БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ • BULGARIAN ACADEMY OF SCIENCES

ПРОБЛЕМИ НА ТЕХНИЧЕСКАТА КИБЕРНЕТИКА И РОБОТИКАТА, **69**
PROBLEMS OF ENGINEERING CYBERNETICS AND ROBOTICS, **69**

София • 2018 • Sofia

# Control of service robot via voice commands

*D. Chikurtev[1, 2], I. Rangelov[2], N. Chivarov[2], N. Shivarov[2], A. Gigov[2]*

*[1]Institute of Information and Communication Technologies – BAS*
*[2]European Polytechnical University - Pernik*

*Email: denis@iinf.bas.bg, rangelov.ivaylo@yahoo.com, nshivarov@code.bg, nedko@code.bg,*
alexx_5226@abv.bg

***Abstract:*** *this paper represents development and application of method for controlling service robot via voice commands. We created web based user interface and implemented voice recognition method, which recognizes given commands from the user and send control commands to the service robot. This method is based on Google Cloud Speech API and uses HTML, Python Django and ROS. Our goal is to achieve complex robot control based on the recognized commands from the users.*

***Keywords:*** *speech recognition, Google Cloud Speech API, HTML, Python Django ROS, robot control*

## 1. Introduction

Service robots controlled by voice command are very effective especially for physically challenged people [12]. They can also be very useful in industries and at places where human life is endangere [7]. In the home or office, robots have to be interactive. People needs assitants that can understand their wishes and bring to them different objects [8].

The advancements in technology in speech recognition over the past few years have led to increased research on user voice commands. (Some results of design

implementation can be found in home appliances operating by voice command, some design implementation of robots and forklift navigation based on voice command).( It gives exact concept of controlling a robot by voice.) Robot is capable of understanding and synthesizing human speech for communication [4].

Service robots usually provide different types of control and interaction [6]. They are controlling directly by physical joysticks. Others provide specific user interfaces controlled distantly by tablets, phones, laptops [11]. There are systems that recognizes gestures also. However, most of the people needs to speak with the robots, so the robot have to recognize their speech and to talk to them. Most of the speech recognition systems needs a powerfull algorithms and neural network learning to be able to work [14]. In addition, some robots can make their own decisions, according to the needs of their owner [5].

Speech is an ideal method for robotic control and communication. This is a kind of direct teleoperation of the robot [9]. The speech-recognition circuit we will outline, functions independently from the robot's main intelligence [central processing unit (CPU)]. This is a good thing because it doesn't take any of the robot's main CPU processing power for word recognition. The speech recognition software is speaker independent allowing any human voice to operate the robot. The special feature of the user interface is the ability of the software to allow any particular user its use without training.

We are using voice to control the service mobile robot motion by giving specific voice commands. The speech recognition software running on a PC is capable of identifying any voice commands, issued by a user. After processing the speech, the necessary motion instructions are given to the mobile platform via the embedded computer of the robot.

## 2. Speech recognition systems

To develop speech recognition system is very complicated and hard working. We realized that there are some speech recognition systems, which we can use for our robot [1], [3], [15].

- Pocketsphinx: a free, real-time continuous speech recognition system

  Pocketsphinks is a package of CMU Sphinx toolkit. PocketSphinx is a library that depends on another library called SphinxBase, which provides common functionality across all CMUSphinx projects. There are some ready to use language models but for our purposes we needed Bulgarian language, which has not a model. To develop custume model we have to build our dictionary, then to build language model. The next steps are adapting the default acoustic model and training an acoustic model for CMUSphinx [2].

- Google Cloud Speech API

Google Cloud Speech API enables developers to convert audio to text by applying powerful neural network models in an easy to use API [13]. The API recognizes over 110 languages and variants, to support your global user base. You can transcribe the text of users dictating to an application's microphone, enable command-and-control through voice, or transcribe audio files, among many other use cases. Recognize audio uploaded in the request, and integrate with your audio storage on Google Cloud Storage, by using the same technology Google uses to power its own products [15].

## 3. Application of the speech recognition

Comparing Pocketsphinx with Google API, we decided to use Google API for our speech recognition. To apply Google speech recognition API in our web based user interface we have written some script in HTML, JavaScript and Python-Django [10].

We added new block in our main page of the user interface named Mobile Base Control. That block is for controlling the mobile base of the robot. There are four buttons for movement directions and empty box with microphone icon (fig. 1). The buttons are for simple button control.

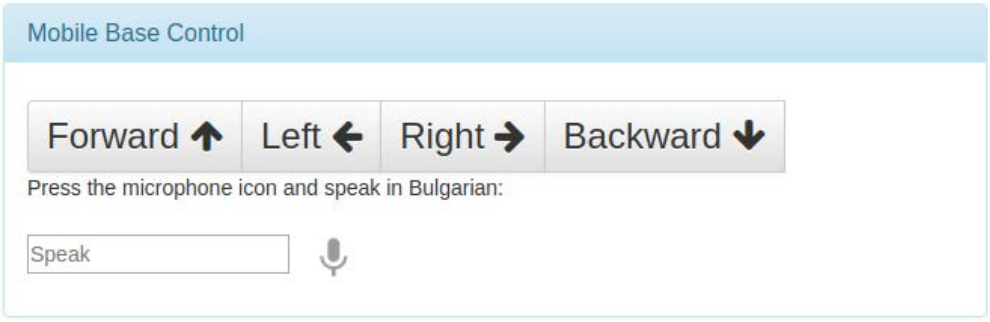


**Figure 1.** Speech icons in the User Interface

When the user wants to speak to the robot, he has to press the microphone and then to speak to the device for about 5 seconds. Google API recognizes the speech and as respond writes the recognized words in the empty box. Therefore, the user can see what the system has recognized. After that, we compare the recognized words with our ready commands for control. When there is matching the interface send control commands to the robot. In the figure 2, we represent the global communication of the user, user interface, Google API and the robot. Our system relies entirely on Google API for the recognition, then processing recognized words.

In more details, we have written the code for Google API in JavaScript and our recognition function name is 'startDictation'. The function calls 'webkitSpeechRecognition', which is the Google API and waits for the API to return result. We write the result in the empty box for the user. For comparing the result

with our phrases, we use another function ‘checkCommand’. That function is called in the ‘startDictation’ function and compare the words directly in the JavaScript.

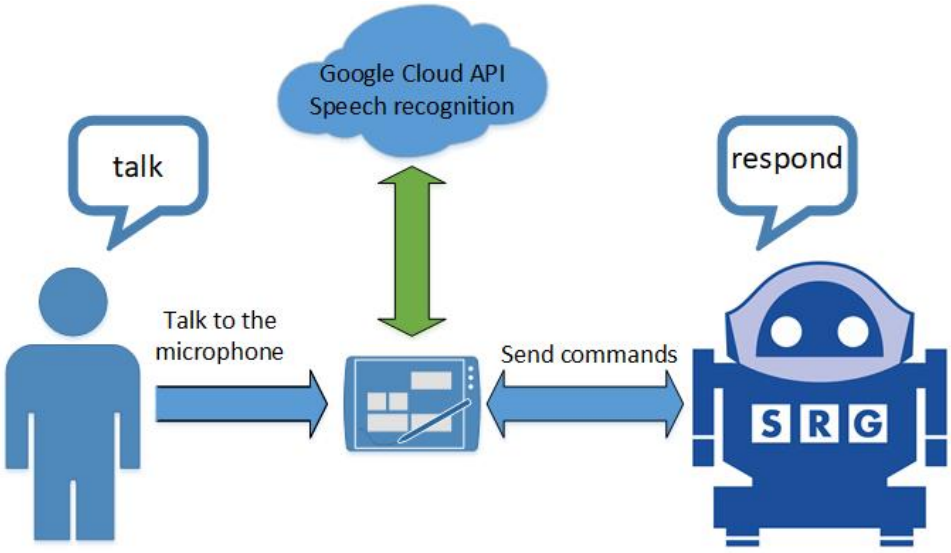


**Figure 2.** Communication - User, User Interface, Google API and Robot.

There are some additional options for our recognition. We can change the language if the recognition in our function. We can use all 110 languages that Google API supports. The only thing that we have to change is the value of the “recognition.lang = "";”. For Bulgarian the value is “bg-BG”, for English is “en-EN” etc. Of course, if we change recognition language we will have to change the language of our command phrases.

## 4. Experiments and results

To test our system, we have developed the following tests. We first checked whether the Google API successfully recognized Bulgarian. In addition, we tested voice recognition for Bulgarian language with different voices, which was very important to check that different voices were recognized. We managed to succeed almost 93% voice recognition on different voices. Then we saved four ready phrases to recognize - forward, backward, left and right. Then we checked that our program was successful in recognizing the words that were returned to us by the API. The last step is to send commands to the robot. We also monitor the performance of the system.

All systems successfully performed their tasks. Google API recognizes Bulgarian. Our functions compares the recognized words and control functions sends commands to the robot.

## 5. Conclusion

Speech communication between humans and machines is very important and applicable. The application of ready-made voice recognition systems contributes to the rapid development of voice communication between people and robots. This type of communication provide better interaction and tele-control. In future, our goal is to create semi-autonomous control of the robot by using phrases that are more complicated in our control system. This will help us to develop interactive system.


### Acknowledgments

This research was carry out as part of the project “Telecontrolled Service Robots for Increasing the Quality of Life of Elderly and Disabled, № DN 07/23 – 15.12.2016”, financed by the Bulgarian National Science Fund.


## References


1. D. Huggins-Daines, M. Kumar, A. Chan, A. Black, M. Ravishankar, A. Rudnicky, "Pocketsphinx: A free real-time continuous speech recognition system for hand-held devices" in IEEE Trans. Acoust. Speech Signal Processing, IEEE, vol. 1, pp. 185-188, 2006.
2. Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar, “A Review on Speech Recognition Technique” in International Journal of Computer Applications (0975 – 8887), Volume 10– No.3, November 2010
3. W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, A. Stolcke, “The Microsoft 2017 Conversational Speech Recognition System”, Microsoft AI and Research Technical Report MSR-TR-2017-39 August 2017
4. Richard P.Lippmann, “Speech recognition by machines and humans”, Speech Communication Journal, Volume 22, Issue 1, July 1997, Pages 1-15
5. Sidobre D. et al. (2012) Human–Robot Interaction. In: Siciliano B. (eds) Advanced Bimanual Manipulation. Springer Tracts in Advanced Robotics, vol. 80. Springer, Berlin, Heidelberg
6. R. Murphy, T. Nomura, A. Billard, J. Burke, "Human–robot interaction", *IEEE Robot. Automat. Mag.*, vol. 17, pp. 85-89, 2010.
7. A. Khamis, D.M. Rivero, F. Rodriguez, M. Salichs, "Pattern-based architecture for building mobile robotics remote laboratories", *Robotics and Automation 2003. Proceedings. ICRA '03. IEEE International Conference on*, vol. 3, pp. 3284-3289 vol.3, 2003, ISSN 1050-4729.
8. J. Stückler and S. Behnke. Integrating indoor mobility, object manipulation, and intuitive interaction for domestic service tasks. In Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids), 2009.
9. Sebastian Muszynski, Jörg Stückler, Sven Behnke, "Adjustable autonomy for mobile teleoperation of personal service robots", *RO-MAN 2012 IEEE*, pp. 933-940, 2012, ISSN 1944-9445.
10. D. Chikurtev, K. Yovchev, E. Chikurtev, Design and functionality of Web User interface for control of service mobile robot through the Internet, PROBLEMS OF ENGINEERING CYBERNETICS AND ROBOTICS, Vol 67, Sofia, 2016, p. 51-60, ISSN 0204-9848.
11. Nayden Chivarov, Denis Chikurtev, Kaloyan Yovchev, Stefan Shivarov - Cost-Oriented Mobile Robot Assistant for Disabled Care; TECIS 2015, 16th IFAC Conference on Technology, Culture and International Stability, Sozopol, Bulgaria, 24-27 September 2015, SJR – 0.263.

12. Chivarov, N.; Shivarov, S.; Yovchev, K.; Chikurtev, D.; Shivarov, N. - Intelligent Modular Service Mobile Robot ROBCO 12 for Elderly And Disabled Persons Care; Robotics in Alpe-Adria-Danube Region (RAAD), 2014 23rd International Conference on, Smolenice, Slovakia, 3-5 Sept. 2014, Print ISBN: 978-1-4799-6797-1. p. 343-348.
13. J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar, and B. Strope, ""Your word is my command": Google search by voice: A case study," in Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics, chapter 4, pp. 61-90. Springer, 2010.
14. L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in ICASSP, 2013.
15. Google cloud speech API - https://cloud.google.com/speech/?ds_rl=1245734&gclid=Cj0KCQiA38jRBRCQARIsACEqIevF-UGUJmmViBVaRyaQFzzvvUdJ1qEzNI5lNZjciHoXH8iSs788euoaAmlQEALw_wcB&dclid=CL6vmZGLidgCFUYTGwod7-sLdA

# Управление сервисным роботом с помощью голосовых команд

*Денис Чикуртев[1, 2], Ивайло Рангелов[2], Найден Шиваров[2], Недко Шиваров[2], Александър Гигов[2]*

*[1]Institute of Information and Communication Technologies – BAS*
*[2]European Polytechnical University - Pernik*

*Email: denis@iinf.bas.bg, rangelov.ivaylo@yahoo.com, nshivarov@code.bg, nedko@code.bg, alexx_5226@abv.bg*

***Аннотация:*** *этот документ представляет собой разработку и применение метода управления служебным роботом посредством голосовых команд. Мы создали веб-интерфейс пользователя и реализовали метод распознавания голоса, который распознает данные команды от пользователя и отправляет управляющие команды сервисному роботу. Этот метод основан на Google Cloud Speech API и использует HTML, Python Django и ROS. Наша цель - добиться комплексного управления роботом на основе признанных команд от пользователей.*

***Ключевые слова:*** *распознавание речи, API Google Cloud Speech, HTML, Python Django ROS, управление роботом*