

Integrating Advance Speech Recognition and Human Attribute Detection for Enhanced Receptionist Tasks in RoboCup@Home

Koshun Arimura

*Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*

Yuga Yano

*Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*

Takuya Kawabata

*Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*

Hakaru Tamukoh

*Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*

*Email: arimura.koshun523@mail.kyutech.jp, yano.yuuga158@mail.kyutech.jp, kawabata.takuya790@mail.kyutech.jp,
tamukoh@brain.kyutech.ac.jp*

Abstract

RoboCup@Home is held to integrate service robots into society. It includes a task called “Receptionist” that evaluates Human-Robot Interaction. In this task, a robot must ask guests for their names and favorite drinks and guide them to available seats. Additionally, the robot must introduce the guest’s features, such as their clothing, to others. We developed a system integrating speech recognition and human attribute detection to achieve these functions. The robot can determine which seat a person is sitting in by detecting the person’s skeletal coordinates. Additionally, the robot can identify individuals by recognizing human attributes. To verify the effectiveness of the developed system, we participated in the Receptionist task at RoboCup@Home 2024. We won first place in our league and demonstrated the effectiveness of our system.

Keywords: Human-robot interaction, Speech Recognition, Human feature detection

1. Introduction

In recent years, research in Human-Robot Interaction (HRI) has explored how humans and robots should interact in shared spaces [1], [2], [3]. RoboCup@Home [4] is a global competition that evaluates robotic capabilities across nine tasks, including HRI, to accelerate service robots’ development for society. Among these, the task that places particular importance on the HRI is a Receptionist task, as shown in Fig. 1. This task simulates a party scenario in home environment, where a robot guides guests to vacant seats and introduces them to a host.

The Receptionist task requires two primary functions of a robot. One is speech recognition. The robot asks guests for their names and favorite drinks and then receives their voice responses. The other is human position and feature detection. The robot should recognize which chairs are occupied to guide guests to vacant seats. Furthermore, the robot must identify who is seated in which chair and introduce guests to the host or another guest. To achieve that, the robot is required to associate names with individual characteristics and memorize these details.



Fig. 1 Functions required by the Receptionist task

Based on the above competition rule, the Receptionist task demands an advanced system integrating speech and human recognition. Therefore, it is meaningful for improving HRI in home environments. Furthermore, this task is carried out in a dynamic environment where new chairs may be added right before the task starts, or guests can freely move between seats. For these reasons, the environment makes the task highly practical.

A solution for the Receptionist task was proposed in [5]. This method got the highest score in RoboCup@Home 2022 but faced some issues when dealing with additional chairs and guests switching seats. In this study, we developed an integrated system that extended the conventional method to address these challenges and complete the Receptionist task. To verify the effectiveness of the proposed system, we participated

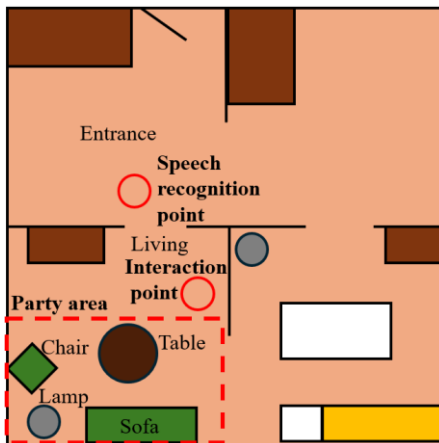


Fig. 2 Map of the arena where the Receptionist task takes place

in RoboCup@Home 2024, held in the Netherlands. In the Receptionist task, the proposed system achieved the highest score compared to other teams using the same robot and demonstrated its effect.

2. RoboCup@Home

2.1. Domestic standard platform league

The Domestic Standard Platform League (DSPL) is a division of the RoboCup@Home competition where teams compete using a standardized robot platform. In this league, participants have to focus on software development and algorithm design within a unified hardware environment. Thanks to this standardized platform, competition results are evaluated based on the quality of the software implementation rather than the robot's physical performance. In this study, we validated the effectiveness of the proposed system in the Receptionist task of the DSPL league.

2.2. Human support robot

In RoboCup@Home, we used the Human Support Robot (HSR) [6] developed by TOYOTA MOTOR CORPORATION. This robot has a camera and microphone on its head, which makes it suitable for performing HRI tasks. Furthermore, it has an arm that helps it point to specific locations to convey positional information.

2.3. Receptionist task

Fig. 2 shows the map for the Receptionist task. Guest 1 enters the entrance room at the task's start, and the robot greets Guest 1. The robot asks Guest 1's name and favorite drink. The robot then guides Guest 1 to the living room and suggests a vacant seat. Since one host is already sitting in the living room, the robot introduces Guest 1's name and favorite drink to the host.

Next, the robot returns to the entrance room to greet Guest 2. It asks Guest 2's name and favorite drink in the

same way and then guides Guest 2 to a vacant seat. The robot then introduces Guest 2's name and favorite drink to Guest 1 and the host. Finally, the robot introduces Guest 1's features to Guest 2, such as their clothing.

Throughout the task, the robot needs to look toward the person talking. Since people sitting in chairs might move to different seats during the task, the robot must identify who is sitting in which chair. In the living room, some chair's positions are known in advance, and others are arranged in the room right before the task starts. That means that the robot knows the positions of the disclosed chairs beforehand but doesn't know the positions of the newly added chairs at the task's start.

3. Implemented skills

3.1. Speech recognition

We utilize Whisper [7] as a speech recognition model to identify people's names and favorite drinks. Whisper is a highly accurate, multilingual speech recognition model. However, Whisper sometimes misrecognize words with similar pronunciations as different words. To address this issue, we predefined a list of names and drinks and calculated the similarity between the recognized text and each element in the list. We adopted the corresponding string as the recognized text when the similarity exceeded a certain threshold.

3.2. Human pose estimation

We utilize OpenPose [8] as a skeleton detection model to estimate human positions. OpenPose detects key points of the human body from 2D images captured by a camera. It also supports multi-person detection within a single frame. Furthermore, by integrating the 2D key point with depth images captured by an RGB-D camera, we calculate the 3D skeletal coordinates of individuals. This method can judge whether a person is in the entrance room or identify the locations of people in the living room.

3.3. Human attribute estimation

We utilize Class-Specific Region Attention (CSRA) [9] to estimate human attributes. CSRA enhances the recognition of appearance, such as clothing, by applying spatial attention scores tailored to each attribute. These scores highlight the specific positions and looks of objects within an image. In the Receptionist task, the robot should introduce the looks of Guest 1 to Guest 2. To achieve this function, our system used these features obtained through CSRA. In this system, the robot selects features with high likelihood scores above a set threshold and says these selected features in speech to describe human attributes. The robot can recognize and articulate 14 attributes, including gender and clothing. In this study, the system was extended to identify people even after they changed seats by associating their features with their names.

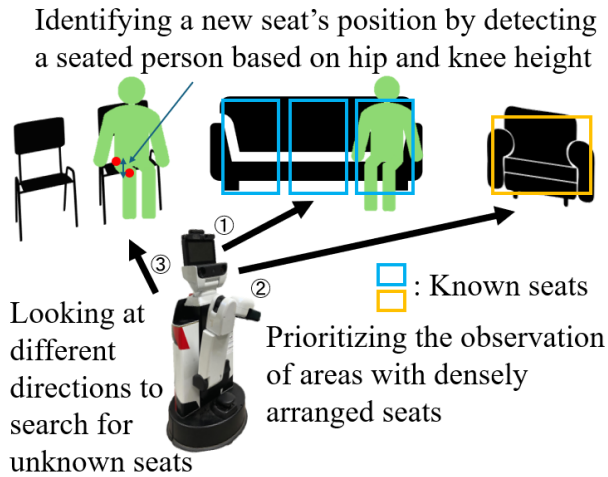


Fig. 3 Vacant seats detection system

4. Proposed system

4.1. Welcoming guests at the entrance

First, the robot detects the guest using OpenPose and requests the guest to stand in front of the robot through verbal instructions. Next, the robot asks the guest for their name, repeats the recognized name, and confirms its accuracy by asking for a "Yes" or "No" response. Similarly, the robot inquires about the guest's favorite drink and verifies the recognition. The robot also uses CSRA to detect the guest's attributes during this process. The robot relates these attributes with the recognized name. This approach is designed to avoid the mismatch between names and attributes.

4.2. Showing guest an available seat at the living

Fig. 3 shows an overview diagram of the vacant seat detection system. First, the robot moves to the interaction point in the living room. It then surveys each seat in the living room to detect people. If the distance between a chair's predefined coordinates and a detected person's central coordinates is below a certain threshold, the robot determines that the seat is occupied. Our system employs a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [10] algorithm to avoid excessive time consumption from sequentially checking each seat. Specifically, it prioritizes checking the central coordinates of densely clustered chairs, reducing the required detections and shortening processing time. Finally, the robot suggests the vacant seat nearest to the guest.

The technique of using distance between people and chairs is effective for chairs whose positions are predefined. However, new chairs are added to this competition before the task starts. Therefore, the task begins even if the robot does not know the positions of the newly added chairs. To address this issue, we implemented a method that estimates the sitting posture using skeletal data and regards the position as the

Table. 1 Result of the Receptionist task

Action	Score	1st try	2nd try
Main Goal			
Navigate the guest to the other guest	2 × 15	30	15
Look in the direction of navigation	2 × 50	100	50
Introduction a new guest to others	2 × 50	100	50
Offer a free seat to the new guest	2 × 100	200	100
Look at the person talking	2 × 25	50	50
Look at the person the robot is introducing the guest to	2 × 50	100	50
Qualitative robot social performance	50		
Bonus Rewards			
Open the entrance door for a guest	2 × 100		
Describe the first guest to the second guest (per correct attribute)	4 × 30		
Describe the first guest to the second guest (per incorrect attribute)	4 × -30		
Use standard microphone	2×5	10	5
Penalties			
Wrong guest information was memorized (continue with wrong name or drink)	-50		-50
Persistent inappropriate gaze (away from conventional partner)	-50		
Persistent gaze not in the direction of the navigation while moving	-10		
Score per try	960	590	270

coordinates of the newly added chair. Specifically, if the height difference between a person's hip and knee key points is below a certain threshold, the person is assumed to be sitting. This approach accurately identifies the position of a chair as the one a person is sitting on, even if the chair's location is unknown.

4.3. Introducing a guest to others

When the robot introduces a guest to the host or other guests, it needs to point toward the person being introduced while looking at the others. Therefore, the robot must accurately identify who is sitting in each seat. To achieve this challenge, we matched the attributes obtained by using CSRA to determine the person's name. This method enables the robot to identify the individuals seated in each position.

5. Competition result

We tested the efficacy of the proposed system through the Receptionist task in RoboCup@Home 2024. In the Receptionist task, points are awarded based on subtasks such as speech recognition, seat guidance, and guest introductions for two guests. The results of our first and second trials are shown in Table. 1. In the first trial, we completed most interactions but ran out of time and needed more time to finish the guest introduction phase. To address this issue, we replaced the Yes/No speech recognition with a touch function using the robot's hand to save time in the second trial. However, during the trial, an error occurred in the speech recognition process, and we were ultimately unable to complete the task within the time limit. Despite these challenges, our system ranked first in the DSPL league for the Receptionist task.

6. Conclusion

We extended the conventional Receptionist task system to handle dynamic environments, such as the addition of chairs and people changing seats. We tested its performance in the Receptionist task of RoboCup@Home 2024. As a result, our system achieved first place in the DSPL league and demonstrated its effect.

Acknowledgments

This research is based on results from a JPNP16007 project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This research received support from JSPS KAKENHI Grant Number 23H03468 and 23K18495, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

References

1. Y. Yano, I. Matsumoto, Y. Fukuda, T. Ono and H. Tamukoh, Proposal for Solution of Human Interaction Task in RoboCup@Home, JSAI SIG on AI Challenge, SIG-Challenge-060-02, 2022.
2. K. Yamao, D. Kanaoka, K. Isomoto, A. Mizutani, Y. Tanaka and H. Tamukoh, Development of A SayCan-based Task Planning System Capable of Handling Abstract Nouns, Proceedings of the 2024 International Conference on Artificial Life and Robotics, 2024, pp. 430-434.
3. Y. Yano, A. Mizutani, Y. Fukuda, D. Kanaoka, T. Ono and H. Tamukoh, Unified Understanding of Environment, Task, and Human for Human-Robot Interaction in Real-World Environments, The 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2024), 2024.
4. RoboCup at Home, RoboCup@Home Official Website, available at: <https://athome.robocup.org/> (accessed December 5, 2024)
5. Y. Yano, Y. Fukuda, T. Ono and H. Tamukoh, Flexible Human-Robot Interaction in Domestic Environment Using Semantic Map, Proceedings of the 2023 International Conference on Artificial Life and Robotics, 2023, pp. 409-414.
6. T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara and K. Murase, Development of Human Support Robot as the research platform of a domestic mobile manipulator, ROBOMECH Journal, Vol. 6, Art. No. 4, 2019.
7. A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey and I. Sutskever, Robust Speech Recognition via Large-Scale Weak Supervision, OpenAI, 2022. Available at: <https://github.com/openai/whisper> (accessed December 5, 2024)
8. Z. Cao, T. Simon, S.-E. Wei and Y. Sheikh, Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 7291-7299.
9. K. Zhu and J. Wu, Residual Attention: A Simple but Effective Method for Multi-Label Recognition, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 184-193.
10. M. Ester, H.-P. Kriegel, J. Sander and x. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Proceedings of the 2nd International

Conference on Knowledge Discovery and Data Mining (KDD), AAAI Press, 1996, pp. 226-231

Authors Introduction

Mr. Koshun Arimura



He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2024. He is currently a master's student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include home service robots.

Mr. Yuga Yano



He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2022. He received his M.Eng. from Kyushu Institute of Technology, Japan, in 2024. He is currently a Ph.D. student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include image processing, autonomous driving, and home service robots.

Mr. Takuya Kawabata



He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2023. He is currently a master's student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include neuromorphic physical computing.

Prof. Hakaru Tamukoh



He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.
