control system for robotics that could provide a high percentage of error-free recognition and classification of speech signals under the influence of external noise and interference in the communication channel and a small delay in processing and transmitting information for fast real-time operation of the system. These aspects require the development and implementation of modern methods for: 1) processing, encoding and recognition of speech signals [4]; 2) adaptive filtering [5]; 3) classification using machine learning algorithms and neural networks; taking into account the balance between the mathematical complexity of calculations and the speed of this type of system, which will affect the overall performance of the system in a real environment [6].

It is with these aspects in mind that the authors of this article have proposed a balanced voice control system for robotics in a noisy environment, which will be explained in detail in Section 3.

II. LITERATURE REVIEW AND PROBLEM STATEMENT

Despite their widespread use and the above advantages, existing voice control systems for robotics have a number of serious drawbacks, which this research article aims to analyze [7].

These include, first of all, the low resolution of speech signal recognition methods and a significant percentage of errors of both the first kind (mistakenly rejected voice commands that have a negative

classification result but are authentic) and the most dangerous second kind (voice commands that are mistakenly considered correctly classified but are not authentic). The situation is especially complicated by the recognition and classification of speech signals in real conditions, accompanied by a set of unfavorable external factors that will directly affect the efficiency of the robotics voice control system [8].

A voice control system for robotics operating in real-world conditions faces the following serious challenges. Firstly, this classification of speech signals causes all sorts of hardware distortions and interference due to the peculiarities of equipment and devices for recording, processing and storing information. Secondly, external acoustic noise inevitably superimposes on the speech signal, which can significantly distort individual informative characteristics. In view of this, voice control systems for robotics, which have demonstrated quite high efficiency in laboratory conditions, may show much lower reliability when analyzing speech information with external noise. Finally, in a number of tasks, classification has to be performed under very difficult conditions of overlapping voices of several speakers, in particular, with similar acoustic characteristics. It should be noted that there have been virtually no studies of speech signal classification capabilities for this most difficult case [9].

Typically, distortion of speech signals is associated with the speaker, environmental noise, distortion of the microphone system (including electromagnetic interference), distortion arising in the recording channel and in the communication channel during the transmission of the speech signal, as well as distortion during its processing by special software (Fig. 1) [10].

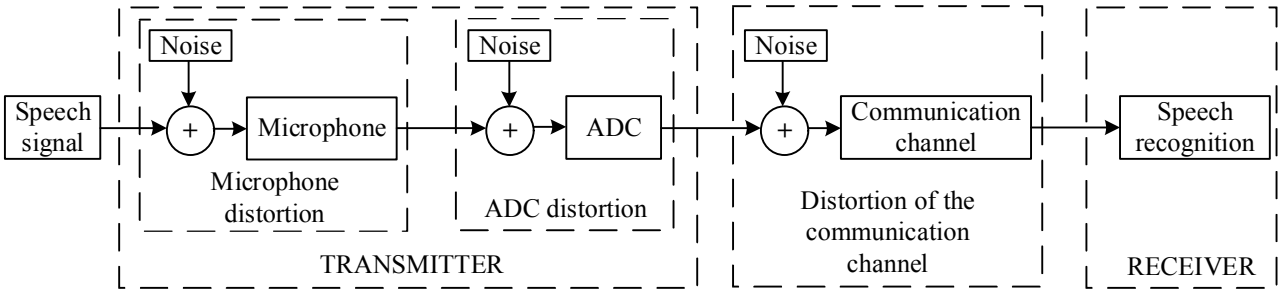

Fig. 1. The effect of noise and interference on different parts of the speech signal

Thus, speech signal processing involves a set of technical, algorithmic, and mathematical methods that cover all stages, from voice recording to voice data classification. The discussed difficulties and shortcomings lead to the conclusion that further development of voice control systems for robotics requires the development of new approaches aimed at processing large amounts of experimental data, their effective analysis, and reliable classification. This indicates the relevance of research on the development of new mathematical methods for processing, analyzing and classifying speech signals that would ensure the reliability and accuracy of voice command classification under the influence of noise and interference in the communication channel of information and telecommunication networks. This study is aimed at analyzing and solving the above scientific problems [11].

III. PROPOSED SYSTEM

The proposed voice control system for robotics has two modes of operation: learning mode and recognition mode. These modes are included in the block diagram of the voice control system for robotics (Fig. 2), whose task is to perform the following steps: 1) dividing the speech signal into time frames and selecting areas of active speech with finding the values of the change in short-term energy and the number of zero crossings between adjacent frames of the speech signal (Short-Time Energy and Zero-Crossing Rate, STE-ZCR) [12]; 2) adaptive wavelet filtering of the speech signal (Adaptive Wavelet Thresholding, AWT). Adaptive Wavelet Thresholding (AWT) to solve the problem of noise removal, where it is necessary to conduct adaptive generation of microlocal thresholds, which will reduce the impact of additive noise on the pure form of the speech signal and the selection of recognition features, where mel-frequency cepstral coefficients (MFCC) are used as informative recognition features in the voice control of robotics [13]. The classification of MFCC recognition features is based on mixtures of Gaussian distributions and the Gaussian Mixture Model and Support Vector Machine (GMM-SVM) using the Campbell linear kernel and the principal component method with a projection on latent structures, which together will increase the reliability of classification, which is manifested in the reduction of errors of the 1st and 2nd kind [14].

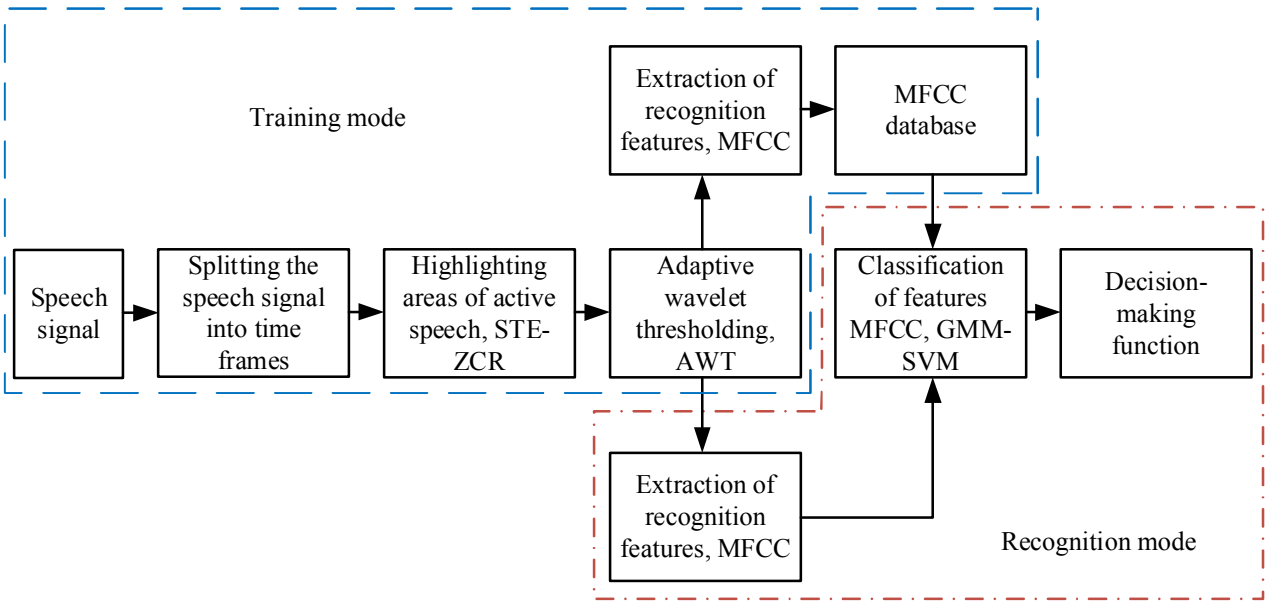

Fig. 2. Block diagram of the robotics voice control system based on MFCC and GMM-SVM

The frame duration of a speech signal should be small enough to allow the sequence of frames to accurately reflect the short-term dynamics of its change, and large enough to allow the sequence of frames to accurately reflect its long-term dynamics. According to the conditions for selecting the frame duration of a speech signal, its frame duration must be not less than the period of the fundamental tone $T_{FT} = 1/f_{\text{FT}} = 10$ ms, where $f_{\text{FT}} \geq 100$ Hz is the frequency of the fundamental tone [15].

The next step is to consider the algorithm for dividing the speech signal into vocalized and non-vocalized segments and segments of silence (pause). This algorithm is based on the assumption that the speech signal is a non-stationary process with significant changes in the short-term energy and the number of zero-crossings between adjacent frames (Short-Time Energy and Zero-Crossing Rate, STE-ZCR) [16].

The algorithm contains 7 blocks.

Block 1. Input speech signal $x(m)$, $m = \overline{0, N-1}$.

Block 2. Splitting the speech signal into 16 ms frames.

Block 3. Calculation of the values of short-term energy E_n and the number of zero crossings Z_n of the nth frame. For example, the short-term energy is equal to $E_n = \sum_{m=n-N+1}^{n} x^2(m)$, where n is the frame number; $n = \overline{0, L}$; L is the number of frames; $M = LN$ is the number of samples of the speech signal.

The short-term function of the average number of zero crossings, or zero intersections, is to compare the signs of neighboring counts. For example,

$$z_n = \sum_{m=-\infty}^{\infty} |\operatorname{sgn}(x(m)) - \operatorname{sgn}(x(m-1))|\, \omega(n-m),$$

where $W(m) = \begin{cases} 1/2, & 0 \leq m \leq N-1, \\ 0, & \end{cases}$ and

$\operatorname{sgn}(X(m)) = \begin{cases} 1, & X(m) > 0, \\ -1, & X(m) < 0, \end{cases}$ is a sign function.

Blocks 4, 6. Setting thresholds E_{thr} and Z_{thr} for E_n and Z_n.

Block 5. Checking the fulfillment of the condition $E_n < E_{thr}$?: yes is the nth frame belongs to the silence segment; no is to block 7.

Block 7. Check if the condition $Z_n < Z_{thr}$? is met: yes is the nth frame belongs to a vocalized segment; no is the nth frame belongs to a non-vocalized segment.

To reduce errors in deciding whether an area is vocalized, it is proposed to use the following ratio

$$R_{rms} = \frac{E_{rms}}{Z_n},$$

where $E_{rms} = \sqrt{\overline{x^2}(m)} = \sqrt{\frac{1}{N}\sum_{m=1}^{N} x^2(m)}$ is the root mean square value of the speech signal.

Vocalized speech is characterized by a large E_{rms} and a small Z_n, and non-vocalized speech is characterized by a small E_{rms} and a large Z_n, so it is fair to say that R_{rms} is large for a vocalized frame and small for a non-vocalized frame.

The next step is to perform Adaptive Wavelet Thresholding (AWT) on the speech signal to eliminate the influence of noise on speech recognition. In this case, a set of coupled mirror filters decomposes the signal in a discrete domain according to an orthogonal wavelet basis $\{\psi_{j,m}\}$ into several frequency bands [17].

Let us represent the model of the speech signal $f(t)$ distorted by additive noise as

$$X(t) = f(t) + \eta(t).$$

Then, when such a signal is decomposed by a set of conjugate mirror filters on some discrete orthogonal basis $\{\psi_m\}$, gives:

$$WX[m] = Wf[m] + W\eta[m].$$

Let's introduce a linear operator D that evaluates $Wf[m]$ by $WX[m]$ using the function $d_m(x)$. The resulting evaluation is

$$\widetilde{F} = DX = \sum_{m=0}^{N-1} d_m(WX[m])\psi_m.$$

When $d_m(x)$ is a threshold function, the risk of this assessment can be minimized.

Let $r_t(x,T)$ be the risk of a threshold estimate computed with threshold T. Then the estimate $\tilde{r}_t(x,T)$ of the risk $r_t(x,T)$ should be calculated from the speech signal $X(t)$, which is distorted by noise. The value of the threshold T in this case is optimized by minimizing $\tilde{r}_t(x,T)$.

To find the value of $\tilde{T}$ that minimizes the estimate of $\tilde{r}_t(x,T)$, N the coefficients of the data $WX[m]$ are sorted by decreasing amplitude. Then, the wavelet decomposition coefficients ranked in this way form an ordered set $\{WX^r[k]\}_{1\le k\le N}$, where any $WX^r[k] = WX[m_k]$ is the corresponding coefficient of the rank k: $|WX^r[k]| \ge |WX^r[k+1]|$ [18].

Let l be some index such that $|WX^r[l]| \le T < |WX^r[l-1]|$, then we can assume

$$\tilde{r}_t(f,T) = \sum_{k=1}^{N} |WX^r[k]|^2 - (N-l)\sigma^2 + l(\sigma^2 + T^2),$$

where σ^2 is the variance of the noise component.

Then to minimize $\tilde{r}_t(x,T)$, you must choose $T = |WX^r[l]|$.

At the next stage of the operation of the voice control system for robotics based on MFCC and GMM-SVM, it is necessary to consider the algorithm for finding MFCC [19]:

1) The signal $s[t]$, is divided into K frames by N counts, which intersect by 1/2 the frame length: $s[t] \to S_n[t]$.

2) A discrete Fourier transform is performed in each frame: $\{\mathrm{Re}\, X_n[k],\ \mathrm{Im}\, X_n[k]\} = FFT(S_n[i])$, where $k = 1,\ldots\ldots, M$, $M = N/2$.

3) Find the power spectral density of the speech signal:

$$P_n[k] = A_n[k]^2,\ A_n[k] = \sqrt{\mathrm{Re}\, X_n[k]^2 + \mathrm{Im}\, X_n[k]^2}.$$

4) Multiply the samples of the speech signal power spectral density by the generated triangular filter bank [20] and taking the logarithm of the power spectral density of the speech signal:

$$X_n[i] = \ln\left(\sum_{k=1}^{M} P_n[k]H_i[k]\right),\ i = 1,\ldots\ldots,P.$$

5) Perform a discrete cosine transform to the logarithmic energy of the speech signal spectrum:

$$C_n[j] = \sum_{k=1}^{P} X_n[k]\cos\left(j\left(k-\frac{1}{2}\right)\frac{\pi}{P}\right),$$
$$i = 1,\ldots\ldots,P,\ j = 1,\ldots\ldots,J,$$

where $C_n[j]$ is the MFCC array; J is the desired number of coefficients ($J < P$).

The next step is to classify the MFCC speech signal features based on GMM-SVM.

For the input vector $\vec{x}$, the density of the Gaussian mixture is the weighted sum of M components of the mixture, and it is given by expression [21]:

$$p_i(\vec{x}\,|\,\lambda) = \sum_{i=1}^{M} \alpha_i p_i(\vec{x}),$$

where $\vec{x}$ is an N-dimensional random vector; $p_i(\vec{x}),\ i = 1,\ldots,M$ are the components of the mixture and α_i, $\{i = 1,\ldots..,M\}$ are the weights of the mixture, and each density component is a Gaussian function.

The weights of the components of the mixture satisfy the bond: $\sum_{i=1}^{M} \alpha_i = 1$.

The GMM is parameterized by a set of parameters defined for each i component of the mixture: mean vectors $\vec{\mu}_i$, covariance matrices $\sum_i$

and weights α_i. These parameters are all represented by a notation system:

$$\lambda = \{\alpha_i, \vec{\mu}_i, \Sigma_i\},\ i = 1, ..., M\ .$$

The goal of training the GMM model is to obtain GMM parameters λ that give a better fit to the experimental distribution of the training vectors $X = \{\vec{x}_1, ..., \vec{x}_T\}$.

Consider the problem of classification in the plane of two non-overlapping classes, in which objects are described by n-dimensional real vectors: $X \in R^N$, $Y \in \{-1, +1\}$.

Then we define a linear threshold classifier [22]:

$$Y(x) = \text{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \text{sign}\left(\langle w, x\rangle - w_0\right),$$

where $x = (x_1, \ldots, x_n)$ is the feature description of the object X; vector $w = (w_1, ..., w_n) \in R^n$ is the scalar threshold $w_0 \in R^n$. Equation $Y(x)$ describes the hyperplane that separates classes in space R^n $\langle w, x\rangle = w_0$.

If we assume that the feature descriptions of objects are vectors $\psi(x_i)$ rather than vectors x_i, then SVM construction is performed in much the same way as before. The only difference is that the scalar product $< x, x' >$ in the space X is replaced by the scalar product $< \psi(x), \psi(x') >$ in the space H.

This means that when building an SVM, the scalar product $< x, x' >$ can be formally replaced by the kernel $K(x, x')$. Since the kernel is generally nonlinear, this replacement leads to a significant expansion of the class of valid algorithms a: $X \to Y$ [23].

For example, the Campbell linear kernel is often used for GMM-SVM speech signal classification systems:

$$K_{lin}(s^a, s^b) = \sum_{i=1}^{N}\left(\sqrt{w_i}\sum_i^{-\frac{1}{2}}\mu_i^a\right)\left(\sqrt{w_i}\sum_i^{-\frac{1}{2}}\mu_i^b\right)^t.$$

IV. Research Results

To mathematically model the distortion of the speech signal in the communication channel, we applied an oversampling algorithm based on the use of the discrete Fourier transform [24].

Let the input speech signal be characterized by a finite number of samples $a(n)$. At the first step of the algorithm, the coefficients $A(k)$ of the direct Fourier transform were calculated:

$$A(k) = \sum_{n=1}^{N} a(n) \cdot e^{-j2\pi\frac{k}{N}n},\ k = 1, 2, ..., N\ .$$

In the second step, zero components were inserted into the area near the sample number $N/2$ of the spectrum, the number of which was set by the values of the initial number of samples N and the number of samples in the resampled signal M.

The coefficients $H(i)$ of the resampled spectrum were determined by the following formula:

$$\begin{cases} H(i) = A(i),\ 1 \le i \le \dfrac{N+1}{2}, \\ H(i) = 0,\ \dfrac{N+1}{2} + 1 \le i \le \dfrac{N+1}{2} + M - N, \\ H(i) = A(i - M + N),\ \dfrac{N+1}{2} + M - N \le i \le M. \end{cases}$$

The final step of the algorithm calculated the $h(m)$ counts of the inverse discrete Fourier transform with normalization:

$$h(m) = \frac{1}{M}\sum_{k=1}^{M} H(k) \cdot e^{-j2\pi\frac{k}{M}m},\ m = 1, 2, ..., M.$$

The nonlinear distortion coefficient K [25] is used as a value that quantitatively characterizes distortion:

$$K = \sqrt{\frac{1}{L}\sum_{l=1}^{L} H^2(l)} \Big/ \sqrt{\frac{1}{N}\sum_{k=1}^{N} A^2(k)}\ ,$$

where $H(l)$ is the spectral components of the output speech signal that are not present in the spectrum of the input speech signal $A(k)$, L is the number of spectral components $H(l)$.

In Figure 3 illustrates a segment of the frequency spectrum of the input speech signal and the corresponding frequency spectrum of the distorted speech signal ($K = 0.8$) for the frequency range from $f = 0$ Hz to $f = 4000$ Hz, for which the distortion was most noticeable.

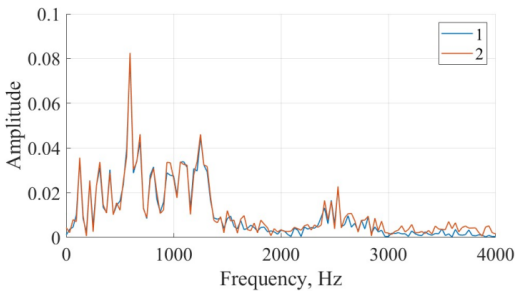

Fig. 3. Segments of frequency spectra of speech signals: 1 is spectrum of the input speech signal; 2 is spectrum of the distorted speech signal

The data obtained in the calculations were presented in the form of classification graphs in the space of the first principal components (PCs), which allow us to clearly interpret the effectiveness of the developed voice control system for robotics.

As a graphical illustration, Fig. 4 shows segments of resampled spectra that were subjected to distortions corresponding to the coefficients $K = 0.2;\ 0.4;\ 0.6;\ 0.8$.

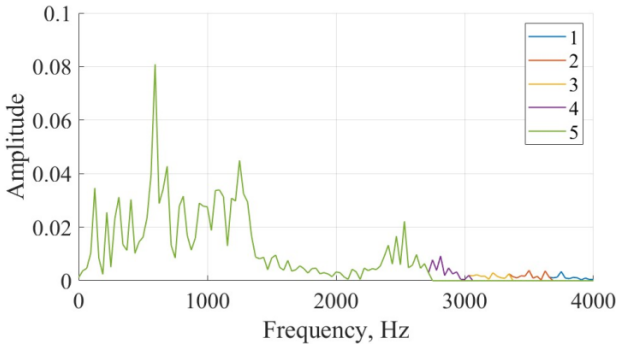

Fig. 4. Segments of resampled speech signal spectra: 1 is $K = 0$; 2 is $K = 0.2$; 3 is $K = 0.4$; 4 is $K = 0.6$; 5 is $K = 0.8$

The graph of classification performance indicators calculated for the input and distorted speech signals is shown in Fig. 5. Here, the input speech signals are not represented by filled red dots, and the distorted speech signals with different values of the distortion coefficient K are represented by green filled dots (point 1 is $K = 0.2$; 2 is $K = 0.4$; 3 is $K = 0.6$; 4 is $K = 0.8$).

Figure 5 shows that at distortion ratios of $K = 0.2$ and $K = 0.4$ (points 1 and 2), the speech signal was classified correctly. At the coefficients $K = 0.6$ and $K = 0.8$ (points 3 and 4, located outside the ellipse), the positive classification was no longer achieved.

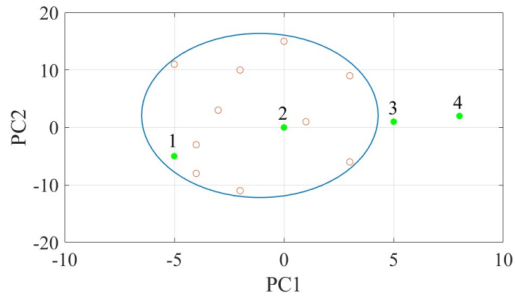

Fig. 5. The effect of distortion in the communication channel on the classification of speech signals

V. CONCLUSIONS

The paper analyzes the effectiveness of the developed voice control system for robotics based on MFCC and GMM-SVM under the influence of interference in the communication channel. The system allows characterizing individual features of speech signals with their subsequent classification and making a reliable decision on the interpretation and execution of voice commands by robotic equipment.

The proposed voice control system for robotics based on MFCC and GMM-SVM is implemented using the following technologies: 1) selection of active speech areas with finding the values of the change in short-term energy and the number of zero crossings between adjacent frames of the speech signal; 2) adaptive wavelet filtering of the speech signal to solve the problem of noise removal, where it is necessary to conduct adaptive generation of microlocal thresholds, which will reduce the effect of additive noise on the pure form of the speech signal; 3) identification of recognition features, where fine-frequency cepstral coefficients are used as informative features of speech signal recognition in robotics voice control; 4) classification of recognition features based on mixtures of Gaussian distributions and the support vector method using the linear Campbell kernel and the principal component method with a projection on latent structures, which will increase the reliability of identification, which is manifested in the reduction of errors of the 1st and 2nd kind.

The influence of the type and magnitude of external noise, various interferences and distortions on the recognition of speech signals transmitted through communication channels of information and telecommunication networks for the tasks of voice control of robotics is investigated. A methodology is proposed that allows classification of speech signals under noise by mathematical modeling of distortions through the use of a subsampling algorithm. This approach is based on the use of a discrete Fourier transform and allows increasing the sampling rate of a speech signal by a given integer or fractional number of times, where the nonlinear distortion coefficient is used as a value that quantitatively characterizes the distortion. The mathematical modeling of speech signal distortion made it possible to quantify the magnitude of these distortions, at which the correct classification is possible. This shows that the proposed approach to assessing the effects of distortion can be used to analyze the reliability of voice control systems for robotics.

Thus, the systematic study made it possible to identify the effect of external noise on the efficiency of the developed voice control system for robotics based on MFCC and GMM-SVM under the influence of interference in the communication

channel, which can be used in the development and testing of remote voice interface systems in information and telecommunication networks.

REFERENCES

[1] P. Liu *et al*., "Design of Bionic Robot Based on Voice Remote Control," *2023 42nd Chinese Control Conference (CCC)*, Tianjin, China, 2023, pp. 4226–4231, https://doi.org/10.23919/CCC58697.2023.10240762

[2] 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

[3] O. Lavrynenko, G. Konakhovych and D. Bakhtiiarov, "Method of voice control functions of the UAV," *2016 4th International Conference on Methods and Systems of Navigation and Motion Control (MSNMC)*, Kiev, Ukraine, 2016, pp. 47–50, https://doi.org/10.1109/MSNMC.2016.7783103

[4] L. Y. Yong, S. Gobee and V. Durairajah, "An Interactive System to Control a Humanoid Robot using Vision and Voice," *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Dharan, Nepal, 2022, pp. 895–898, https://doi.org/10.1109/I-SMAC55078.2022.9987307

[5] 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

[6] 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

[7] M. Norda, C. Engel, J. Rennies, J. -E. Appell, S. C. Lange and A. Hahn, "Evaluating the Efficiency of Voice Control as Human Machine Interface in Production," in *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 4817–4828, July 2024, https://doi.org/10.1109/TASE.2023.3302951

[8] 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

[9] M. M and V. R S, "A Review on Quality Speech Recognition Under Noisy Environment," *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2023, pp. 545–548, https://doi.org/10.1109/ICACCS57279.2023.10112783

[10] 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

[11] 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

[12] 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.

[13] T. Kim, J. Chang and J. H. Ko, "ADA-VAD: Unpaired Adversarial Domain Adaptation for Noise-Robust Voice Activity Detection," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, Singapore, 2022, pp. 7327–7331, https://doi.org/10.1109/ICASSP43922.2022.9746755

[14] 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.

[15] 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.

[16] M. Bächle, M. Schambach and F. Puente León., "Signal-Adapted Analytic Wavelet Packets in Arbitrary Dimensions," *2020 28th European Signal Processing Conference (EUSIPCO)*, Amsterdam, Netherlands, 2021, pp. 2230–2234, https://doi.org/10.23919/Eusipco47968.2020.9287575

[17] 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

[18] R. Odarchenko, O. Lavrynenko, D. Bakhtiiarov, S. Dorozhynskyi and V. A. O. Zharova, "Empirical Wavelet Transform in Speech Signal Compression Problems," *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 2021, pp. 599–602, https://doi.org/10.1109/PICST54195.2021.9772156

[19] 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

[20] 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

[21] O. Lavrynenko, A. Pinchuk, H. Martyniuk, A. Fesenko, S. Yarotsky and M. Aleksander, "Remote Voice User Verification System for Access to IoT Services Based on 5G Technologies," *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 2023, pp. 1042–1048, https://doi.org/10.1109/IDAACS58523.2023.10348955

[22] 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

[23] 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

[24] 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

[25] O. Yu. Lavrynenko, D. I. Bakhtiiarov, B. S. Chumachenko, O. G. Holubnychyi, G. F. Konakhovych and V. V. Antonov, "Application of Daubechies wavelet analysis in problems of acoustic detection of UAVs," *CEUR Workshop Proceedings*, 2024, vol. 3662, pp. 125–143.

Received August 05, 2024

Lavrynenko Oleksandr. ORCID 0000-0002-7738-161X. PhD in Engineering. Associate Professor. Associate Professor of the Department of Telecommunication and Radio Electronic Systems.
Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, 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 30.
E-mail: oleksandrlavrynenko@gmail.com

О. Ю. Лавриненко. Система голосового управління робототехнікою в шумовому середовищі

У роботі проведено аналіз ефективності розробленої системи голосового управління робототехнікою на основі MFCC і GMM-SVM в умовах впливу завад у каналі зв'язку. Система дає змогу характеризувати індивідуальні особливості мовних сигналів із подальшою їхньою класифікацією та ухваленням достовірного рішення щодо інтерпретації та виконання голосових команд роботизованою технікою. Запропоновану систему голосового управління робототехнікою на основі MFCC і GMM-SVM реалізовано за допомогою таких технологій: 1) виділення ділянок активної мови за допомогою розрахунку короткочасної енергії та кількості перетинів нуля між суміжними кадрами мовного сигналу; 2) адаптивна вейвлет-фільтрації мовного сигналу, де необхідно провести генерацію порогових значень, що дасть змогу зменшити вплив адитивного шуму; 3) виділення ознак розпізнавання, в якості яких використовуються мел-частотні кепстральні коефіцієнти; 4) класифікація ознак розпізнавання на основі сумішей Гауссових розподілів та методу опорних векторів з використанням лінійного ядра Кампбелла та методу головних компонент з проекцією на латентні структури, що забезпечить зменшення помилок 1-го та 2-го роду.

Ключові слова: мовні сигнали; голосове управління; адаптивна вейвлет-фільтрація; мел-частотні кепстральні коефіцієнти; суміші Гаусових розподілів; метод опорних векторів; канал зв'язку; коефіцієнт нелінійних спотворень.

Лавриненко Олександр Юрійович. ORCID 0000-0002-7738-161X. Кандидат технічних наук. Доцент. Доцент кафедри телекомунікаційних та радіоелектронних систем.
Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, 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 30.
E-mail: oleksandrlavrynenko@gmail.com