TELECOMMUNICATIONS AND RADIO ENGINEERING

UDC 621.391 (045)
DOI:10.18372/1990-5548.63.14529

[1]**O. Y. Lavrynenko,**
[2]**G. F. Konakhovych,**
[3]**D. I. Bakhtiiarov**

PROTECTED VOICE CONTROL SYSTEM OF UNMANNED AERIAL VEHICLE

[1,2,3]Faculty of Air Navigation, Electronics and Telecommunications,
National Aviation University, Kyiv, Ukraine
E-mails: [1]oleksandrlavrynenko@gmail.com, [2,3]tks@nau.edu.ua

Abstract—*A system for recognizing steganographic-transformed voice commands for controlling an unmanned aerial vehicle based on cepstral analysis has been developed, which allows efficient recognition and covert transmission of commands to an unmanned aerial vehicle by converting voice control commands into a kind of steganographic feature vector, which implies information hiding voice control of an unmanned aerial vehicle. A mathematical model of the algorithm for calculating small-frequency cepstral coefficients and a classifier for recognizing voice control commands has been synthesized to solve the problem of semantic identification and ensure the secrecy of control information of an unmanned aerial vehicle in a communication channel. The detailed results of preliminary experimental studies of the developed voice command recognition system and the computational algorithm of mel-frequency cepstral coefficients in MATLAB in the case of the identification of control commands "up", "down", "right", "left" spoken by different subjects.*

Index Terms—Steganographic voice transform; voice control commands; voice command recognition; cepstral analysis; cepstral coefficients; Fourier transform; wavelet transform; cosine transform; voice control system; unmanned aerial vehicle.

I. INTRODUCTION

Currently, there is an active development of the defense industry of Ukraine, one of the main and main tasks of which is the introduction of radically new high-tech means in the field of military communications, namely, the creation of a secure voice-activated radio control system for unmanned aerial vehicle (UAV) functions to solve the tasks in the military intelligence purposes. In this regard, the need for alternative, more natural new UAV control methods is becoming increasingly relevant. One of the most natural control methods for humans is through voice commands.

II. PROBLEM STATEMENT

We formulate the hypothesis of this research. Since, the calculation of the Daubechies wavelet filter involves finding the coefficients of the polynomial

$$P_k = \frac{\prod_{i=-N+1}^{N}\left(\frac{1}{2}-i\right)}{\prod_{i=-N+1}^{N}(k-i)}, \quad k = 1, \ldots, N$$

which for all values $k \neq i$ form a vector

$$P = \begin{pmatrix} P_N & 0 & P_{N-1} & 0 & \ldots & 0 & P_1 & 1 & P_1 & 0 & P_2 & 0 & \ldots & 0 & P_N \end{pmatrix},$$

length of $4N-1$.

Then, the coefficients of the low-frequency wavelet filter are as follows

$$R = \sqrt{2}\begin{pmatrix} P_1 & \ldots & P_{2N} \end{pmatrix},$$

$$D = \begin{pmatrix} R_{2N} & \ldots & R_1 \end{pmatrix},$$

length of $2N$, and the high-frequency coefficients are determined by calculating the quadrature-mirror filter as follows

$$W = \begin{pmatrix} R_{2N} & -R_{2N-1} & R_{2N-2} & -R_{2N-3} & \ldots & -R_4 & R_3 & -R_2 & R_1 \end{pmatrix}, \qquad V = \begin{pmatrix} W_{2N} & \ldots & W_1 \end{pmatrix}.$$

As a result, using the operation of mathematical convolution of the values of the speech signal X with the values of the coefficients of the wavelet filter low and high frequencies D and V with subsequent gamming, it will allow us to highlight the semantic features of speech, and also to intentionally distort speech semantics with the possibility of reverse recovery in order to prevent unauthorized access to UAV control. The mathematical formalization of the above statement is determined by the following expressions

$$Z_k = F_i\left(\sum_{j=\max(1,\,k+1-2N)}^{\min(k,\,L)} \mathrm{X}_j D_i\right),$$

$$Y_k = F_i\left(\sum_{j=\max(1,\,k+1-2N)}^{\min(k,\,L)} \mathrm{X}_j V_i\right),$$

where F_i is the gamming function, at $k = 1,\ \ldots,\ L + 2N - 1$, $i = k + 1 - j$. Thus, the found vectors of the values of the coefficients Z and Y after double thinning $\downarrow 2$ will be the encrypted semantic component of the speech signal X, which plays a fundamental role in UAV voice control systems. Confirmation of the hypothesis put forward will allow for the protection of control information in the UAV communication channel on one mathematical apparatus, and also using a sufficiently large level of decomposition of the wavelet transform, we will gain a gain in the recognition efficiency of voice commands for UAV control.

III. Analysis of Existing Research and Publications

There are studies conducted in recent years in the field of speech recognition and voice control of UAVs, the results of which were reviewed and analyzed in the publications [1] – [4]. The publications have been developed voice control systems of the UAV, which significantly increases the efficiency of speech recognition, where the recognition coefficient is 97% in a system based on neural networks [1], 96% in a system based on hidden Markov models [2], 92% in a system based on **cepstral** coefficients with subsequent classification by the least square method [3], 98% in a system based on wavelet transform using neural networks [4]. In the presented UAV voice control systems [1] – [4], the problem of ensuring UAV secure control via a radio communication channel using voice control commands is not solved, which does not allow protecting control information from unauthorized access by intruders in order to intercept UAV control. The problem appears then, as soon as we try to integrate some kind of encryption algorithm, immediately exceeds the boundary value of the delay of more than 300 ms, due to the total delay of speech recognition and encryption algorithms, which usually increases the delay by two, i.e. about 600 ms, which does not allow the use of this system in UAV control in real time. In connection with this problem, it was decided to develop a system that would work at a recognition coefficient of about 97% with data encryption in the communication channel with a total delay of processing and data transfer of not more than 300 ms. The next important point that this work differs from existing is the merging of two methods for processing speech signals, namely, it is the MFCC and WT in order to increase the recognition efficiency of the semantic features of voice commands and on the basis of the same mathematical apparatus to provide protection without resorting to different encryption algorithms.

IV. Protected Voice Control System of UAV

The paper presents the developed system for the secure voice control of UAV based on WT using MFCC as semantic features of speech control commands (Fig. 1).

This scientific article discusses an approach to solving the problem of recognition of voice control commands, using the distribution of MFCC and WT for semantic identification of voice commands with subsequent encryption. The basic theoretical information on the cepstral analysis is presented, an algorithm for calculating the MFCC is presented (Fig. 2). The digital speech stream of the spoken command coming from the output of the analog-to-digital converter (ADC) is divided into 20-millisecond segments of speech containing 160 samples (1280 bits). At the first stage, to suppress noise and normalize the frequency band, the digitized voice command signal is fed to the Butterworth bandpass filter.

Since the speech signal is a non-stationary random process, it was proposed to use discrete wavelet transform to process it, the input of which receives samples of the signals of speech commands, and the wavelet coefficients (WC) are formed at the output. We turn to the circuit shown in Fig. 3. The signal $f(k)$ is fed to a low-pass filter (LPF) and a high-pass filter (HPF), in which the convolution (digital filtering) is calculated by the formula: $y(k) = \sum_{l=0}^{2n-1} f(k)q(k-l)$, where $2n$ is the number of samples of the impulse response of the wavelet filter $q(k)$. Accordingly, the output of the filters will be the high-frequency $y_H(k)$ and low-frequency $y_L(k)$ components of the signal.

When moving from the current level of wavelet decomposition to the next one, decimation $\downarrow 2$ with factor 2 is performed, i.e., the signal is thinned out by a factor of two, after which WC approximations a_{1k} and details d_{1k} are formed. At the next level of decomposition, WC approximations a_{1k} are applied to the low-pass and high-pass filters instead of the signal $f(k)$, and the WC details d_{1k} remain unchanged [5].

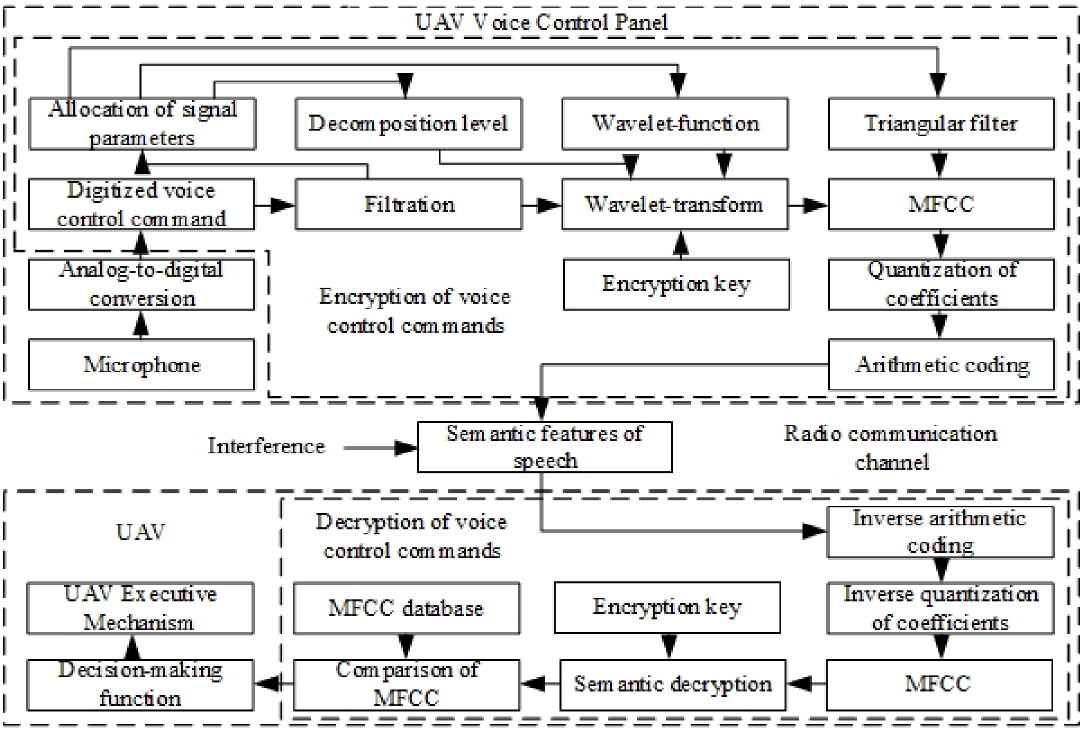

Fig. 1. Protected Voice Control System of UAV

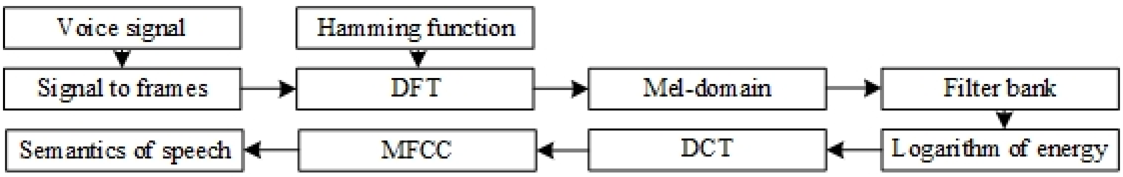

Fig. 2. Algorithm for calculation of MFCC for UAV voice control

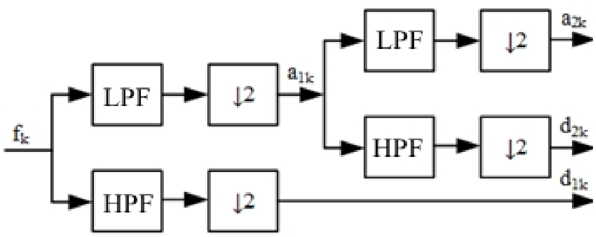

Fig. 3. Wavelet-coefficient calculation algorithm

In practice, the Hamming window, which has the following form, is often used as a weight function:

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

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

Then the DFT of the weighted speech signal can be written in the form of a formula:

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

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

Transfer to the mel-frequency region is carried out according to the formula:

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

The formula for separation into triangular filters will be as follows:

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

where, N_f is the number of mel-filters; $M(F)$ is the frequency translation in mel; $M(F_{\max} - F_{\min})$ is the frequency range.

We form triangular filters according to the following formula:

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

where $H_m[k]$ are filter coefficients.

The obtained values of the spectrum energy are presented as follows:

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

The final step in the calculation of the MFCC is the DCT, which is given by the following formula:

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

The resulting set of values is called MFCC [6].

Figure 4 shows calculated by developed algorithm of MFCC experimental samples of voice commands control subject #1: “Up”, “Down”, “Right”, “Left”.

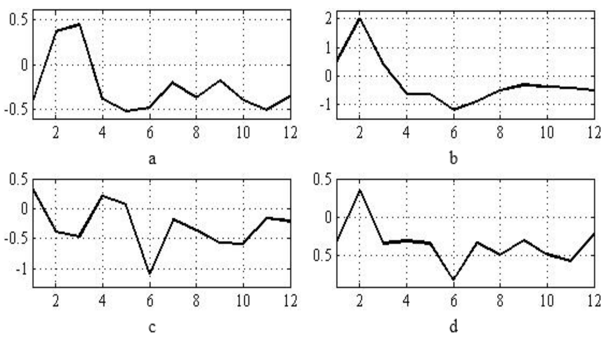

Fig. 4. MFCC voice commands control subject #1: (a) – “Up”; (b) – “Down”; (c) – “Right”; (d) – “Left”

In this system for evaluation of the results of automatic recognition of voice commands a classifier built by the criterion of minimum distance is used. As such indicator figure stands variance of the difference of the expectation value of MFCC saved in base of standard voice samples with expectation value of MFCC at the testing system level.

V. Results

In the article, the developed Protected Voice Control System of UAV has been studied and modeling has been performed in the MATLAB software package, in particular, the compression coefficient (C), bit rate (BR), correlation coefficient (CC), signal-to-noise ratio (SNR) were evaluated, peak signal-to-noise ratio (PSNR) and root-mean-square error (RMSE), which are the main indicators of the quality of recognition of speech command signals. The result of scientific and experimental studies of the Protected Voice Control System of UAV is described below (Table I).

From the above research results (Table I), it is clearly seen that the optimal solution according to the given criteria for the quality of speech recognition for voice commands is: “Up” – C = 444 times, BR = = 144 bit/s, CC = 0.95, SNR = 21 dB, PSNR = 38 dB, RMSE = 0.04; “Down” – C = 444 times, BR = =144 bit/s, CC = 0.94, SNR = 23 dB, PSNR = =41 dB, RMSE = 0.09; “Right” – C = 444 times, BR = 144 bit/s, CC = 0.97, SNR = 26 dB, PSNR = =43 dB, RMSE = 0.06; “Left” – C = 444 times, BR = =144 bit/s, CC = 0.91, SNR = 22 dB, PSNR = 38 dB, RMSE = 0.1. This shows rather good results, while preserving the individual semantic components of the spoken commands of the control subject, which makes it possible to accurately recognize voice commands for controlling UAV functions and decrypting MSCC of legal control subjects on the receiving side of the voice control system for UAV functions. An experimental study showed that the developed system can work with a speech recognition coefficient in the range from 94–97%, with a signal-to-noise ratio of about 22 dB, without exceeding the boundary value of the processing and data transfer delay of 300 ms, this allows the system to operate in real time.

TABLE I TEST RESULTS OF PROTECTED VOICE CONTROL SYSTEM OF UAV

Learning	**Testing**			
Voice commands	"Up"	"Down"	"Right"	"Left"
"Up"	C = 444 times BR = 144 bit/s CC = 0.95 SNR = 21 dB PSNR = 38 dB RMSE = 0.04	C = 444 times BR = 144 bit/s CC = 0.32 SNR = 6 dB PSNR = 11 dB RMSE = 0.38	C = 444 times BR = 144 bit/s CC = 0.12 SNR = 4 dB PSNR = 9 dB RMSE = 0.52	C = 444 times BR = 144 bit/s CC = 0.41 SNR = 7 dB PSNR = 15 dB RMSE = 0.45
"Down"	C = 444 times BR = 144 bit/s CC = 0.37 SNR = 8 dB PSNR = 14 dB RMSE = 0.33	C = 444 times BR = 144 bit/s CC = 0.94 SNR = 23 dB PSNR = 41 dB RMSE = 0.09	C = 444 times BR = 144 bit/s CC = 0.22 SNR = 3 dB PSNR = 7 dB RMSE = 0.38	C = 444 times BR = 144 bit/s CC = 0.34 SNR = 6 dB PSNR = 13 dB RMSE = 0.43
"Right"	C = 444 times BR = 144 bit/s CC = 0.14 SNR = 9 dB PSNR = 16 dB RMSE = 0.52	C = 444 times BR = 144 bit/s CC = 0.27 SNR = 7 dB PSNR = 11 dB RMSE = 0.39	C = 444 times BR = 144 bit/s CC = 0.97 SNR = 26 dB PSNR = 43 dB RMSE = 0.06	C = 444 times BR = 144 bit/s CC = 0.18 SNR = 5 dB PSNR = 14 dB RMSE = 0.67
"Left"	C = 444 times BR = 144 bit/s CC = 0.45 SNR = 3 dB PSNR = 9 dB RMSE = 0.47	C = 444 times BR = 144 bit/s CC = 0.38 SNR = 4 dB PSNR = 10 dB RMSE = 0.42	C = 444 times BR = 144 bit/s CC = 0.24 SNR = 8 dB PSNR = 14 dB RMSE = 0.27	C = 444 times BR = 144 bit/s CC = 0.91 SNR = 22 dB PSNR = 38 dB RMSE = 0.1

VI. CONCLUSIONS

This paper introduces voice command recognition system based on steganographic-cepstral analysis, which allows to increase the efficiency of voice recognition. The paper details the results of preliminary experimental studies on which conclusions on the desirability of further research and practical application of the developed voice command recognition system based on steganographic-cepstral analysis algorithm for calculation of MFCC, as well as justification of scientific importance of research. The comparative assessment of the calculated values from the selected minimum distance criterion, which is the main indicator of the quality of voice recognition test has been carried out.

Also the article proposes and explores a protected voice radio control system of UAV's functions, which is based on the WT and the identification of relevant semantic features of speech recognition. A software package has been developed that includes tools for compiling a base of reference voice images of control subjects for training and testing a recognition system for steganographic-transformed voice control commands for UAV based on cepstral analysis and computer models of the proposed methods and algorithms for recognizing voice control commands in the MATLAB environment. An algorithm is presented for calculating MFCCs, which act as the main recognition features and are the result of steganographic speech transformed, where a classifier is used to evaluate the results of automatic recognition of voice control commands, built according to the minimum distance criteria in the role of which are C, BR, CC, SNR, PSNR, RMSE. The results of an experimental study allow us to conclude the feasibility of further practical application of the developed Protected Voice Control System of UAV.

REFERENCES

[1] A. R. Fayjie, A. Ramezani, D. Oualid and D. J. Lee, "Voice enabled smart drone control," *Ubiquitous and Future Networks (ICUFN), IEEE 9th International Conference*, 2017, pp. 119–121. https://doi.org/10.1109/ICUFN.2017.7993759

[2] L. Jie, C. Jian and W. Lei, "Design of multi-mode UAV human-computer interaction system," *Unmanned Systems (ICUS), IEEE International*

Conference, 2017, pp. 353–357. https://doi.org/10.1109/ICUS.2017.8278368

[3] S. W. Jung, K. W. Sung, M. Y. Park, E. U. Kang, W. J. Hwang, J. D. Won, W. S. Lee and S. H. Han, "A study on robust control of mobile robot by voice command," *Control, Automation and Systems (CAS), IEEE 13th International Conference*, 2013, pp. 657–659. https://doi.org/10.1109/ICCAS.2013.6703950

[4] S. Zhang, X. Liu, J. Yu, L. Zhang and X. Zhou, "Research on Multi-modal Interactive Control for Quadrotor UAV," *Networking, Sensing and Control (NSC), IEEE 16th International Conference*, 2019, pp. 329–334. https://doi.org/10.1109/ICNSC.2019.8743337

[5] O. Lavrynenko, G. Konakhovych, and D. Bakhtiiarov, "Method of Voice Control Functions of the UAV," *Methods and Systems of Navigation and Motion Control (MSNMC), IEEE 4th International Conference*, 2016, pp. 47–50. https://doi.org/10.1109/MSNMC.2016.7783103

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

Received January 16, 2020

Lavrynenko Oleksandr. Post-graduate student.
Faculty of Air Navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2014).
Research area: speech recognition, digital signal processing, information security.
Publications: more than 15.
E-mail: oleksandrlavrynenko@gmail.com

Konakhovych Georgy. Doctor of Engineering Science. Professor.
Faculty of Air Navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (1968).
Research area: information security systems, telecommunication systems.
Publications: more than 100.
E-mail: tks@nau.edu.ua

Bakhtiiarov Denis. Post-graduate student.
Faculty of Air Navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2013).
Research area: information security systems, telecommunication systems.
Publications: more than 15.
E-mail: tks@nau.edu.ua

О. Ю. Лавриненко, Г. Ф. Конахович, Д. І. Бахтіяров. Система захищеного голосового керування безпілотним літальним апаратом

Розроблено систему розпізнавання стеганографічно-перетворених голосових команд керування безпілотним літальним апаратом на основі кепстрального аналізу, яка дозволяє забезпечити ефективне розпізнавання і секретну передачу команд на борт безпілотного літального апарату, за допомогою перетворення голосових команд керування в свого роду стеганографічний вектор ознак, що має на увазі приховування інформації голосового керування безпілотним літальним апаратом. Синтезована математична модель алгоритму обчислення мел-частотних кепстральних коефіцієнтів і класифікатора розпізнавання голосових команд керування для вирішення завдання семантичної ідентифікації та забезпечення скритності інформації керування безпілотного літального апарату в каналі зв'язку. Детально викладені отримані результати попередніх експериментальних досліджень розробленої системи розпізнавання голосових команд і алгоритму обчислення мел-частотних кепстральних коефіцієнтів в середовищі MATLAB на прикладі ідентифікації вимовлених різними суб'єктами керування команд: «Вгору», «Вниз», «Вправо», «Вліво».

Ключові слова: стеганографічні перетворення мови, голосові команди керування, розпізнавання голосових команд, кепстральний аналіз, кепстральні коефіцієнти; перетворення Фур'є; вейвлет-перетворення; косинусне перетворення; система голосового керування; безпілотний літальний апарат.

Лавриненко Олександр Юрійович. Аспірант.
Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2014).
Напрямок наукової діяльності: розпізнавання мови, цифрова обробка сигналів, захист інформації.
Кількість публікацій: більше 15.
E-mail: oleksandrlavrynenko@gmail.com

Конахович Георгій Филимонович. Доктор технічних наук. Професор.
Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (1968).
Напрямок наукової діяльності: системи захисту інформації, телекомунікаційні системи.
Кількість публікацій: більше 100.
E-mail: tks@nau.edu.ua

Бахтіяров Денис Ілшатович. Аспірант.
Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2013).
Напрямок наукової діяльності: системи захисту інформації, телекомунікаційні системи.
Кількість публікацій: більше 15.
E-mail: tks@nau.edu.ua

А. Ю. Лавриненко, Г. Ф. Конахович, Д. И. Бахтияров. Система защищённого голосового управления беспилотным летательным аппаратом

Разработана система распознавания стеганографически-преобразованных голосовых команд управления беспилотным летательным аппаратом на основе кепстрального анализа, которая позволяет обеспечить эффективное распознавание и скрытную передачу команд на борт беспилотного летательного аппарата, посредством преобразования голосовых команд управления в своего рода стеганографический вектор признаков, что подразумевает скрытие информации голосового управления беспилотным летательным аппаратом. Синтезирована математическая модель алгоритма вычисления мел-частотных кепстральных коэффициентов и классификатора распознавания голосовых команд управления для решения задачи семантической идентификации и обеспечения скрытности информации управления беспилотного летательного аппарата в канале связи. Подробно изложены полученные результаты предварительных экспериментальных исследований разработанной системы распознавания голосовых команд и алгоритма вычисления мел-частотных кепстральных коэффициентов в среде MATLAB на примере идентификации произнесенных разными субъектами управления команд: «Вверх», «Вниз», «Вправо», «Влево».

Ключевые слова: стеганографическое преобразование речи; голосовые команды управления; распознавания голосовых команд; кепстральный анализ; кепстральные коэффициенты; преобразование Фурье; вейвлет-преобразование; косинусное преобразование; система голосового управления; беспилотный летательный аппарат.

Лавриненко Александр Юрьевич. Аспирант.
Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, (2014).
Направление научной деятельности: распознавания речи, цифровая обработка сигналов, защита информации.
Количество публикаций: более 15.
E-mail: oleksandrlavrynenko@gmail.com

Конахович Георгий Филимонович. Доктор технических наук. Профессор.
Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, (1968).
Направление научной деятельности: системы защиты информации, телекоммуникационные системы.
Количество публикаций: более 100.
E-mail: tks@nau.edu.ua

Бахтияров Денис Илшатович. Аспирант.
Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Национальный Авиационный Университет, Киев, Украина, (2013).
Направление научной деятельности: системы защиты информации, телекоммуникационные системы.
Количество публикаций: более 15.
E-mail: tks@nau.edu.ua