Chang, M. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

Learning Conversation with a Mobile Robot

Wing-Kwong WONG*, Wen-Jia KUO , Ting-Sheng WEI, Kai-Ping CHEN
Depart. of Electronic Engin., National Yunlin University of Science and Technology, Douliu, Taiwan
*wongwk@yuntech.edu.tw

Abstract: This study proposes to design a mobile robot to assist language learning. This should motivate learners to practice speaking and listening in order to improve their English skill. This project provides a low-cost robot for all students who are willing to spend time learning English inside and outside the classroom. This research proposes to deliver a robot with wheels, cameras, microphones and speakers. Raspberry Pi 3 and Arduino provide the computation power to interact with and control the robot. The robot converses with English learners to accomplish specific tasks. In one task, a learner asks the robot to travel on a street map with simple roads and locations such as a library, a park, and an elementary school. The utterances of the learner are recognized with a Google cloud service, and the voice response of the robot is generated with text-to-speech technology. When the robot walks on a street map, it does image processing to follow lanes and recognize intersections and a number of prescribed locations. An empirical experiment was run to study the learning process of learners. Some interesting results were found.

Keywords: Robot-assisted Language Learning, Language Teaching, Mobile Robot, Image Processing, Speech Recognition

1. Introduction

With the huge political and economic influence of English-speaking countries, English has become tone of the most important communication languages in the world. According to statistics, "English is the most widely learned second language and is either the official language or one of the official languages in almost 60 sovereign states. [...] It is estimated that there are over 2 billion speakers of English".[1] English is the main language in the fields of books, music, international business, science, air traffic control and news media. In a transnational, cross-cultural society, English language skills are even more important. Therefore, in today's globalized society, international language—English—has become a capability that should be learned.

Most people in Taiwan also regard English as the first foreign language. In the syllabus of the 12-year Basic Education issued by the Ministry of Education, English is one of the major subjects. Students must learn English from the third grade of the Elementary School.[2] However, English is a big obstacle for many students. This maybe due to the fact that English is not our native language, and there is no good English learning environment. As a result, many students fail to achieve satisfying results and reject learning it (Shih, 2016). The anxiety of English communication is a well-known problem for the learners. Listening and speaking have been relatively weak, making students afraid to speak English in English classes. A survey by Hamouda (2013) suggests that the anxiety of English learners might be due to the following reasons: low English proficiency, fear of speaking in front of others, fear of making mistakes, negative evaluation, shyness and lack of confidence and preparation.

Robot-Assisted language learning provides a novel approach and has gained more and more attention as robot technologies become more mature and available at lower cost. The development of hardware and the advancement of technology, including artificial intelligence and machine learning, have made this teaching method more feasible. Robot-assisted language learning might offer several

1 English language, at https://en.wikipedia.org/wiki/English_language (last visited 04/20/2019).

2 National Academy for Educational Research, at https://english.naer.edu.tw/ (last visited 04/08/2019).

advantages over traditional teaching. For example, for teaching that requires repeated practice, robots are not as tired and impatient as teachers due to long hours of teaching, and the students have less fear of getting negative feedback from others. Moreover, in robot-assisted language learning, the data of learners can be stored and used as a reference for teaching.

There have been many related studies on the research and development of language robots. In Turkey, Kose (2011) did a study on sign language teaching for 106 preschool children using the NAO H25 humanoid robot. The purpose was to assess children's ability to learn sign language from robots. The study was very successful and the children achieved a high percentage of correctness in a test. In South Korea, Lee (2011) designed an English course to let students interact with robots. After post-testing and comparison, it was found that although there was no significant improvement in listening, the improvement in speaking was significant. It also increased student satisfaction, interest, self-confidence and motivation. In recent years, there are similar research projects in Taiwan. Huang (2017) explored the impact of using English robots on the effectiveness of learning English for young children. Kang (2017) also did a study on the effectiveness of robot-assisted language learning for Taiwanese primary school children's language learning. This result indicated that the students in the experimental group were more focused, confident and had more positive attitude than the students in the control group. It can be known from the above research results that robots are used in secondary language learning, and robots can have better learning outcomes for students than traditional teaching methods.

A goal of this research is to evaluate our design of activities of robot-assisted language learning. In our design, students use English to communicate with robots. After voice recognition, the robot responds verbally or executes the task according to the instructions. The task location is on a simplified street map. Students can order the robot to move to a designated location on the map, or command the robot to go forward, turn right, or turn left. Through this oral and visual interaction with the robot, students can learn English in a different way than before. The evaluation results of the learners are discussed in this paper.

2. Dialog Mechanism and Image Processing of Robot

2.1 Dialog Process

The dialog between the user and the robot is controlled according to a flowchart implemented as a Python program (Figure 1). When a learning session starts, the robot constantly listens for a wake-up word from the user. In our case, the wake-up word is “YunBot”, which is the name of the robot. When the robot hears the wake-up word, the sound input of the next five seconds will be recorded and sent to a speech recognition service of Google Cloud, which will return the recognized text. For now, there is an assumption of short input utterances of less than five seconds. The recognized text is matched with a number of prescribed inputs. Once a match is found, a response text is converted into a voice output.

According to the different dialogues, it can be divided into three types. The general dialogue does not need to do other things. If it is a place dialogue, it will enter the next state. First calculate the path to the specified location, and then find out whether there is an ArUco tag from the captured image. If there is, determine whether the tag ID is the specified location ID, and if so, end the move and return to the conversation state. Next, judge whether you encounter an intersection, because after the intersection, you should turn according to the path calculated beforehand, and then follow the lane to move. If the dialogue used is a mobile conversation, it is judged whether the direction indicated by the user has a lane to walk, and if it is possible, the movement starts. It is necessary to capture the image to make the judgment of the ArUco tag at the intersection and location. If it is found, return to the conversation state, wait for the next user's instruction, and if not, follow the lane movement.

Among them, the path of the street map, the calculation of the moving path, the detection of the ArUco tag, the detection of the intersection, the calculation of the lane image, etc., the data (image) is transmitted to the computer. Because the processing speed of the computer is much faster than that of the Raspberry Pi, the Raspberry Pi is prevented from being overloaded and the reading frame rate is reduced. The rest is handled on the Raspberry Pi.

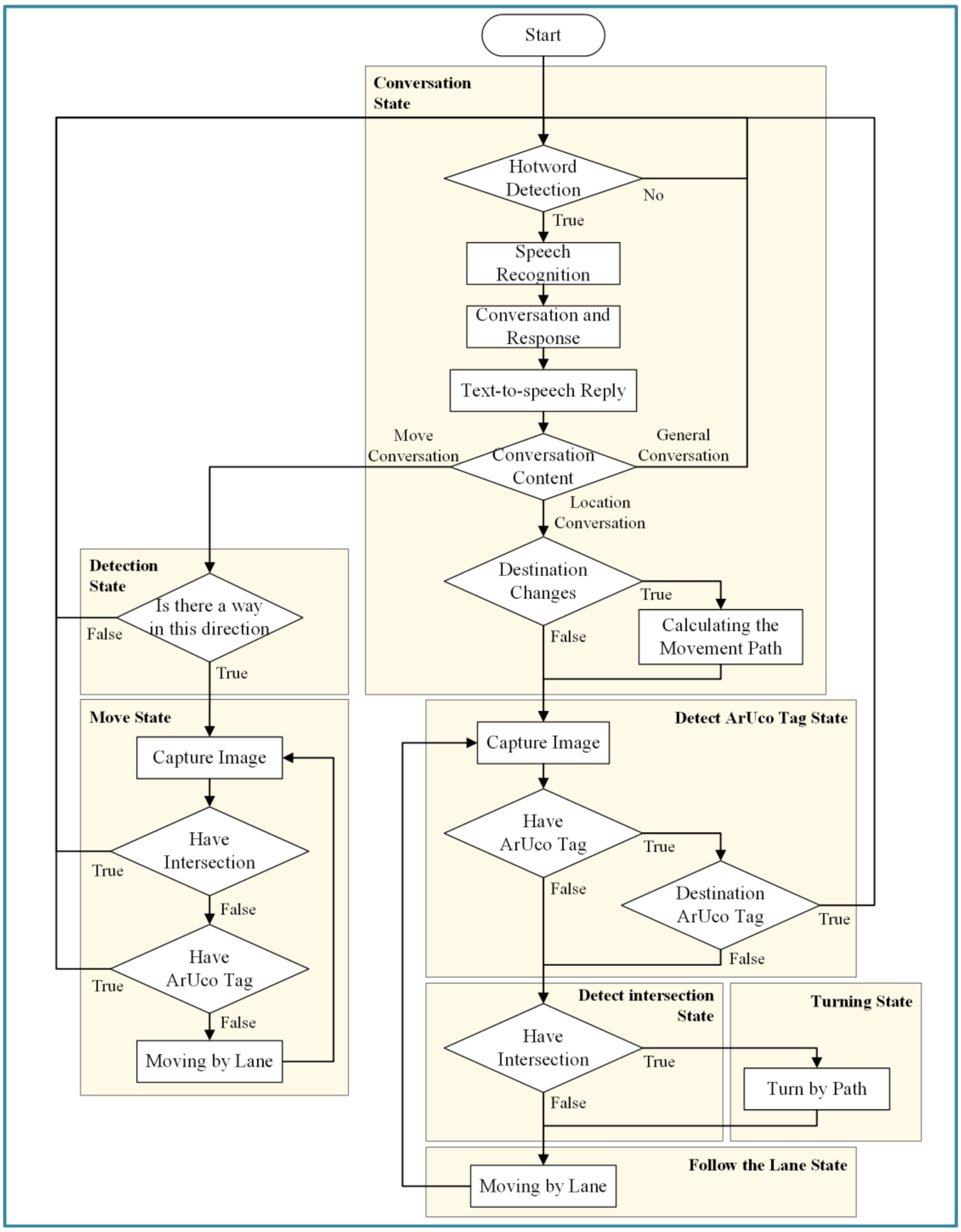

Figure 1. Robot program overall flow chart

2.2 Dialogue and Response Mechanism

The dialogue and response mechanism works as follows. First, the previous input string of the user is saved for subsequent decision making, and then the robot waits for the user's input. After an utterance is received and converted to a text string with Google cloud service of speech recognition, the string is preprocessed. String preprocessing involves removing punctuation and duplicate white spaces, and converting all English to uppercase to make it easier to match to target strings. This completes processing the user's input.

In the response stage, the string of the previous response is also saved at the beginning, and the state is set to "TO UNDERSTAND". Then the robot determines whether the user enters a null value or repeatedly enters a null value. If so, the state is set to "NULL INPUT" or "NULL INPUT REPETITION" accordingly. In both cases, the robot reminds the user that no input is received so that the user knows that he should say something to the robot.

At this point, the robot must have obtained some input string from the user. Next, the input string is matched to some templates in the language database. If no match is found, the state is set to

"BOT DOESN'T UNDERSTAND" and the robot picks one expression from several choices all of which mean that the robot does not understand the user's utterance. Otherwise, the matched template generally provides multiple choices of responses. Then one choice is randomly selected and the choice would be pre-processed. For example, if the user says "YOU ARE MY ROBOT", a matched template can be "YOU ARE $REMAIN_STRING". One response of this template is to prepend "SO, YOU THINK THAT I'M" to $REMAIN_STRING, resulting in an utterance "SO, YOU THINK THAT I'M MY ROBOT." This sounds illogical because of the clause "I'M MY ROBOT," which should be "I'M YOUR ROBOT." This correction can be achieved if there is a check on whether there are pronouns in any of the input string. Any first person pronoun is replaced by a second person pronoun and vice versa. For example, "I" is replaced by "YOU", "MY" is replaced by "YOUR", etc. In addition, any subsequent be-verb following the pronoun should be changed accordingly. For example, "AM" is replaced by "ARE". In the previous example, the response statement becomes "SO, YOU THINK THAT I'M YOUR ROBOT." Then the robot determines whether the response has been used in the last 10 responses. If yes, the current response is replaced by another random choice provided by the matched template. Otherwise, the conversation record is saved, and the response is uttered. Table 1 lists some of the user's utterances and the possible response of the robot.

All dialogue can be roughly divided into three types. The first one is a general conversation that requires no motion from the robot. The robot only needs to find the corresponding response and convert it into voice output. It does not need to control other hardware, like "What can you do?", "Nice to meet you." In order to avoid repeating the same response, which might bore the student, one of the equivalent responses is chosen at random. For example, the student says "How are you", the robot may respond "Fine, thank you.", "I'm very well.", "Pretty good! ", or "I'm doing OK.", and if the selected response was used before, another random choice of response will be selected.

Table 1

General Conversation

Student talk	Robot choices of responses
General dialogue	
How are you?, What's up?, Are you alright?	Fine, thank you./I'm very well. Thank you./Pretty good!/Never better!/So far so good./I'm doing OK./Just so-so./Not bad./As usual.,/Nothing special.
What's your name?	My name is Yunbot./I am Yunbot.
Nice to meet you.	Nice to meet you too/Thank you. It's very nice to meet you as well.
Location dialogue	
Go to the School.	Ok, go to the School. (Action: Move to the School.)
Go to the Library.	Ok, go to the Library. (Action: Move to the Library.)
Move dialogue	
Go forward.	Ok, Go forward. (Action: Go forward.)
Turn right.	Ok, Turn right. (Action: Turn right.)

The second type is about asking the robot to get to a destination. For example, if the user says "Go to the school.", the robot will answer "OK, go to the school.", the destination of school will be noted by the robot, which then plans the path from the current location to the school. After the planning is done, the robot starts moving again. When moving, the robot will continuously recognize the boundary lines of the lane on the ground through a camera, so that it will try to stay inside the lane. When it comes to an intersection, it will stop and then turn to the correct direction according to its plan, and finally will arrive at the destination. Upon arrival, the robot will say "I have arrived at the school."

The last type of dialog is an interactive, step-by-step instructive conversation. When the user commands the robot to move, it will respond verbally and then move according to the command. When the robot visually detects an ArUco tag, which specifies a location, or a road intersection, the robot will stop. At an ArUco location, the robot will say it has arrived at the location. At an intersection, the robot will say "I am at an intersection. What should I do now?" The user can also ask the robot to go forward, or turn left or right. Then the robot will determine if there is a road in the stated direction. If yes, then it moves to that direction. Otherwise, it will say there is no road and waits for further instruction. The process repeats until it reaches the destination.

2.3 Image Processing

When people see, closer objects look bigger and further objects look bigger. This is the same with a camera. In order to recognize the two white lines bounding a lane with a camera, the images must be preprocessed with perspective transformation, resulting in images of resolution 420*240 where the two lines are parallel (Figure 2).

Next, the white lanes are separated from the black background with color filtering. Because the three channels of the original RGB color space are not very stable with different levels of brightness and saturation, color filtering is difficult with RGB images. So the RGB color space is first converted to the HSV color space, which includes hue (H), saturation (S), and brightness value (V). With HSV space, color filtering can be done more effectively.

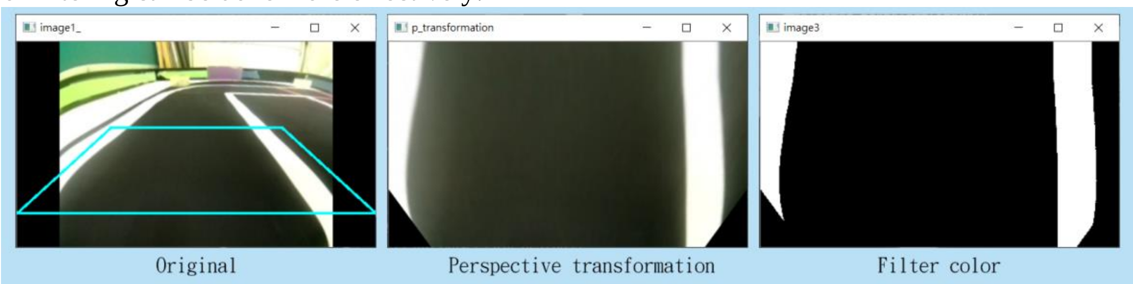

Figure 2. Perspective transformation and Filter color result

After perspective transformation and color filtering, the images are further processed to find the left and right boundary lines of the lane on which the robot moves. The robot always tries to stay in the center of the lane. Another task is the detection of lane curvature with image processing. Upon detecting a curvature, the robot turns to move along the curvature, aiming to stay within the two bounding lines,.

3. Empirical Results

In order to test the integrity of the study, the system was tested by five subjects, who were graduate students in the laboratory. According to the three dialogue modes explained above, there were three tasks: general conversation where the robot did not move, a command to ask the robot to move to a destination, an interactive, step-by-step dialog commanding the robot whenever it stopped at an intersection or a location on the map.

In the empirical experiment, there were several problems. Sometimes a subject forgot to wake up the robot first before talking to the robot. Moreover, because there were only few samples recorded for the wake-up word “YunBot”, different accents of the subjects caused the failure of speech recognition by the robot. After the wake-up word was recognized, the robot would only record the subject’s utterance for five seconds, which was not be enough for longer utterances.

For the general dialogue task, there was sometimes a problem of low volume from the subject so short utterances sometimes failed to be recognized. For example, the sound of “Hello” was recognized as “Hole”. On the contrary, longer utterances such as "How are you?" were usually recognized correctly. Another problem was the small size of the current dialog templates so that some utterances of the subjects were not found in the prepared templates.

The move-to-destination dialog task was relatively easy because only one command was needed from the subject. The interactive, step-by-step dialogue task was easy since the subjects used only three kinds of instructions: go forward, turn left, and turn right.

4. Conclusion and future work

The main goal of this research is to build a robot to assist language learning in a game-like manner involving motion from the robot, where communication can break down and mistakes can be made by both the robot and the subject. There are three modes of conversation: general dialogue, command to move to a destination, and interactive, step-by-step instruction dialogue. In a preliminary empirical experiment, five subjects of graduate students conversed with the robot. Despite a number of problems during their conversation, the subjects found the learning experience novel and interesting.

After the experiment, the subjects made a few suggestions. First, they wished the robot could provide hints on what utterances the subjects could say at various situations. The subjects wished there could be more instructions on how the robot moved, for example, "move backward". Overall speaking, practicing dialog with a robot in the above scenarios were novel and interesting to the subjects, compared to the traditional ways of learning.

In future work, more artificial intelligence and deep learning techniques can be employed to improve the performance of image detection and speech recognition. More interesting tasks should be designed for children so that the robot can be used in real language learning sessions in elementary schools. Like it or not, robots are going to play more roles in real life, in addition to language learning.

Acknowledgements

The authors are grateful to the support of the Ministry of Science and Technology, Taiwan under grant MOST 108-2511-H-224 -007 -MY3.

References

Hamouda, A. (2013). An Exploration of Causes of Saudi Students' Reluctance to Participate in the English Language Classroom. International Journal of English Language Education, 1(1). doi:https://doi.org/10.5296/ijele.v1i1

Huang, Ta-Yang (2017). A Study on Robot-based English Learning for Infancy/探討運用英語機器人對於幼兒學習英語成效之影響. Department of Software Engineering and Management, National Kaohsiung Normal University. Kaohsiung. Retrieved from https://hdl.handle.net/11296/897nf8

Jeonghye, H., Miheon, J., Vicki, J., & Jun, J. H. (2008). Comparative Study on the Educational Use of Home Robots for Children. Journal of Information Processing Systems, 4(4), pp. 159-168. doi:10.3745/JIPS.2008.4.4.159

Kang, Shih-Min (2017). Effects of Using Robot Assisted Language Learning for Promoting Taiwanese Young EFL Learners' Intonation Acquisition/機器人輔助語言學習對台灣國小學童語調習得成效之研究. Graduate Institute of Children's English, National Changhua University of Education. Changhua County. Retrieved from https://hdl.handle.net/11296/h9a9hf

Kose, H., & Yorganci, R. (2011). Tale of a robot: Humanoid robot assisted sign language tutoring. 2011 11th IEEE-RAS International Conference on Humanoid Robots, (pp. 105-111). Bled.

Lee, S., Hyungjong, N., Jonghoon, L., Kyusong, L., Gary, L. G., Seongdae, S., & Munsang, K. (2011). On the effectiveness of Robot-Assisted Language Learning. ReCALL, 23(1), pp. 25-58. doi:10.1017/S0958344010000273

Shih, Yo-Ting (2016). A Study of English Learning Motivation and Learning Disturbances: a theoretical perspective/英語學習動機與學習困擾:理論的觀點. Taiwan Educational Review Monthly, 5(7), 111-128

WORKSHOP 16 - The 2nd Workshop on Innovative Technologies for Enhancing Interactions and Learning Motivation