Control Actions Using Voice and Gestures at the Level of the Operating System

Serhii Kulibaba, Oleg Kurchenko and Liudmyla Zubyk

Taras Shevchenko National University of Kyiv, 60 Volodymyrska str., Kyiv, 01601, Ukraine

Abstract

The task of developing a tool that will perform various user tasks at the level of the operating system with the help of voice and / or gestures is under consideration. Processing of input data will take place at the expense of the created modules, which are associated with ready-made solutions. Analyzing the status of the issue, physical remote control devices were seen that could interact with the operating system, but no software was found. The application of this tool can be in various spheres of activity - commercial and general. The purpose of this application is to provide some part of the community with the opportunity to use most of the functions in various applications, in particular, the general use of a certain operating system.

Keywords [1]

Voice, gesture, control, command.

1. Introduction

Currently, modern technologies are rapidly developing and are in demand among society. It is possible to notice a variety of sensors that allow performing actions automatically, software for performing a number of tasks [1, 2].

It has been found that a certain part of the society cannot use most of the applications due to the respective defects. Software or application developers pay little attention to maintaining or developing projects that could solve common problems.

Therefore, it was decided to develop a model that would allow solving most problems in the use of information technologies at the level of the operating system with the help of a single software tool.

2. Analysis of publications, state of the issue and statement of the problem

2.1 Analysis of research and publications

Using voice assistants is becoming a normal everyday thing. Various companies are implementing voice assistants in their applications. The reason for this is to simplify the use of the product due to additional technologies. In [3], the principle of using ready-made solutions in voice recognition is considered. In addition to voice assistants, gesture management is rapidly developing. Currently, companies are implementing such solutions in their cars in order to provide their customers with a certain convenience in using their product.

The work [4] shows an example of the application of gesture control, which is adapted to another system.

Dynamical System Modeling and Stability Investigation (DSMSI-2023), December 19-21, 2023, Kyiv, Ukraine
EMAIL: kulibseryyy@gmail.com (S. Kulibaba); oleg.kurchenko@knu.ua (O. Kurchenko); zubyk.liudmyla@knu.ua (L. Zubyk)
ORCID: 0000-0002-7316-1214 (S. Kulibaba); 0000-0002-3507-2392 (O. Kurchenko); 0000-0002-2087-5379 (L. Zubyk)

2.2 Analysis of the issue in the applied industry

At the moment, there are various sensors and software that allow solving a number of people's problems. But the search for software that would allow solving specific problems in the use of different applications from different manufacturers was not found. Therefore, it was decided to develop a model that can solve different problems of different users using voice and gestures.
Thanks to ready-made solutions in a certain programming language, as well as their combination, you can create a universal and unique solution that can be in demand - this is a feature of this model.

2.3 Formulation of the problem

Different products are adapted for different tasks [5, 6]. But no universal one was noticed. Thanks to the use of innovative technologies, it is possible to create a software application that could solve a number of problems without additional costs from the user.

3. Application development

3.1 Application of ready solutions

Different programming languages have similar data libraries. Thanks to this, you can solve similar problems in different programming languages.

Computer vision is in demand among the famous. Its application has various directions: automation of actions, tracking of objects, processing of input data, etc. [7, 8].

As an example, the Python programming language and the OpenCV data library are used. This library allows you to process both streaming videos and images [9].

Another off-the-shelf solution that is needed is voice recognition. Thanks to voice recognition, it is possible to adapt certain actions to the application [10]. There are several libraries that allow you to implement voice recognition. Among the well-known are Pocket Sphinx and Speech Recognition [11, 12]. Pocket Sphinx allows you to create a model that will recognize only those words that are listed in the dictionary [13]. Speech Recognition is a model that is trained using a set of specified words. The set of words that includes Speech Recognition can be enough for most tasks.

3.2 Structure of the project

The project contains various directories and executable files, as well as additional configuration files for saving user settings.

The structure of the project is divided in such a way that it would be easy to navigate through it. There are SOLID design principles, where the first principle is applied to the project structure (Single Responsibility Principle), which indicates the division of responsibility between components [14].

Figure 1 shows a diagram of the application components.

Libraries. Two libraries are used - OpenCV and Speech Recognition. The diagram shows the connections between the libraries and other system components.

Voice recognition. This directory contains a basic set of functions and classes that will allow recognition and processing of input data, where the specified user action can be performed as a result. Since the set of necessary words can be large, the commands themselves are placed in a separate file, where the system can be flexibly scaled in the future.

Gesture recognition. A directory that contains the necessary file that will help recognize gestures.

Component interfaces. This directory contains the interfaces of all project classes.

The precondition for creating interfaces is the second principle of SOLID – classes are open to extension, but closed to modifications (Open-Closed Principle).

Main and configuration files. The main file (main.py) contains a set of functions that work compositionally with speech and voice recognition modules. To save the user configuration, a separate file (.conf) is created.

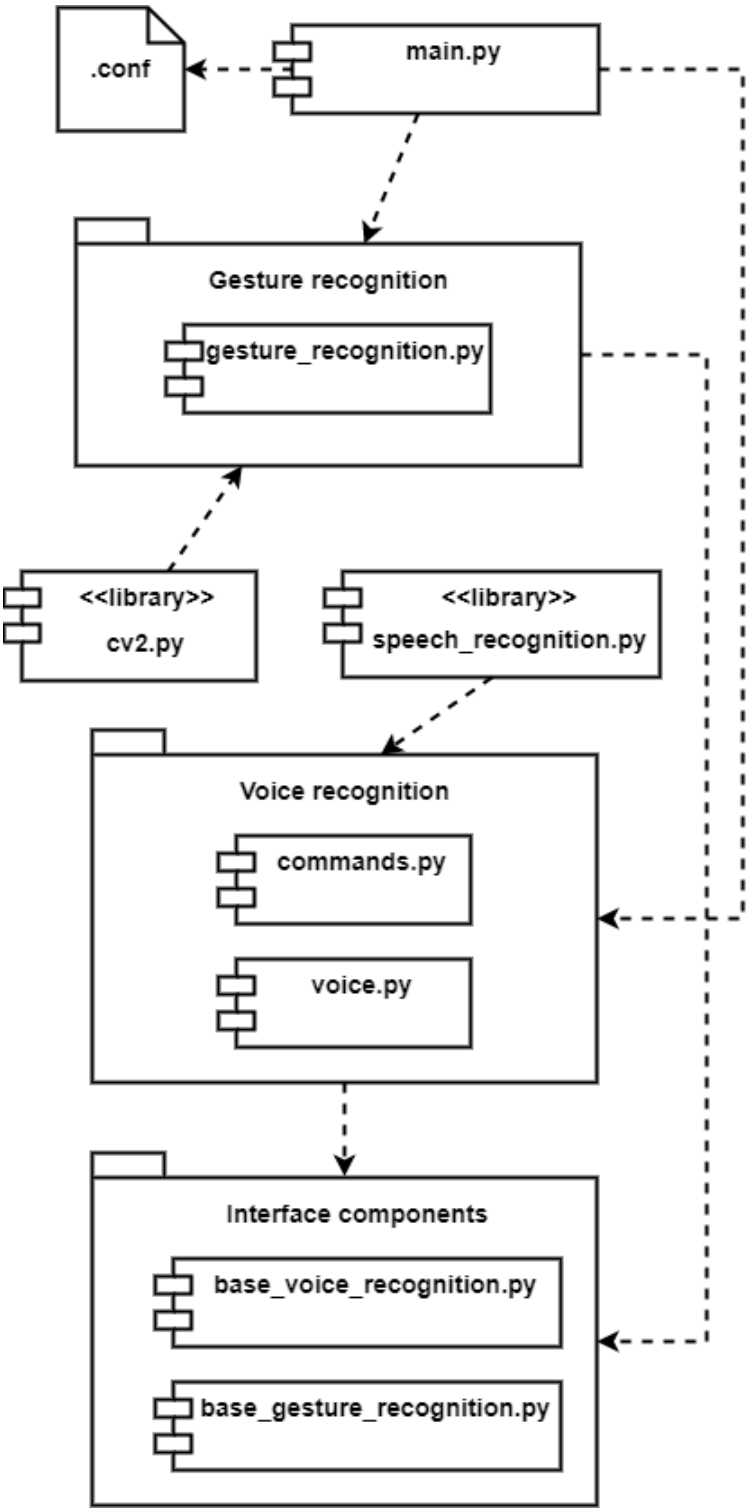

Figure 1: Component diagram

3.3 Principle of operation of the application

At the beginning of working with the program, the user has the opportunity to make some settings or directly start managing actions automatically (Fig. 2). The setup includes several stages: determining the position of the eyes, nose, and main voice commands. When the setup process is complete, the data will be saved to a separate file. The setting is required in order to reduce the possibility of incorrect processing of user commands.

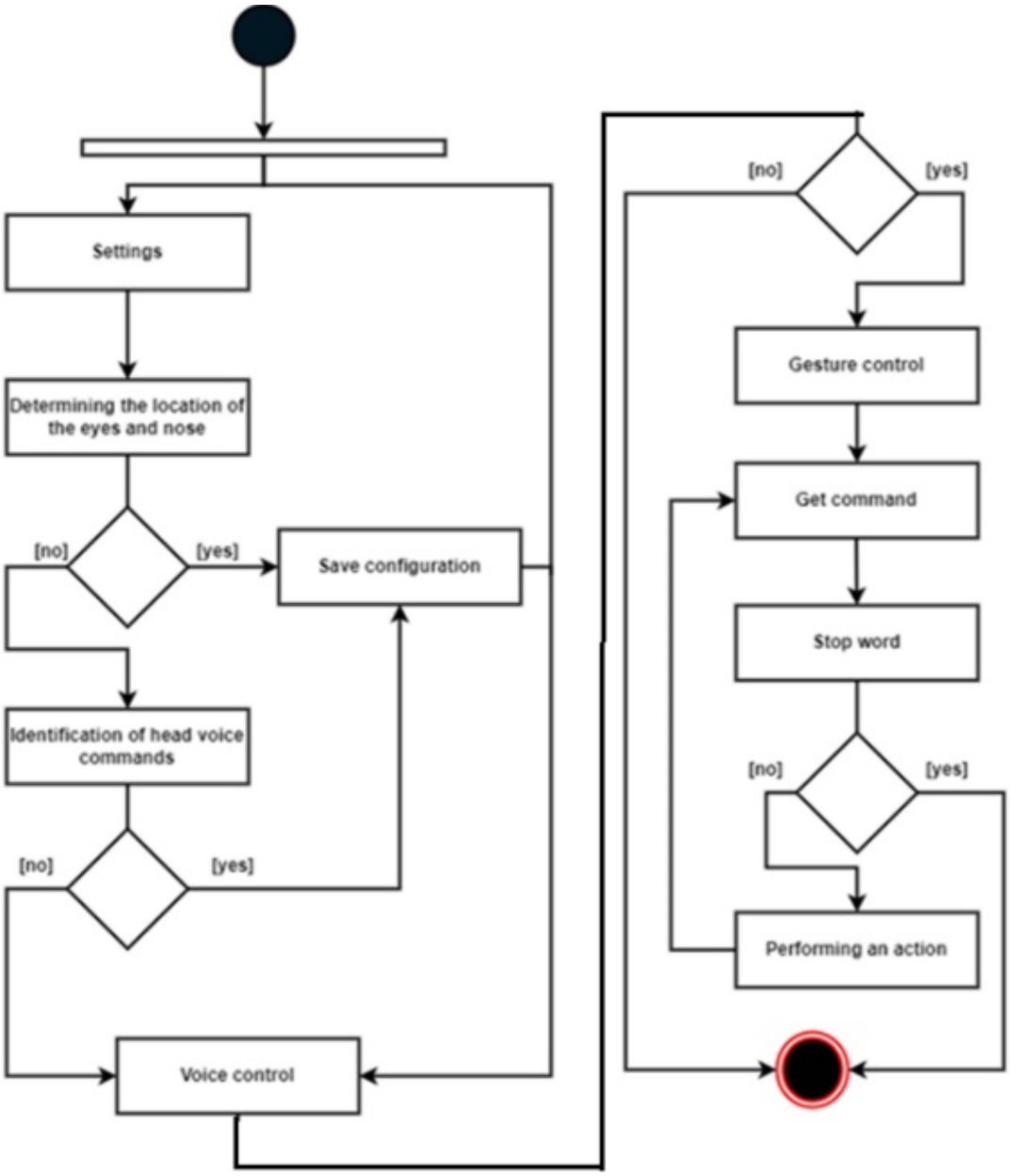

Figure 2: Activity diagram

After the setup is complete, you can start working with the application. You can use both voice control and gesture control.

The program can run until the "stop word" is triggered. The voice recognition module will wait for the corresponding word when it needs to pause the program execution process. Otherwise, all commands will be processed.

Since the application runs in a single thread, only one of the action management methods can be executed at a time. In this case, you need to use 2 separate streams for gesture control and voice control. Thanks to this, the user can perform actions faster and more conveniently.

4. Conclusions

Voice assistants and gesture control are implemented in various products or systems of companies. Having analyzed the state of the problem of the use of existing applications in society, it was found necessary to create some automation to solve most of them.

This work reflects the principle of project construction, which allows you to manage actions with the help of various existing means. Thanks to the combination of several components, a new unique and universal goal can be achieved.

For further scaling of the project, SOLID design principles were used, thanks to which the application is built in such a way that the structure of any module can be changed without problems.

5. References

[1] P. Jhunjhunwala, U. D. Atmojo and V. Vyatkin, "Towards Implementation of Interoperable Smart Sensor Services in IEC 61499 for Process Automation," 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 2020, pp. 1409-1412, doi: 10.1109/ETFA46521.2020.9211925.

[2] A. Arunachalam, R. Raghuraman, P. Obed Paul and J. Vishnupriyan, "A System for Energy Management and Home Automation," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 2021, pp. 1-3, doi: 10.1109/ICSCAN53069.2021.9526526.

[3] R. Sivapriyan, N. Sakshi and T. Vishnu Priya, "Comparative Analysis of Smart Voice Assistants," 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2021, pp. 1-6, doi: 10.1109/CSITSS54238.2021.9683722.

[4] Z. Zou, Q. Wu, Y. Zhang and K. Wen, "Design of Smart Car Control System for Gesture Recognition Based on Arduino," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2021, pp. 695-699, doi: 10.1109/ICCECE51280.2021.9342137.

[5] S. S. Pulleti, S. Inturi, H. V. Valluru, H. Oc, P. K. Putta and M. V. Jaliparthi, "Arduino Based Voice Controlled Wheelchair For Physically Challenged Persons," 2023 9th International Conference on Electrical Energy Systems (ICEES), Chennai, India, 2023, pp. 499-503, doi: 10.1109/ICEES57979.2023.10110267.

[6] R. Kristof, V. Ciupe, C. Moldovan, I. Maniu, M. Banda and A. -M. Stoian, "Arduino mobile robot with Myo Armband gesture control," 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 2019, pp. 000294-000297, doi: 10.1109/SACI46893.2019.9111627.

[7] S. S, D. Vijila and M. Shastika, "Air xylophone Using OpenCV," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-6, doi: 10.1109/ICSES55317.2022.9914191.

[8] G. Dai and P. Wang, "Design of intelligent car based on WiFi video capture and OpenCV gesture control," 2017 Chinese Automation Congress (CAC), Jinan, China, 2017, pp. 4103-4107, doi: 10.1109/CAC.2017.8243499.

[9] S. Gulati, A. K. Rastogi, M. Virmani, R. Jana, R. Pradhan and C. Gupta, "Paint / Writing Application through WebCam using MediaPipe and OpenCV," 2022 2nd International

Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 287-291, doi: 10.1109/ICIPTM54933.2022.9753939.

[10] Kulibaba, Serhii and Popereshnyak, Svitlana and Shcheblanin, Yuri and Kurchenko, Oleg and Mazur, Nataliia (2022) Advanced Communication Model with the Voice Control and the Increased Security Level Cybersecurity Providing in Information and Telecommunication Systems 2022, 3288 (1). pp. 64-72. ISSN 1613-0073.

[11] C. S. Manasa, K. J. Priya and D. Gupta, "Comparison of acoustical models of GMM-HMM based for speech recognition in Hindi using PocketSphinx," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 534-539, doi: 10.1109/ICCMC.2019.8819747.

[12] M. A. Rohan, K. S. Swaroop, B. Mounika, K. Renuka and S. Nivas, "Emotion Recognition Through Speech Signal Using Python," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2020, pp. 342-346, doi: 10.1109/ICSTCEE49637.2020.9277338.

[13] O. Zealouk, M. Hamidi and H. Satori, "Investigation on speech recognition Accuracy via Sphinx toolkits," 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 2022, pp. 1-6, doi: 10.1109/IRASET52964.2022.9738105.

[14] E. Chebanyuk and K. Markov, "An approach to class diagrams verification according to SOLID design principles," 2016 4th International Conference on Model-Driven Engineering and Software Development (MODELSWARD), Rome, Italy, 2016, pp. 435-441.

[15] Palko, D.; Babenko, T.; Bigdan, A.; Kiktev, N.; Hutsol, T.; Kuboń, M.; Hnatiienko, H.; Tabor, S.; Gorbovy, O.; Borusiewicz, A. Cyber Security Risk Modeling in Distributed Information Systems. *Appl. Sci.* 2023, *13*, 2393. https://doi.org/10.3390/app13042393

[16] O. Kalivoshko, V. Kraevsky, K. Burdeha, I. Lyutyy and N. Kiktev, "The Role of Innovation in Economic Growth: Information and Analytical Aspect," *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 2021, pp. 120-124, doi: 10.1109/PICST54195.2021.9772201