The Method of Commands Identification to Voice Control of the Electric Wheelchair

Vasil Dozorskyi[a], Iryna Dediv[a], Sofiia Sverstiuk[b], Vyacheslav Nykytyuk[a] and Andrii Karnaukhov[a]

[a] *Ternopil Ivan Puluj National Technical University, Ruska str., 56, Ternopil, 46001, Ukraine*
[b] *Ternopil National Pedagogical University, 2 Maxyma Kryvonosa St., Ternopil, 46027, Ukraine*

Abstract
The paper shows the relevance of the task of improving the electric wheelchair control system for patients with lost or functionally limited upper limbs. A method of indirect control using voice commands is proposed. At the same time, the biometric principle is used, according to which, in the process of voice signal analysis, signs that are biometric parameters of the patient's speech are identified to prevent the operation of the control system when registering similar commands from outsiders. The main such sign is the value of the main tone frequency. To process voice signals in order to identify the four main control commands, namely forward, backward, left and right, the sliding window method is used, within which the presence of signs of the main tone is evaluated. For a specific person, this value will be individual and present only in those parts of the voice signal that correspond to vowels and consonant vocalized sounds. In this way, it is possible to segment the voice signal into such areas and identify individual voice commands based on pre-set durations of these areas. A threshold function is also used for this, which takes a certain value when the main tone frequency is present in the voice signal structure within each implementation of the sliding window and is equal to zero when there are no signs of this frequency. At the same time, it became possible to identify individual commands with higher accuracy.

Keywords 1
Electric wheelchair, voice control, sliding window, main tone frequency

1. Introduction

Development a voice control system is an urgent task. These systems are useful when working with computerized systems such as voice typing, control of electronic devices in the home, etc. This technology is relevant for people with disabilities, in particular, with disorders of the musculoskeletal system.

Thus, according to the World Health Organization [1], as of 2020, more than 15% of the world's population had some form of disability. According to official statistics in Ukraine [2,3], as of January 1, 2021, there were 2,703,000 people with disabilities in Ukraine (according to the State Statistics Service), including 163,900 children among people with disabilities. It is also stated that in Ukraine there are 222.3 thousand people with disabilities of the first group, 900.8 thousand - of the second group, 1 million 416 thousand - of the third group. In connection with the war in Ukraine, the number of people with disabilities has increased significantly over the past year. Everyone understands this, despite the lack of official data today (according to the conditions of martial law). At the same time, it

is very important for people with limited physical abilities to feel like equal members of society and to be able to move freely.

Today, the industry of goods for the disabled offers special means of transportation – wheelchairs [4]. There are several types of wheelchairs, the choice of which depends on the degree of impaired mobility of the patient. In the past, people with partial mobility could only use mechanical devices. But today, designers are trying to do everything possible to make life easier for patients with disabilities. Today, the preference is given to wheelchairs with an electric drive. Such devices are more comfortable to use, because to activate them, it is enough to press the joystick. A small movement is enough, and a person sitting in such wheelchair can start moving independently without waiting for help.

Using an electric wheelchair does not require much effort. This is what makes such means of transportation popular among people with serious illnesses, after injuries: with weak lower and upper limbs, the body as a whole; with cardiovascular diseases; with spinal injuries resulting in leg problems; with paralysis or paresis of the legs, the whole body; after limb amputation. Limb tremors, partial paralysis or amputations may also limit the use of such wheelchairs, as a result of which it will be difficult or even impossible for the patient to control such wheelchair using a joystick.

However, today, with the development of information technologies and innovative developments, attempts are being made to create alternative methods of wheelchairs controlling. So, the Intel company together with the Brazilian startup Hoobox Robotics presented Wheelie 7 - a smart wheelchair whose movements can be controlled by changing facial expressions [5]. The wheelchair uses a software system with a camera that recognizes ten facial expressions, each of which the chair owner can match with a specific control command.

There is also active development of wheelchairs that are controlled by the mind [6]. But they require a long time to learn and adapt for each individual user.

Another example is the control of an electric wheelchair using signals from sensors that register active muscle contractions (for example, Stephen Hawking's wheelchair, also developed by Intel, and the attached on-board computer wass controlled almost entirely by cheek muscle contractions) or cameras recording eye movements.

Similar technologies include the "Vocalizal" technology described in [7]. In this case, a method of voice control of a smartphone or mobile phone is proposed for disabled people using electric wheelchairs. With voice recognition, Vocalizal constantly listens for key words, eliminating the need to press any buttons before or while using the mobile phone. It offers new and improved voice control features with improved recognition quality. Such a device is attached to a wheelchair and communicates with a mobile phone wirelessly using the Bluetooth protocol.

In the research, an attempt is being made to develop a prototype of the voice control system for an electric wheelchair, in particular, the development of algorithms for the operation of such a system and appropriate methods of processing voice signals to form the necessary command signals, taking into account non-standard and dangerous situations. Some of such situations are described in works [8,9].

In similar systems, which are used today in the field of computer control, for example, voice signals are processed and the signs are selected in their structure, which are then used to identify or distinguish individual commands, which are certain keywords. For this, signs of the phonemic composition of the speech, amplitude and time parameters of the voice signal, its composition, transitions between phonemes, syllables or words can be used.

There are a number of problems when creating such a system. First, the lack of a mathematical model of the semantics of a speech signal, since only probabilistic and heuristic methods can be used to determine the semantics of a speech signal, which do not give an exact result and the accuracy of which is inversely proportional to the number of semantic units. Secondly, the individual characteristics of the announcer: specifics of pronunciation, accents, etc. Thirdly, work with spontaneous speech and the need to determine the presence of a keyword. Fourth, differences in the acoustic environment, noises.

Thus, the proposed system must necessarily take into account all possible risk factors related in particular to the ability of the system to recognize the patient's voice commands in a certain way, not to be sensitive to registered voice signals similar in value from outsiders, to identify only those voice signals, which are actually commands for the system and do not identify commands uttered by the patient in the process of conversation, which do not relate to the process of driving a wheelchair, etc. Taking into account these and similar factors will exclude the possibility of spontaneous start or, on the

contrary, stopping the execution of the control command and creating dangerous situations for both the patient and others. So, if the wheelchair is commanded to move forward, the wheelchair should not stop in the middle of the road when the patient in the context of the conversation utters a word that should mean a command to stop, etc.

Thus, the information system being developed by us should be able to be adjusted exclusively to the acoustic parameters and characteristics of the patient's speech, ignoring the voice signals of people who may be talking nearby. Systems of this type belong to the class of biometric systems [10].

Usually, the recognition process in a biometric system goes through two stages:

• Learning – a physical or behavioral sample is memorized by the system to the database, unique information is determined and a "print" of the voice is created;

• Comparison – the presented sample is compared with the original and an answer is formed based on the obtained result.

The principles of operation of information and complex cyber-physical systems in medical applications were also analyzed [11,12,13]. Accordingly, the principle of operation of the proposed system should be based on the analysis of biometric parameters of the human voice [14]. For this purpose, a method of processing voice signals and algorithms for identifying individual commands were developed.

2. Materials and Methods

The essence of the voice signals processing method for the selection of command words and, accordingly, the formation of control signals based on them is reduced to the following. We need to recognize four commands: "left", "right", "back" and "forward". The work of the system was focused on the possibility of working with the Ukrainian language, so the following four commands (in transliteration) were recognized: [vlivo], [vpravo], [vzad], [vpered].

At the first stage, the user speaks these words into the microphone, the information system processes them and selects informative signs, which will be used to identify these commands in the next step.

First, the signal is recorded, then it is filtered using adaptive methods. Next, the word length of each command is determined. At the next stage, each word is segmented into phonemes. Actually, on the last step, areas with signs of vowels and consonant vocalized sounds are determined. All four commands differ in their phonemic composition for Ukrainian language. So, the first command contains 5 phonemes, of which 2 vowels and 3 consonants are vocalized. The second command contains 6 phonemes, of which 2 are vowels, 3 are vocalized consonants, and one is voiceless. The third command contains 4 phonemes with one vowel and three vocalized consonants. And the fourth command contains 6 phonemes with two vowels, one voiceless and three consonant vocalized phonemes. The alternation of phonemes in commands is different. Therefore, for phoneme segmentation, the value of the main tone frequency, which is also the main biometric characteristic of each person, is used as an informative sign.

At the first stage, the time limits of each command are identified. For this, a sliding window is used, within which the average value of the voice signal is calculated [15,16,17]. Further, this value is compared with the threshold, and if it is equal to or exceeds the threshold, then the value of the threshold function h is taken as 1, otherwise - 0. The algorithm of such processing is shown in Figure 1, a.

At the next stage, the main tone frequency is determined, which is an individual characteristic and is present in vowels and consonants vocalized phonemes. For this, the method of formant analysis was used, when the main tone frequency corresponds to the frequency of the first maximum in the power spectrum. For this, the processing was also carried out using a sliding window. The algorithm of such processing is shown in Figure 1, b.

And at the last stage, the command is identified. For this, the unknown command is processed within the sliding window as follows. The presence of the main tone frequency is assessed. It will be present only in the sections within the duration of the command that correspond to vowels and consonant vocalized sounds. Further, intervals with the presence of such a frequency are formed, and already by their sequence and size, the command is identified, since the sequences of these sections and durations must be different for the four considered commands. Having identified the commands,

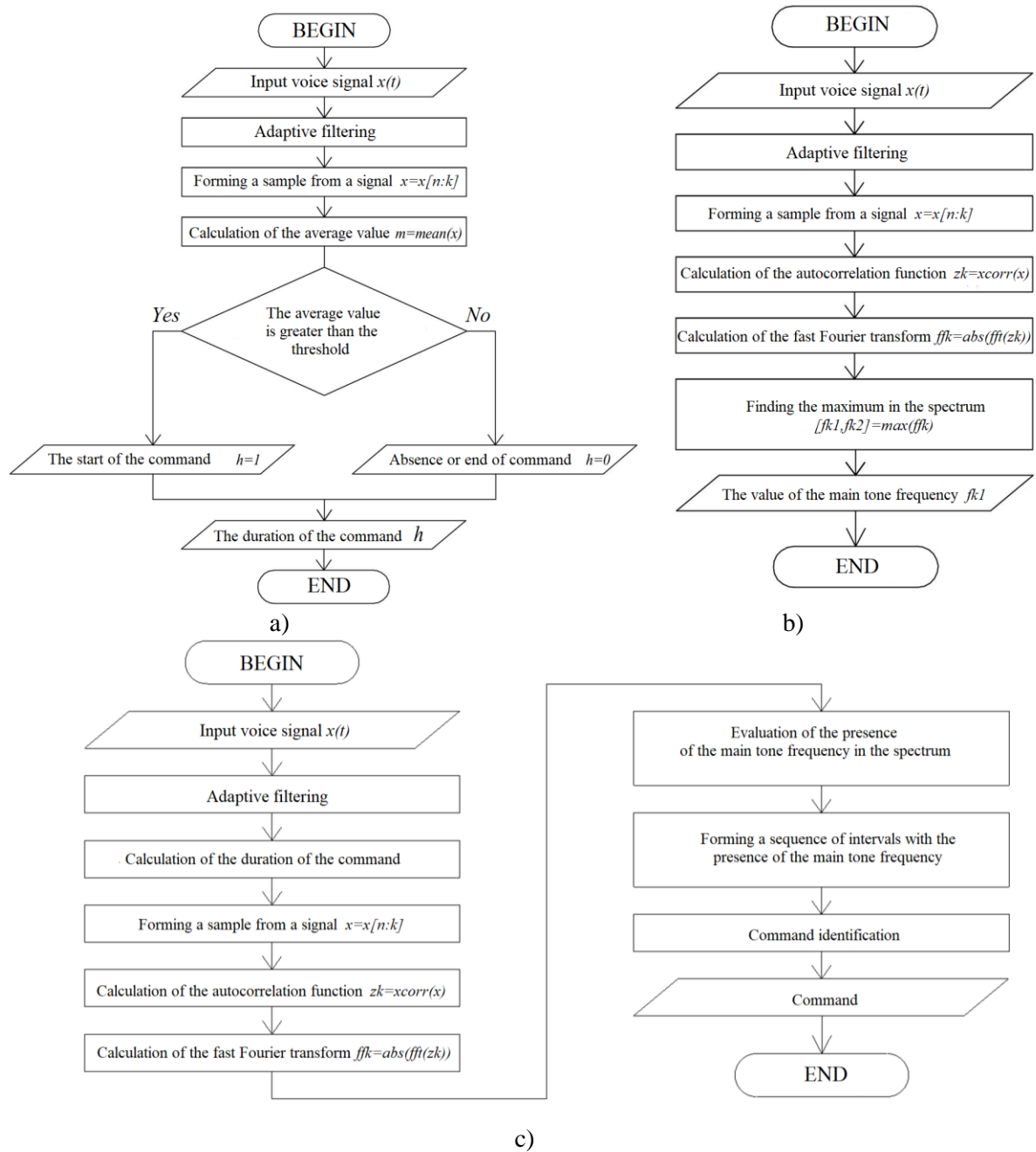generating control signals is not a difficult task. The command recognition algorithm is shown in Figure 1, c.

Figure 1: The algorithm for finding the duration of the command (a), the algorithm for and for estimating the value of the main tone frequency (b), command recognition algorithm (c)

3. Experiment and Results

A standard computer headset and the Adobe Audition program were used to select voice signals with spoken commands. In the same environment, adaptive filtering was carried out by the method of spectral subtraction. The appearance of the signal after such filtering is shown in Figure 2.
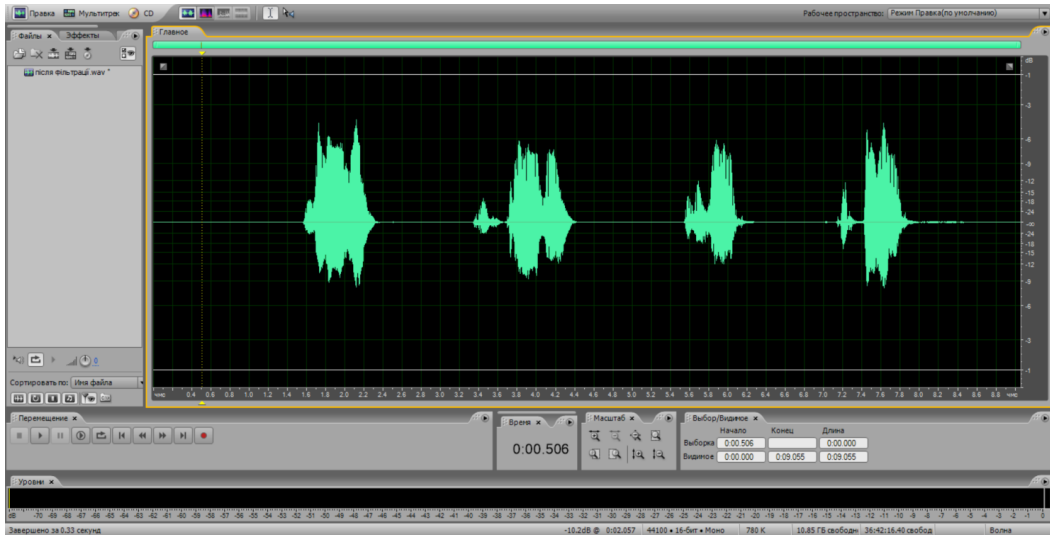
Figure 2: The appearance of the signal after adaptive filtering

Further, the filtered signal was loaded into the Matlab environment, in which its further processing was carried out. The appearance of such a signal is shown in Figure 3.
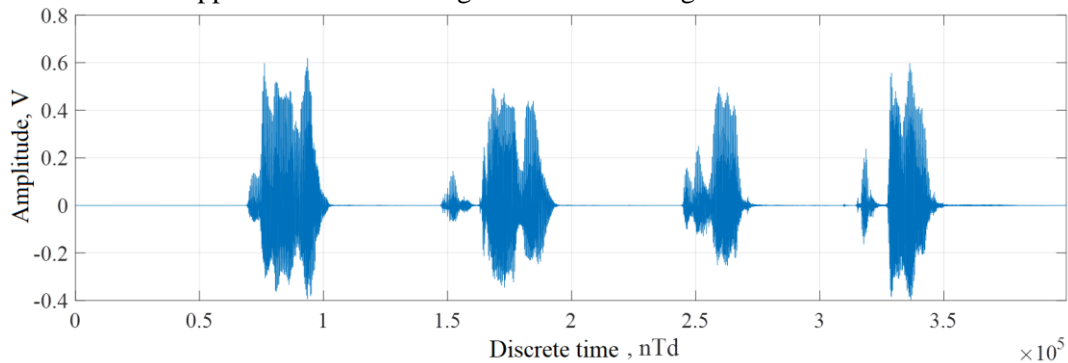

Figure 3: The appearance of voice signal with spoken commands, loaded in Matlab

According to the algorithm in Figure 1, a, the duration of each command is determined. For this, a sliding window with a duration of 5ms was used, which was chosen a priori and was more than 100 times shorter than the implementation of the command. The threshold was also chosen a priori. The view of the value of the threshold function is shown in Figure 4.
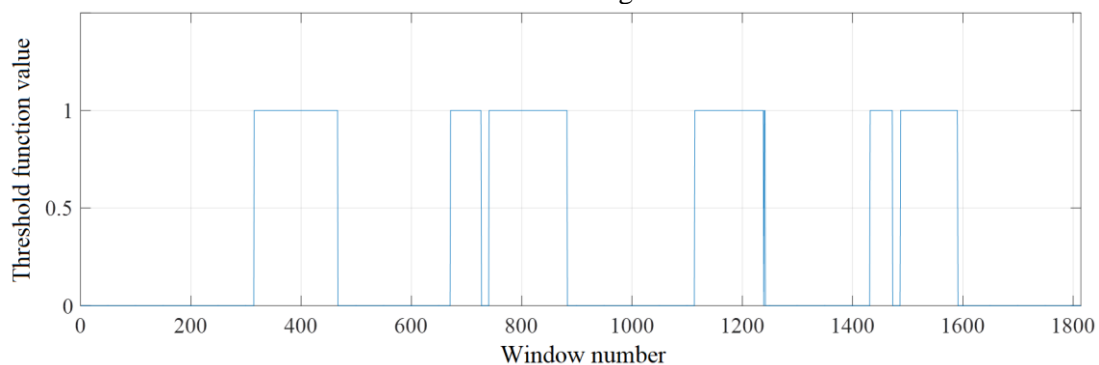

Figure 4: View of the threshold function

As it can be seen, in this way, the duration of voice commands can really be determined by the threshold function.

At the next stage, the value of the main tone frequency was determined, which was 172 Hz for this case (according to the algorithm in Figure 1, b).

At the last stage, phonemic segmentation of each individual command was carried out. In Figure 5, a the appearance of the voice signal of the first command [vlivo] is given.

Further, according to the algorithm described above, estimates of the spectrum were constructed and its averaging was carried out around the values of the main tone frequency. Processing was carried out within the limits of the sliding window. If the main tone frequency was present within these limits, it was manifested in the signal spectrum. These averaged spectrum values around the main tone frequency were plotted on the same time axis according to where the sliding window was placed. The resulting graph is shown in Figure 5, b.

Next, a threshold function was constructed with a threshold of 0.1, which was chosen a priori. Its appearance is shown in Figure 5, c. It can be concluded that the main tone frequency is present in the signal for the full duration of the command, which is true since all phonemes are either vowels or vocalized consonants.

In Figure 6 shows the view of the signal of the second command [vpravo], the graph of the main tone frequency presence in the signal and the view of the threshold function. In this case, there is a gap that corresponds to the voiceless phoneme [p].
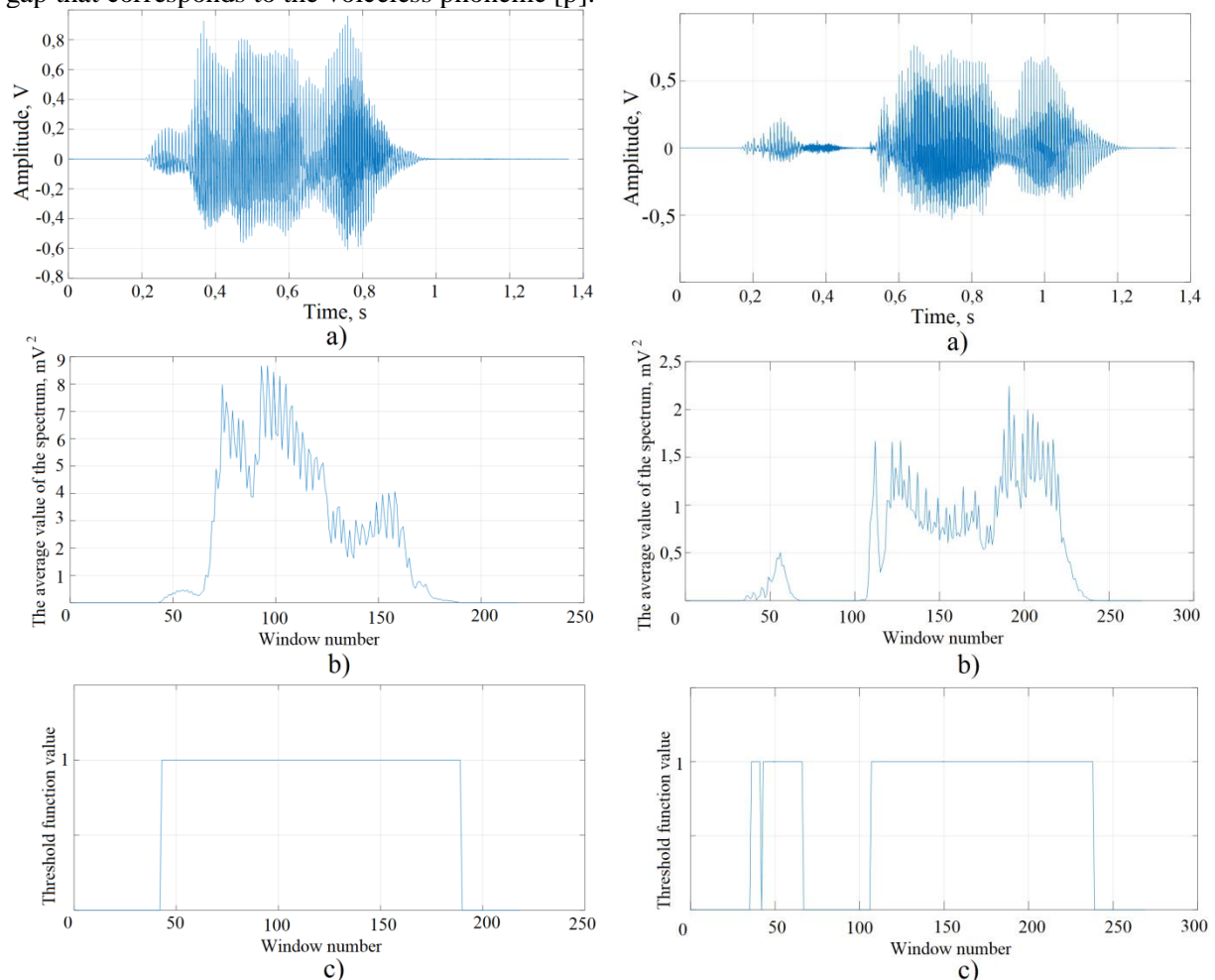
Figure 5: The view of the command signal [vlivo] (a), the graph of the main tone frequency presence in the signal (b) and the view of the threshold function (c)

Figure 6: The view of the command signal [vpravo] (a), the graph of the main tone frequency presence in the signal (b) and the view of the threshold function (c)

In Figure 7 shows the view of the signal of the third command [vzad], the graph of the main tone frequency presence in the signal and the view of the threshold function. In the picture it can be seen an interphonemic gap.

In Figure 8 shows the view of the signal of the third command [vpered], the graph of the main tone frequency presence in the signal and the view of the threshold function.
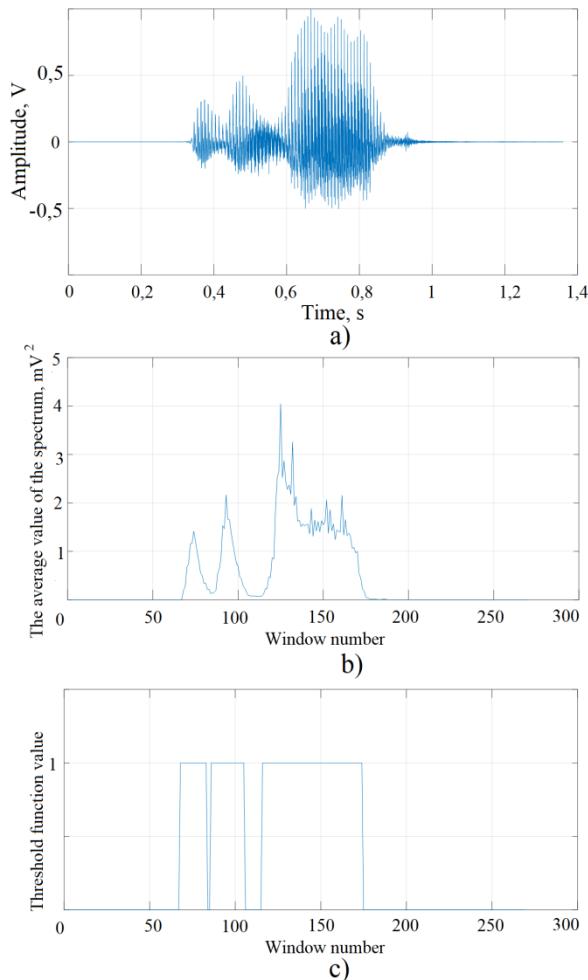
Figure 7: The view of the command signal [vzad] (a), the graph of the main tone frequency presence in the signal (b) and the view of the threshold function (c)
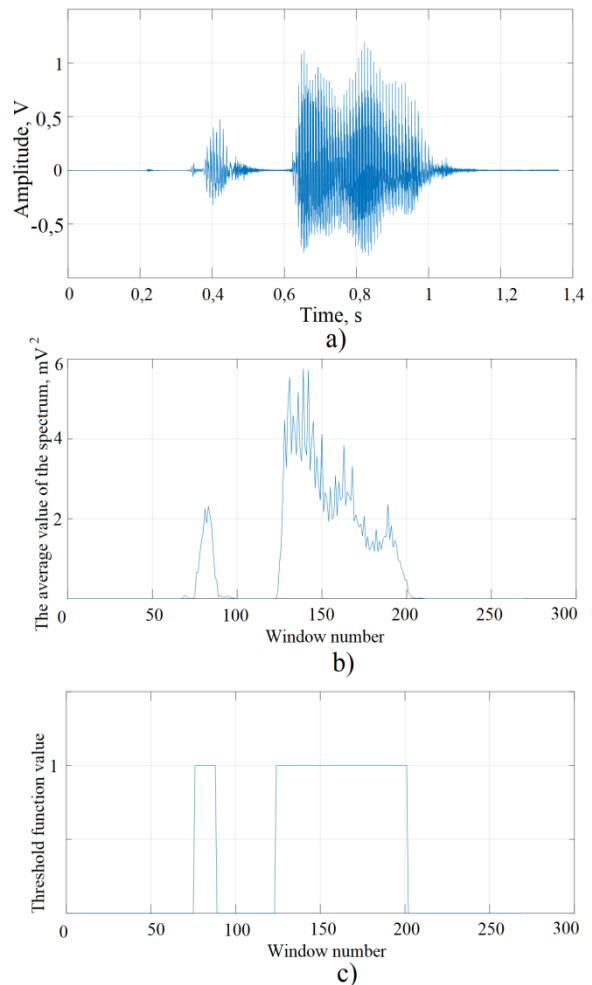
Figure 8: The view of the command signal [vpered] (a), the graph of the main tone frequency presence in the signal (b) and the view of the threshold function (c)

Figure 8 also clearly shows the omission due to the presence of a voiceless phoneme. In addition, analyzing the given graphs for four commands, we can come to the conclusion that based on the values of the threshold function and the durations of the commands, their identification and the subsequent formation of the corresponding control signals can be carried out.

In order for the system not to be sensitive to similar commands that can be uttered by the patient in the context of a conversation, it is suggested to use a certain key word, after the utterance of which the command word must be unambiguously followed. Also, the second keyword will mean the end of the command. These keywords can be selected by the system user individually, but with the lowest possible frequency of occurrence in the speech of the user himself. Ideally, it can be a meaningless set of sounds or numbers. Actually, in this case, it will be possible to avoid the dangerous situations mentioned above.

4. Conclusion

A method of voice signals processing has been developed to recognize four commands in the voice signal. First, the user speaks these words into the microphone, the information system processes them and selects informative signs, which will be used to identify these commands in the future.

First, the signal is recorded, then it is filtered using adaptive methods. Next, the word length of each command is selected. At the next stage, each word is segmented into phonemes. Actually, on the last step, the areas with signs of vowel and consonant vocalized sounds are selected.

The sliding window method and the threshold function were used for the actual processing. At the same time, the values of the energy characteristics of the signal and the excess of their value over the predetermined threshold value were evaluated. During the next processing, the signals were divided into sections that correspond to individual phonemes, in particular vowels and vocalized consonants. A sliding window and a threshold function were also used in this case. Analyzing the results of processing four commands, it was concluded that based on the values of the threshold function and the duration of the commands, their identification and the subsequent formation of the corresponding control signals can be carried out.

5. References

[1] World Health Organization. URL: https://www.who.int/
[2] State Statistics Service of Ukraine. URL: https://www.ukrstat.gov.ua/
[3] National Assembly of People with Disabilities of Ukraine. URL: https://www.facebook.com/vgonaiu/posts/1709214279269040/?locale=hi_IN
[4] Electric wheelchairs. URL: https://techno-med.com.ua/ua/elektrokoljaski
[5] Wheelie 7. URL: https://theindexproject.org/award/nominees/3326
[6] Yes, mind-controlled wheelchairs are a thing. URL: https://psmag.com/environment/mind-controlled-wheelchairs-are-here
[7] Voice control bluetooth mobile phone control system for wheelchair. URL: https://inclusiveinc.org/uk-ua/products/vocalize-wheelchair
[8] Hevko, B.M., Hevko, R.B., Klendii, O.M., Buriak M.V., Dzyadykevych, Y.V., Rozum, R.I. Improvement of machine safety devices. Acta Polytechnica, 58(1), pp. 17-25 (2018)
[9] Buketov, A., Maruschak, P., Sapronov, O., Zinchenko, D., Yatsyuk, V., Panin, S. Enhancing performance characteristics of equipment of sea and river transport by using epoxy composites. Transport, 31(3), pp. 333-342 (2016)
[10] What is Biometric Authentication? By Dean Nicolls, 2019. URL: https://www.jumio.com/what-is-biometric-authentication/
[11] Martsenyuk, V., Klos-Witkowska, A., Sverstiuk, A., Bernas, M., Witos, K. Intelligent big data system based on scientific machine learning of cyber-physical systems of medical and biological processes. CEUR Workshop Proceedings, 2021, Vol. 2864, pp. 34–48. https://ceur-ws.org/Vol-2864/paper4.pdf.
[12] Lypak, H., Rzheuskyi, A., Kunanets, N., Pasichnyk, V. Formation of a Consolidated Information Resource by Means of Cloud Technologies. 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 – Proceedings 8632106, pp. 157-160
[13] Martsenyuk, V., Sverstiuk, A., Bahrii-Zaiats, O., Rudyak, Y., Shelestovskyi, B. Software complex in the study of the mathematical model of cyber-physical systems. CEUR Workshop Proceedings, 2020, Vol. 2762, pp. 87–97. https://ceur-ws.org/Vol-2762/paper5.pdf
[14] Geoffrey S. Meltzner, James T. Heaton, Yunbin Deng, Gianluca De Luca, Serge H. Roy, and Joshua C. Kline: Silent Speech Recognition as an Alternative Communication Device for Persons with Laryngectomy, IEEE/ACM Trans Audio Speech Lang Process, pp. 2386–2398 (2017).
[15] Dozorskyi V.G., Dozorska O.F., Yavorska E.B.: Selection and processing of biosignals for the task of Human Communicative Function Restoration, Kremenchug National University, 4(105), pp. 9-14 (2017).
[16] Vyacheslav Nykytyuk, Vasyl Dozorskyi, Oksana Dozorska: Detection of biomedical signals disruption using a sliding window, Scientific Journal of TNTU, Ternopil, 91(3), pp. 125–133 (2018).
[17] Martsenyuk, V., Sverstiuk, A., Klos-Witkowska, A., Horkunenko, A., Rajba, S. Vector of diagnostic features in the form of decomposition coefficients of statistical estimates using a cyclic random process model of cardiosignal. Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2019, Vol. 1, pp. 298–303. DOI: 10.1109/IDAACS.2019.8924398.