

Results of the Tokuolab team in the preliminary round of the Dialogue Robot Competition 2023

Kai Hisada, Yuna Miyazaki

Abstract—We propose a speech-based dialogue system to perform tasks in the Dialogue Robot Competition 2023. This system consists of various modules, including speech recognition, speech synthesis, image recognition, and robot control, along with our developed dialogue control program. For dialogue control and robot facial expression manipulation, we adopt ChatGPT’s dialogue model and an image recognition module, respectively, with the expectation of fostering empathetic conversations. Although the performance of this system in the Dialogue Robot Competition 2023 did not meet the expected outcomes, it revealed new challenges. Therefore, we provide an analysis and propose solutions in light of these findings.

I. INTRODUCTION

The Dialogue Robot System Competition 2023 is focused on the development of a spoken dialogue system that can handle given tasks while embodying hospitality. In addition to the previously held first and second editions, the details of this competition are outlined in papers [1] and [2], respectively.

The assigned task involves the travel plan decision task for a travel agency, and the goal is to implement a system that can achieve both this task and hospitality. The system utilizes the ChatGPT’s GPT model for dialogue control and employs speech recognition and image recognition modules for controlling the robot’s facial expressions. This is expected to enable the dialogue robot to exhibit empathetic behaviors towards the user.

II. COMPONENTS

The programs used in this project consist of a speech recognition program, a speech synthesis program, an image recognition program, and a robot control program provided by the Dialogue Robot Competition 2023 organizers, as well as a dialogue control program created by our team. Our program is responsible for dialog control and facial expression control of the interactive robot.

A. Interactive control

For the dialogue control, our team used the gpt-3.5-turbo model of ChatGPT to interact with the user.

First, a speech recognition program provided by the organizers of Dialogue Robot Competition 2023 receives the user’s speech text streaming from the microphone input. Next, when the program receives the user’s text, it executes a program that calls ChatGPT’s API in the main loop with the text as an argument, prompting ChatGPT to ”Answer the user’s text + up to 80 characters”. The prompt ”user text + answer within 80 characters” is given to ChatGPT. When ChatGPT responds, the response is given as an argument to

the text-to-speech function provided by the Dialogue Robot Competition 2023 organizers, and the robot speaks.

B. Facial Expression Control of Interactive Robots

We constructed a system that allows an interactive robot to change its facial expression according to the user’s facial expression.

First, an image recognition program provided by the organizers of Dialogue Robot Competition 2023 receives information about the user’s facial expression, facial orientation, age, and gender from the camera image input via streaming, and receives information about the user’s emotions from the recognized information in the main loop receives the information about the user’s emotion from the recognized information. The robot’s facial expression changes based on the user’s emotion. We designed the robot so that if the user’s emotion is happy, the robot’s expression becomes happy; if the user’s emotion is sad, the robot’s expression becomes sad; and if the user’s emotion is angry, the robot’s expression becomes sad.

By doing so, we expect the interactive robot to behave empathetically on the surface toward the user.

III. OPERATION RESULTS

Our team’s results in the preliminary round of Dialogue Robot Competition 2023 were as follows: the user had trouble starting the dialogue, the dialogue robot answered not only with information about the specified region but also about other regions, which was contrary to its main purpose, the dialogue robot did not answer in many cases, and the facial expressions of the dialogue robot did not change. In many cases, the interactive robot did not respond, or it took longer than expected to respond, and the facial expressions of the interactive robot did not change.

IV. DISCUSSION

We discussed the results of our team’s dialogue robot in the preliminary round of the Dialogue Robot Competition 2023. We summarize the causes of the four major problems that occurred and our thoughts on how to solve them.

First, as for the trouble experienced by the user at the start of the dialogue, the cause may be that the robot’s speech at the start of the dialogue was not in line with the purpose of this dialogue. This problem can be improved by improving the content of the robot’s speech at the startup (e.g., ”Please tell us about a sightseeing spot in Kyoto that interests you” to raise the topic of conversation).

Second, the dialogue system did not correctly answer the

travel destination information because it did not incorporate information from the travel destination database. The solution is to embed ChatGPT so that the system can use a vector of travel destination data, evaluate the similarity and relevance of the desired travel destination information in the user's utterance, and output the correct travel destination information.

Third, as for not receiving a response or taking a long time to respond, the causes include the fact that timeout processing was not performed when the API call was made and that when receiving speech text via streaming, if the received text was blank, the API call was not processed, and the response was interrupted in the middle of the call. The solution was to perform a timeout process. The main solution is to perform timeout processing.

Fourth, the facial expressions did not change. The issue stems from a system freeze caused by repeatedly sending commands to change facial expressions. The design of the main loop involved updating facial expressions every time the user's expression was received, potentially leading to a continuous stream of commands to change expressions while they were already in progress. As a solution, it is proposed to limit facial expression changes to only occur at the onset of the user's speech and when the dialogue robot begins speaking, as an example.

V. CONCLUSION

In the preliminary round of the Dialogue Robot Competition 2023, our team's interactive robot received an average impression score of approximately 3.3889 and an average plan evaluation score of 0.5625 in the nine items of the experience questionnaire. The overall ranking of all teams was 9th out of 12 teams. The results of the individual questionnaires do not indicate that the interactive robot provided a good experience for the users.

Based on the results of this qualifying round and the points raised as problems and proposed solutions, we hope to find and solve further problems to further satisfy users and provide better services to dialogue systems and people

ACKNOWLEDGMENT

We would like to express our sincere appreciation to the organizers of the "Dialogue Robot Competition 2023" for providing us with this opportunity, to the JTB for providing us with a place for evaluation, and to all the participants who cooperated in the evaluation.

REFERENCES

- [1] Takashi Minato, Ryuichiro Higashiyama, Kurima Sakai, Tomo Funayama, Hiromitsu Nishizaki, Takayuki Nagai, "Overview of Dialogue Robot Competition 2023" Proceedings of the Dialogue Robot Competition 2023", 12/2013
- [2] Takashi Minato, Ryuichiro Higashiyama, Kurima Sakai, Tomo Funayama, Hiromitsu Nishizaki, Takayuki Nagai, "Design of a competition specifically for spoken dialogue with a humanoid robot ", Advanced Robotics, vol.37, no.21, pp.1349-1363, 2023