Building Human Mobile Robot Audio Communication Interface with Artificial Intelligence and Deep Learning

Snejana Pleshkova-Bekiarska
Faculty of Telecommunication
Technical University
Sofia, Bulgaria
snegpl@tu-sofia.bg

Alexander Bekiarski
Faculty of French Electrical Engineering Education
Technical University
Sofia, Bulgaria
aabbv@tu-sofia.bg

***Abstract*—Voice conversation is the natural way of communication between persons and mobile robots. First of all, the person as a speaker sends usually the appropriate voice commands to navigate and control mobile robot movements. Second, having speaking ability, the mobile robot can interact with human, i.e. can speak to answer or report the person for accurate or wrong understanding and executing of human voice commands. The common way to realize the voice communication between persons and mobile robots is to build an appropriate human mobile robot audio communication interface. This is the main goal of this article and the chosen means to build the proposed human mobile robot audio communication interface are artificial intelligence (AI) and deep learning (DL). The results achieved are presented as the suitable experiments of sending voice commands to the mobile robot and the corresponding answers or reports as speech of mobile robot to inform the person whether or not the voice commands are accurately or wrong understand and executed.**

Keywords—human mobile robot audio interface, speech or voice recognition, artificial intelligence, deep learning

I. Introduction

Without voice or speech recognition it is not possible to have a cognitive interfacing of mobile robot with human. The general bases of building an appropriate human mobile robot audio communication interface are the methods and algorithms for speech or voice recognition. The only difference between "speech" and "voice" recognition is the ability of voice recognition not only to "understand", i.e. to "recognize" the spoked from a person words or phrases, but also to identifier the personality of the speaking person. There are a lot of the methods and algorithms for speech or voice recognition [1], [2], most of them have the real practical implementations in robotics mobile robot human interfaces [3], [4]. But some of the main disadvantages of these methods and algorithms are computational complexity and corresponding difficulties for real time execution. Therefore, the goal in this article is to overcome these disadvantages. This is done with the proposition to apply in the building of human mobile robot audio communication interface the advantages of the artificial intelligence (AI) and neural networks with deep learning (DL) in area of speech and voice recognition.

II. Block Schema of the Proposed Human Mobile Robot Audio Communication Interface

In Fig. 1 is presented the block schema of the proposed human mobile robot audio communication interface developed with the usage of AI and neural networks with DL. It allows the person to send voice command to mobile robot and mobile robot to be able to answer the person for voice command execution. Voice conversation start when the person begin to speak sending voice command to control of mobile robot motion.

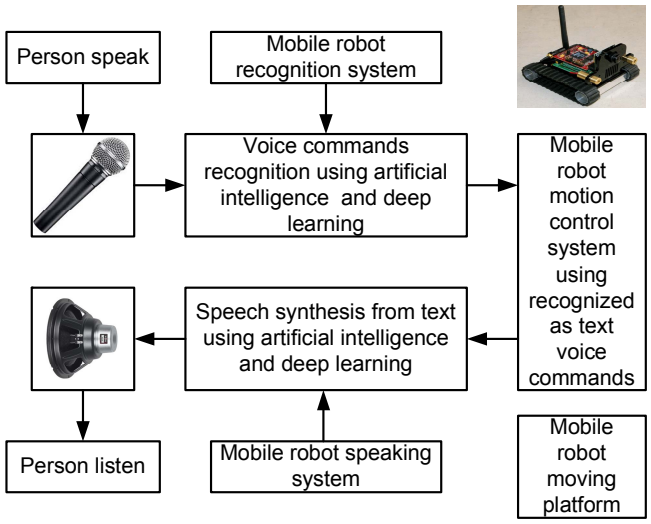

Fig. 1. Block schema of the proposed human mobile robot audio communication interface developed with usage of AI and neural networks with DL

Human voice is received and transformed as sound signal of the voice command from a microphone (Fig.1) mounted on the mobile robot moving platform. From output of the microphone the received sound signal of the voice command is entered in the next block (Fig.1). In this block are applied suitable algorithms for voice command recognition, based on AI and neural networks with DL. The result from recognized voice command, in the form of text, is send in the next block - the mobile robot motion control system (Fig.1), responsible for execution of mobile robot current movement. The information for mobile robot right or wrong understanding and executing voice command is output in the form of text from the block - the mobile robot motion control system (Fig.1). This text is transformed as the speech in the next block (Fig.1) using speech synthesis in text to speech algorithm, based on AI and neural networks with DL. The synthesized speech is output from the loudspeaker (Fig.1) as "the mobile robot voice". In this way the person can observe and control the execution of the issued from him voice

978-1-7281-3467-3/19/$31.00 ©2019 IEEE

command not only as seen the right or wrong mobile robot motion, but also as he is listening "the mobile robot voice". This means, that "the mobile robot voice" is as a voice response of the mobile robot to the person whether the mobile robot has correctly understood the voice command sent by the person and whether the mobile robot has correctly performed the current operation of movement, according to this voice command. The main parts of the presented in Fig. 1 block schema of the proposed human mobile robot audio communication interface are the algorithms for voice commands recognition and speech synthesis, which are the subject of the detailed development in this article with the help of AI and neural networks with DL.

III. Vocie Commands Recognition Algorithm in the Proposed Human Mobile Robot Audio Communication Interface Based on Artificial Intelligence and Deep Learning

The voice command recognition and transformation to text can be considered as concrete application of speech to text methods and algorithms [5]. The only difference is of the limited word and phrases as voice commands to the mobile robot, instead of being used largest vocabulary and sentences analysis in general cases of speech to text applications. Therefore, in this article is proposed to use the existing general speech to text recognition algorithms, but modifying them in the following two ways:

- using real issued voice commands from concrete persons as speakers of voice commands to mobile robot motion navigation and control;
- creating, in the phase of learning, a limited vocabulary or data base only consisting from words and phrases like "go", "go forward", "go back", "turn left", "turn right", etc., and then to use this limited vocabulary or data base in the phase of testing the accuracy of voice recognition and in the phase of real execution of mobile robot motion control with the recognized voice commands.

The basic steps of the proposed voice commands recognition algorithm are shown in Fig.2, as follow:

- enter sound signal from microphone mounted on mobile robot moving platform as the input signal for the algorithm of voice commands recognition;
- to start the algorithm to learn human mobile robot audio interface for voice commands recognition using artificial intelligence and deep learning and creating voice commands data base;
- to execute algorithm for testing the ability and accuracy of human mobile robot audio interface for correct voice commands recognition using artificial intelligence and deep learning algorithm and created voice commands data base;

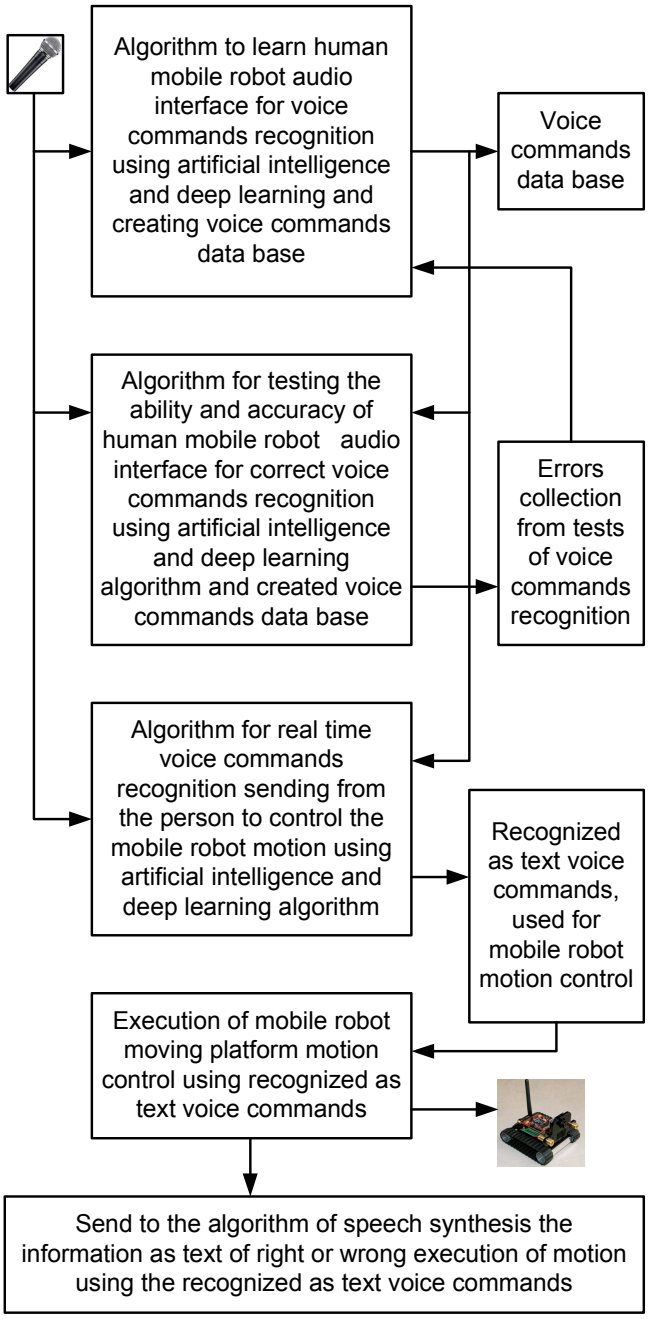

Fig. 2. The proposed voice commands recognition algorithm in human mobile robot audio communication interface developed with usage of AI and neural networks with DL

- to collect the errors from tests of voice commands recognition and if the number of errors exceeds a chosen threshold, then return to the previous step to do the supplementary learning of human mobile robot audio interface for voice commands recognition;
- to execute in real time voice commands recognition algorithm, transforming them as text to control with them the mobile robot motion using artificial intelligence and deep learning;
- to enter the recognized as text voice commands in mobile robot moving control system to form the necessary for the wheels of moving platform of the mobile robot controlling signals;

- to execute the motion of mobile robot moving platform using the recognized voice command in form of text;

- to calculate and collect the information of right or wrong, i.e. the errors, from execution of motion from moving platform of mobile robot using the issued from person and recognized as text voice command;

- to send as text to the speech synthesis algorithm the information for right or wrong, i.e. the errors, from execution of motion from moving platform of mobile robot using the issued from person and recognized as text voice command.

IV. Speech Synthesis Algorithm in the Proposed Human Mobile Robot Audio Communication Interface Based on Artificial Intelligence and Deep Learning

The information, send as text from the presented in Fig.2 voice commands recognition algorithm, for right or wrong execution of motion from moving platform of mobile robot is used in the speech synthesis algorithm, shown in Fig.3, of the proposed human mobile robot audio communication interface. It is proposed this speech synthesis algorithm to be based also on usage of AI and neural networks with DL.

The basic steps of the proposed speech synthesis algorithm are shown in Fig.3, as follow:

- to receive as text the information, entered as text from voice commands recognition algorithm, of right or wrong execution of motion using the recognized as text voice commands;

- to do speech synthesis as text to speech algorithm based on artificial intelligence and deep learning;

- to output the synthesized from text speech to the mobile robot audio speaking system (Audio amplifier with loudspeaker) , to be hear by the person;

- person to hear the mobile robot speech;

- person to decide whether or not to answer of the mobile robot speech, repeat the voice command or send new type of voice command, which to be received from the microphone, mounted on mobile robot moving platform.

V. Experimental Test and Results from Development of the Proposed Human Mobile Robot Audio Communication Interface Based on Artificial Intelligence and Deep Learning

The presented in Fig.2 and in Fig.3 algorithms for voice commands recognition and speech synthesis, as the main parts in development of human mobile robot audio communication interface with usage of AI and neural networks with DL, are subjects of experimental implementations and tests.

There are chosen and applied, in the carried out experiments and test, suitable functions, tools and embedded libraries in standard programming languages dedicated to voice recognition and speech synthesis in area of AI and neural networks with DL. As programming language is chosen the popular in area of AI and neural networks with DL programming language Python [6]. In the Python packages “SpeechRecognition” are included suitable speech recognition functions based on AI and neural networks with DL. In the developed here human mobile robot audio communication interface are used some of these speech recognition functions and also a limited part, only words and phrases like “go”, “go forward”, “go back”, “turn left”, “turn right”, etc. from the big vocabulary or data base existing in Python packages “SpeechRecognition”.

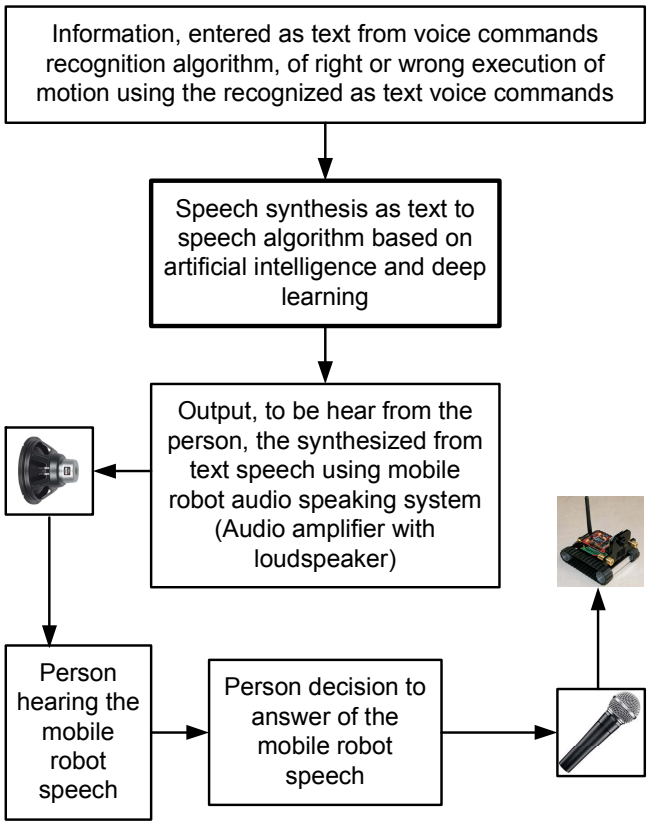

Fig. 3. The proposed speech synthesis algorithm in human mobile robot audio communication interface developed with usage of AI and neural networks with DL

The following functions from the suitable Python packages “SpeechRecognition” and “PyAudio” in area of AI and neural networks with DL are used in the experiments and tests of the proposed in Fig.2 algorithm for voice commands recognition as the part of the development of human mobile robot audio communication interface:

- to include SpeechRecognition Package using the Python function “import speech_recognition as sr”;

- to set microphone mounted on mobile robot moving platform using “PyAudio” function to define microphone as object “mic=sr.Microphone”;

- to define “r=sr.Recognizer()” as object “r” using class Recognizer(), from imported as “sr” SpeechRecognition Python Package;

- using the “PyAudio” function “audio= r.listen(source)” to start the mobile robot to “listen” for the voice command issued from the person;

- the following functions inform the person that the mobile robot “wait” and “listening”, i.e. is ready to receive the voice command from the person:

with sr.Microphone() as source:

print (“Say something!”)

audio= r.listen(source).

The listed above main functions are embedded as execution voice recognition program in the developed human mobile robot audio communication interface and the

results of right work and correct voice commands recognition from mobile robot are demonstrate briefly with the following examples:

- if, for example the person say “go left” as voice commands to mobile robot “to turn in the left”, then the results of correct recognition and understanding from mobile robot of these examples of voice commands are shown in Fig. 4 (above);

- if, for example the person say “go forward” as voice commands to mobile robot “to go forward”, then the results of correct recognition and understanding from mobile robot of these examples of voice commands are shown in Fig. 4 (below)

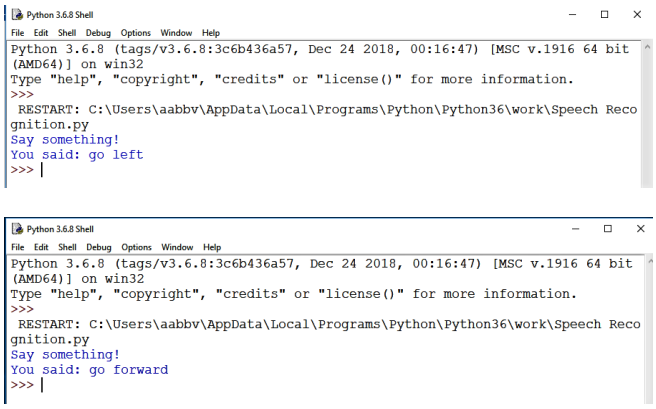

Fig. 4. The results of correct recognition and understanding from mobile robot of these examples of voice commands “go left” and “go forward”, which are send from a person to mobile robot using the developed human mobile robot audio communication interface with embedded in Python tools of AI and neural networks with DL

From Fig.4 it is seen that the result from voice commands recognition is in the form of text. The text is converted as speech using speech synthesis as text to speech algorithm based on AI and neural networks with DL as is outlined in Fig.1 and Fig.3 for the proposed in this article human mobile robot audio communication interface.

The following functions from the suitable Python packages in area of AI and neural networks with DL are used in the experiments and tests of the proposed in Fig.3 algorithm for speech synthesis as the part of the development of human mobile robot audio communication interface:

- to include Python based on AI and neural networks with DL speech synthesis Package gTTS (Google Text to Speech) using the Python function “from gtts import gTTS”;

- to transfer the result from voice commands recognition in the form of text using function (my_tts = "go left"), for example, if the recognized voice command is "go left" or using function (my_tts = “go forward”), if the recognized voice command is “go forward”;

- to execute function tts = gTTS(text=my_tts, lang='en') of Python speech synthesis Package gTTS (Google Text to Speech) based on AI and neural networks with DL;

- the results of speech synthesis are input for mobile robot audio speaking system (Audio amplifier with loudspeaker) to be hear by the person as mobile robot speech answer for right voice recognition, for example of the send from person voice commands “go left” or “go forward”;

- to save the synthesized speech in two files “go left.mp3” and “go forward.mp3” for to be shown in this article as the time diagrams, demonstrating the right synthesized speech from the text examples “go left” and “go forward” of recognized voice commands .

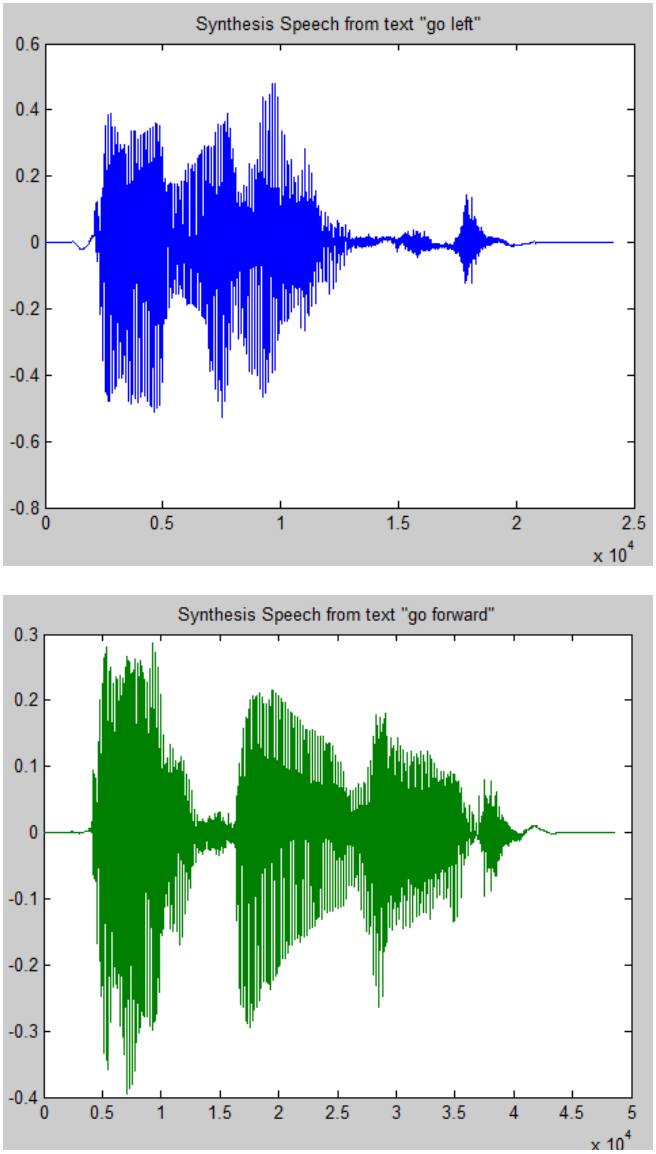

Fig. 5. The results of correct recognition and understanding from mobile robot of these examples of voice commands “go left” and “go forward”, which are send from a person to mobile robot using the developed human mobile robot audio communication interface with embedded in Python tools of AI and neural networks with DL

VI. Conclusion

It can be concluded from tests that the proposed human mobile robot audio communication interface work correct in execution of voice commands recognition and speech synthesis based on AI and neural networks with DL. The mobile robot can receive, recognize, understand and execute issued from the person voice commands and answer with synthesis speech to the person of right or wrong understanding and execution of voice commands in mobile robot motion control tasks. Therefore, the proposed in this article human mobile robot audio communication interface,

based on AI and neural networks with DL, will be used in the future researches to estimate and analyze the accuracy achieved in navigating and control motion of mobile robot using this human mobile robot interface for intelligent interaction between person and the mobile robot.

REFERENCES

[1] D. Jurafsky, and J. H. Martin, "Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition," Stanford University, Third Edition Draft of September 23, 2018.

[2] M. R. Saady, H. El-Borey, El. A. El-Dahshan, and A. Sh. Yahia, "Stand-Alone Intelligent Voice Recognition System," Journal of Signal and Information Processing, vol. 5, pp. 179-190, 2014.

[3] K. Park, S. J. Lee, H. Y. Jung, and Y. Lee, "Human-robot interface using robust speech recognition and user localization based on noise separation device," RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication, 27 September - October 2009.

[4] St. Heinrich, and St. Wermter, "Robust Speech Recognition for Human-Robot Interaction," Proceedings of the IROS2011 Workshop on Cognitive Neuroscience Robotics (CNR), pp. 29-34, San Francisco, CA, USA, September 2011.

[5] V. C. Rajyaguru, "Different Methods Used In Voice Recognition Techniques," International Research Journal of Engineering and Technology (IRJET), vol. 3, July 2016.

[6] D. L. Poole, and Al. K. Mackworth, "Python code for Artificial Intelligence: Foundations of Computational Agents," Version 0.7.6 of January 19, 2019.