

Development of Reception Robot for Airport Guidance

Junya ONISHI¹Ayako MASUDA¹Chihiro OBAYASHI, Ph.D.¹Yuichi KAWASAKI¹Yuki TAKEBE¹

ABSTRACT

We proposed a novel guide robot system that uses robots to provide guidance at Narita International Airport. The system implements the guidance through collaboration between robots having two different service roles: a reception role and a guide role. The reception robot greets guests, shows the route to a destination identified by a guest through voice and touch-screen input, and, by means of the Guide Robot Management System, communicates the destination to the guide robot. As this reception robot was developed, the concept of “Compass to Accompany” in the guide robot system was implemented through the functions of universal design (multilingual support and height adjustment), impasse-free dialogue (voice recognition and dialogue engine), and natural responses from the robot (motion, sound, and display). In user evaluations, 80% of survey participants indicated that they would like to try using the system if it becomes available at an airport.

1. Background

In its 2030 Vision, Honda announced a commitment to “serve people worldwide with the ‘joy of expanding their life’s potential’.” Honda has continued the development of robots such as ASIMO as one part of that endeavor⁽¹⁾.

Among fields interested in employing robots are the delivery, sanitation, and travel businesses. In recent years, having a multilingual response capability for responding to people from diverse countries and regions has become an important issue due to labor shortages in the service and distribution industries, and especially in the travel industry, with an increasing number of international guests coming to Japan. However, it is not easy in Japan to enlist personnel capable of providing information in multiple languages. With the spread of individual information terminals such as smartphones, progress continues with environmental enhancements that provide information in multiple languages; but as these require a certain degree of media literacy, a demand exists for robots capable of communicating in multiple languages.

The objective of the airport guide robot system⁽²⁾ is to show the potential of a new guide system by offering guests at Narita International Airport (hereafter, Narita Airport) the new experience of being guided to their destinations by a robot, while the aim is also to reshape the public image of robots as not merely performing demonstrations but as

coexisting and cooperating with people, performing useful functions in society.

As shown in Fig. 1, the guide robot system performs the guide service while coordinating the two types of robot, reception and guide robots. The reception robot in Fig. 1(b) shakes its body to attract the attention of a guest, and shifts its body toward the guest when approached. When the guest selects a language using a touch-screen terminal, the reception robot faces to the front where the touch screen is located.

As shown in scenario 1 of Fig. 1(a), when the guest mentions a destination the reception robot performs a nodding motion and displays a selection list on the touch-screen display. When the guest touches one of the destinations in the list, detailed information showing how to proceed to the destination is displayed. If, as shown in scenarios 2, 3, and 4 of Fig. 1(a), the guest selects the option to go with a robot, then through the management system⁽³⁾ the guide robot, shown in Fig. 1(c), receives the destination and face-identifying information⁽⁴⁾ for the guest from the reception robot, and accompanies the guest, leading the way to the destination. Although responsibilities are divided between the reception robot and the guide robot, the guide service is provided to the guest quite seamlessly.

This paper discusses the reception robot in the Narita Airport robot system.

¹ Life Creation Center



Fig. 1 Guide service overview

2. Reception Robot Concept and System Configuration

The guide robot service incorporates the concept of a Compass to Accompany. “Accompany” expresses taking care to remain at a comfortable distance and proving to be a reliable presence while “Compass” indicates a tool of unmatched accuracy showing the direction and route to a destination.

During development of the reception robot the foundational concept of a Compass to Accompany was further broken down into concepts (1) through (3) below. Five functions developed to implement these three concepts are multilingual support, height adjustment, voice recognition, a dialogue engine, and motion-sound-display. Figure 2 shows how these concepts and functions correspond to one another.

- (1) Universal design (multilingual support, height adjustment)
- (2) Impasse-free dialogue (voice recognition, dialogue engine)
- (3) Natural robot response (motion, sound, display)

Universal design is configured so as to be responsive not only to foreign guests but also, since it is used in a public

space, to users of wheelchairs, and is implemented with the features of multilingual support as well as height adjustment to accommodate an individual sitting in a wheelchair. Impasse-free dialogue is implemented with voice recognition and a dialogue engine to enable information to be obtained regardless of media literacy skill levels. To provide the guest with the experience of a robot engaging in communication, responses to actions by the guest are implemented through motion, sound, and display, which enables natural responses by the robot.

Figure 3 shows the system configuration that implements these functions in the reception robot. The system is configured with a camera and a microphone, an input system based on a Laser Range Finder (LRF), a recognition system for recognizing inputted images, voice, and point group data, a motion generation system, and a search system.

When the LRF detects that a guest has approached or when there is vocal input to the microphone, motion and sound are generated in the robot, achieving the concept of natural robot response. Vocal input from the microphone is applied to a destination search in the multilingual support function and potential destinations are displayed. During

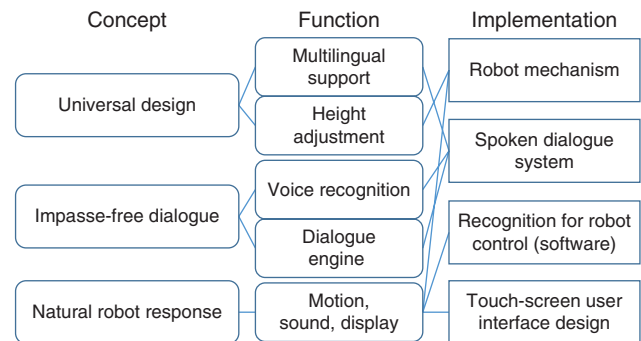


Fig. 2 Table of correspondences among concepts, functions, and implementations

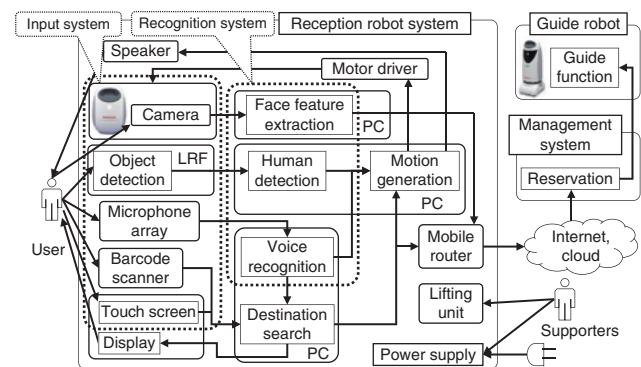


Fig. 3 System diagram

response to a person in a wheelchair the microphone is set by staff to a height convenient for use.

The reception robot communicates with the management system and the guide robot for coordination purposes, communicating the guest's destination to the guide robot to deliver seamless guidance.

The five functions mentioned above are achieved through a combination of the system configuration shown in Fig. 3 and the four features of a robot mechanism, a spoken dialogue system, recognition for robot control, and a touch-screen user interface. These four features are shown in the third column in Fig. 2 together with their relationships to the five functions, and are explained in Chapters 3 through 6. Chapter 7 describes the results of an impressions evaluation conducted to understand what aspects of these five functions called for enhancement.

3. Robot Mechanism

3.1. Robot Motion Mechanism

When a guest approaches the reception robot and operates the touch screen to begin a search for a destination, the reception robot executes motion suited to that scenario. Figure 4 illustrates an overview of actual motions.

A mechanism to accomplish motions of this sort is shown in Fig. 5. To make the robot bow or tilt there is a structure capable of rotating around each of three axes, X, Y, and Z, which are centered at the bottom of the robot and are orthogonal to one another. In addition, by configuring the drive mechanism to use links it is possible to minimize the curvature of a wire harness, which makes it possible for the robot to be used at the airport over a long period of time. In the section where the three axes intersect a universal joint is employed, which connects through the links to a bearing that slides vertically, and communicates forward-and-back and side-to-side motions while enabling rotation. These

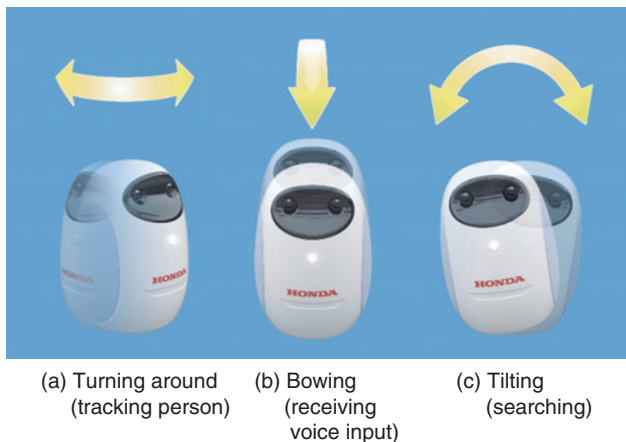


Fig. 4 Robot motion examples

portions of the mechanism were developed in a joint effort with ARRK Corporation.

By the use of this type of motion mechanism the concept of natural robot response was implemented.

3.2. Height Adjustment

To provide universal design, a height adjustment feature was provided to make it easy for people of various heights and for wheelchair users to see the touch screen and to enhance sound collection capabilities. This height adjustment feature is shown in Fig. 6.

In order to accommodate heights ranging from wheelchair users to tall foreign visitors, the touch screen was specified to be adjustable in height between 927 and 1310 mm. A commercially available height-adjusting unit intended for desk use was employed.

Use of this height-adjustment mechanism enabled realization of the universal design concept.

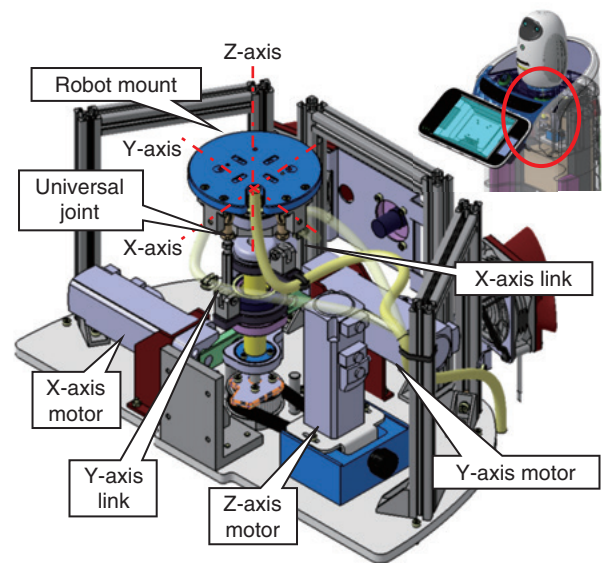


Fig. 5 Robot drive mechanism

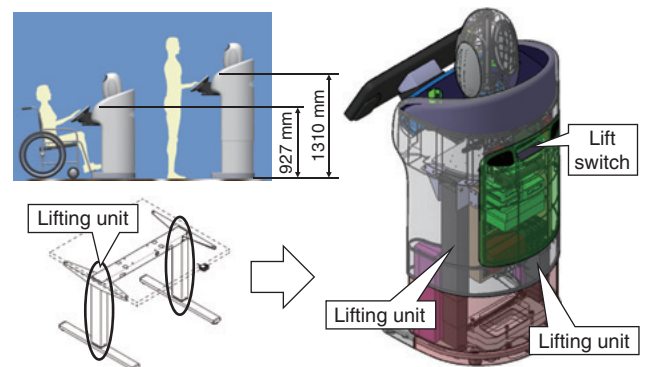


Fig. 6 Lifting mechanism

4. Spoken Dialogue System (Microphone, Dialogue Engine)

Figure 7 shows the sequence of the voice dialogue search system. When the guest speaks, the microphone converts the sound into a voice signal, and only the signal for the human voiced segment is extracted by Voice Activity Detection (VAD). After being processed, the voice is transmitted to the dialogue engine Houndify, and once the destination has been recognized and understood, search results are returned and displayed on the touch screen. When the guest makes a selection from the resulting list, information is obtained from the Narita Airport official app (NariNAVI), and relevant details (including a map) can be displayed.

4.1. Microphone

The microphone used for voice recognition needs to perform recognition accurately in the noisy environment of the airport. A design requirement of the voice recognition and dialogue engine is that the Character Error Rate (CER) for voice recognition results in a noise environment mimicking the inside of an airport be set to 10% or less, and the task incompleteness rate, which is the voice recognition result when sentence meaning has been taken into account, be set to 15% or less. These settings are based on the experience of Fairy Devices Inc., a company that develops voice recognition and voice dialogue technology, with

respect to voice response services in public facilities.

For the microphone array, a bar-shaped 16-ch. multi-microphone system produced by Fairy Devices was used. Using a microphone array makes it possible to approximate the direction of a sound source even in the midst of interfering noise, and to select only sound that comes from sources in a specific direction. Parameters are adjusted to isolate sound coming from the presumed position of the guest from noise in the environment and extract only the guest's speaking voice, achieving a CER of about 2 to 4% and a task incompleteness rate of about 11%, which is considered wholly sufficient to perform voice recognition smoothly. Voices in the vicinity of the microphones were assumed to have an input sound pressure between 65 and 75 dB.

4.2. Dialogue Engine

In order to respond to diverse guests, the dialogue engine used a modified version of the voice AI platform Houndify, which was provided by SoundHound Inc. on request. This dialogue engine performs at a level of anticipated dialogue so as to be able to recognize expressions and vocabulary close to everyday speech in the four languages of Japanese, English, Chinese, and Korean. It performs at a level that users indicated they would like to be able to converse at with a robot, and is able to recognize about 70% both within that standard and outside of that standard. One setting requirement for voice recognition and the dialogue engine, established after the results of user evaluations described in Chapter 7 were obtained, was that the response time between voice input and the display of results be within three seconds. Conventionally, data has been taken from a voice file, temporarily converted into text, and then applied to a dialogue engine, but Houndify, the dialogue engine that was used, generates a response directly from the voice file, so response time was shortened and results were displayed within three seconds of voice input.

5. Recognition for Robot Control (Software)

In order to achieve a function whereby the robot turns its face to a guest for the purpose of drawing the guest near, simple configurations were implemented for the detection of people in the environment, user selections, and robot posture.

LRF has been proposed⁽⁵⁾⁻⁽⁷⁾ as a method of detecting people in the environment using leg or torso detection. One can imagine that in an environment where a reception robot would be installed there would be multiple people moving or standing about in the vicinity of the robot; and it was thus decided to detect torsos using LRF, which was set within a height range that allowed individuals' torsos to be line-scanned. The robot had a height-adjustment function but the head of the robot would remain at a standard height until a guest stood in a position to operate the touch screen. It was confirmed that at the standard height LRF detection

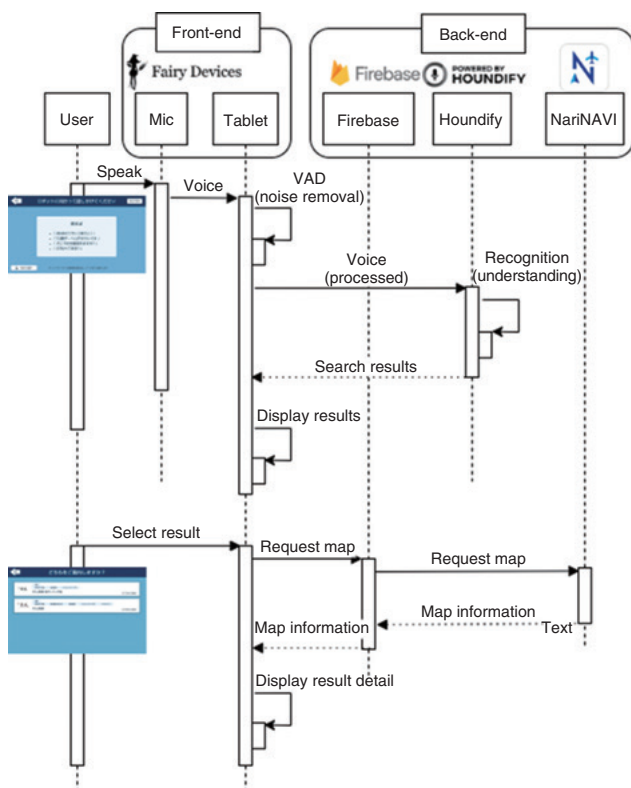


Fig. 7 Sequence diagram for dialogue search engine

was possible from the level of a user in a wheelchair to the height of a tall individual. An overview of the sensing area and installation position is shown in Fig. 8 from a perspective above the head of the robot.

Figure 9 is a schematic drawing relating to the detection of people. Among those point groups sensed by the LRF, a set of points was regarded as a cluster if the points differed from those nearby within a threshold value. It was assumed that points associated with a human torso would form a roughly elliptical shape, which was judged to be human based on the width and orientation of the cluster. Specifically, the distance between edge points in a cluster $\|v_w\|$ shown in Eq. (1) is within the width of a human chest.

$$l_{\min} < \|v_w\| < l_{\max} \quad (1)$$

Here, $l_{\max}$ and $l_{\min}$ are the highest and lowest values, respectively, for chest width. Both were empirically adjusted based on human chest width measurements.

One can expect that someone interested in the reception robot will orient their body toward the robot. With respect to the direction a person is facing, a cluster believed to be

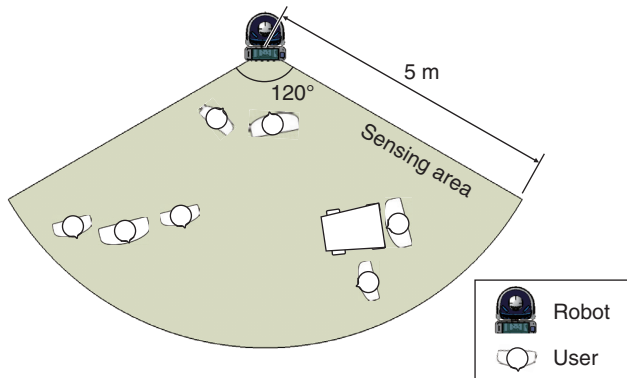


Fig. 8 LRF sensing area

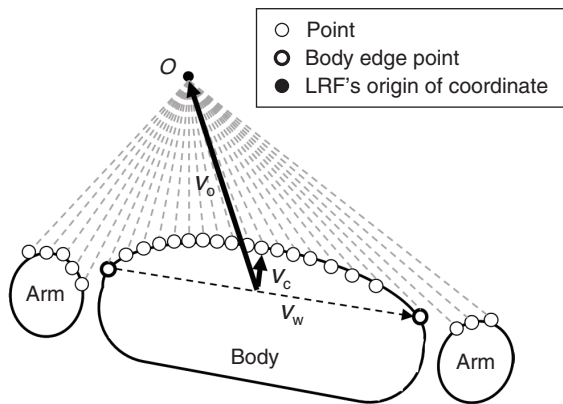


Fig. 9 Definition of vectors for detecting people from point cloud data

associated with a human is treated as a person facing the direction of the robot when the degree of cosign similarity between two vectors, shown in Fig. 9 as a vector v_c in the direction of the center of the cluster and a vector v_o in the direction of the LRF point of origin O , both vectors emanating from the midpoint between the two edge points of the cluster judged to be a person, is as expressed in Eq. (2).

$$\cos(v_o, v_c) > 0 \quad (2)$$

Among individuals facing the robot, the person who is the shortest distance v_o away from the robot is the person selected as the target user. Attitude control for the robot is performed by an attitude angle control system that takes as input the attitude angle that the robot should adopt as determined by the positional relationship between the target user and the robot. When the target user operates the touch screen, the operations described in Section 3.1. are performed.

6. Touch-screen User Interface Design

6.1. Design Considerations

In order to achieve impasse-free dialogue with all guests, the front of the robot is equipped with a touch screen. The guest uses the touch screen to select a language for interacting with the guide service and can choose to proceed with searching for a destination through dialogue with the robot or by using a list displayed on the touch screen. This ability to respond to the needs of various guests is one feature of the system.

6.2. User Interface Flow

The configuration of the touch screen and the user interface flow are based on results from a survey of user evaluations, described in Chapter 7. When a guest faces the touch screen, selects a language from the tutorial screen shown in Fig. 10, scenario 1, and acknowledges the consent form that is presented concerning personal data protection and cautions relating to the robot, a voice chat screen is displayed, as shown in Fig. 10, scenario 2. Because one objective of the reception service is to make conversing with a robot a fun experience, performing a search by way of vocal dialogue is given precedence over searching by means of a list. Thereafter, a list of search results is displayed as in Fig. 10, scenario 3. The search term is displayed at the top of the list so that the guest can check that voice recognition was performed correctly. When the guest selects a destination from the list and taps on the display, the display transitions to a results screen that provides detailed information about the destination, as in Fig. 10, scenario 4. The search proceeds from beginning to end with as few as three taps, a great step forward in useability. The results screen uses NariNAVI to display a map of the route from the guest's current position to the destination. Landmarks appear on the map, making it easier for the guest to follow the route. Finally, the guest

chooses to be accompanied to the destination by a robot or to proceed without escort, and a reception service closing screen is displayed.

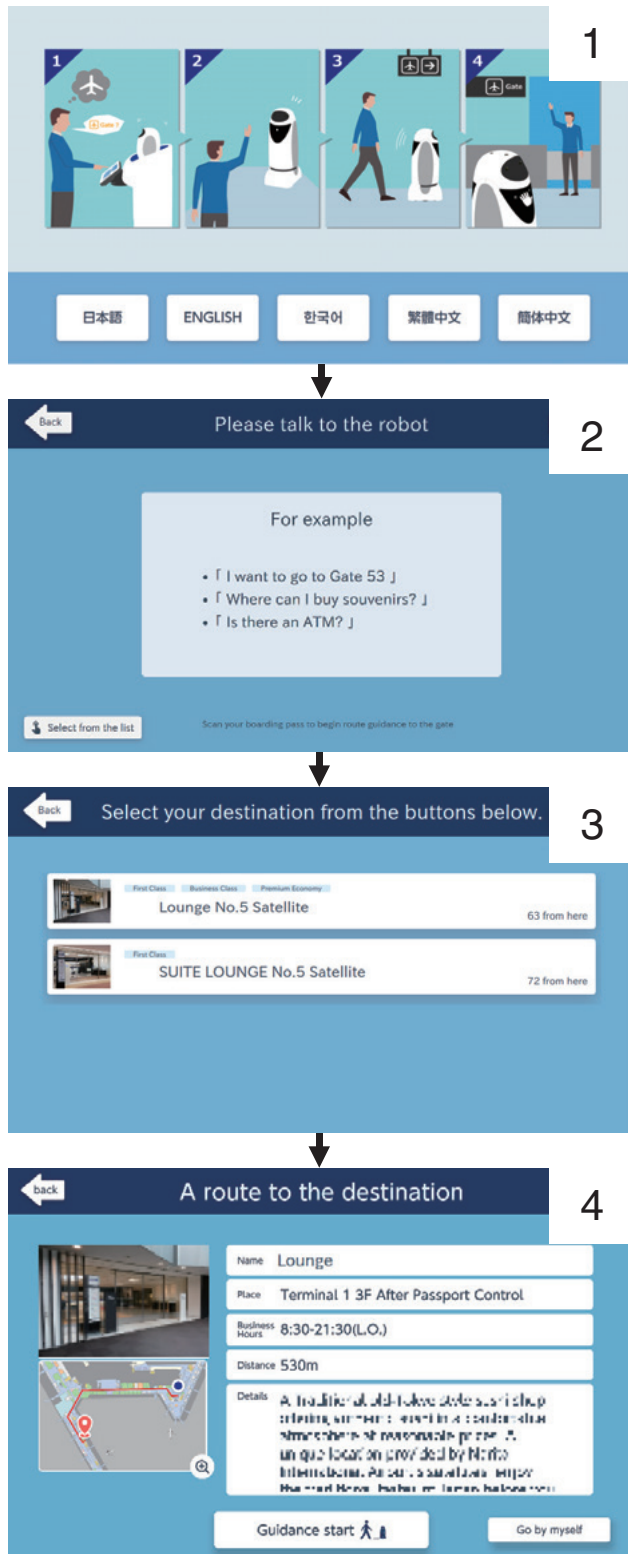


Fig. 10 User interface flow

7. User Evaluation Survey

7.1. Objective

In order to understand what enhancements the functions described in Chapter 2 needed, this survey asked about impressions received, with a focus on the following points:

1. Touch screen usability
2. Impressions of the robots
3. Willingness to use the reception service in the future

7.2. Survey Method

Survey participants included five Japanese (two men, three women) and five foreign nationals (three men, two women). The Japanese were Japanese citizens whose native language was Japanese, and the non-Japanese were native English speakers with foreign citizenship. Efforts were made to achieve as much diversity as possible with respect to age and frequency of experience interacting with a voice recognition terminal or robots.

The survey was conducted between January and February of 2020 in the Wako Building at Honda R&D Co., Ltd. An environment very closely resembling the one in which the guide service would be offered was created in a conference room inside the building. The survey involved a usability test and interview. Survey participants attempted to execute six tasks using the reception service. One such example was to request to be accompanied by a robot after communicating a desire to go to a lounge. At the beginning of the survey, administrators explained the nature of the survey to the participants and obtained their informed consent.

Prior research⁽⁹⁾ was consulted for the evaluation of touch screen usability, and a questionnaire suited to evaluation of the reception service in this instance was used. After the participants had completed all of the tasks, they were asked to evaluate their experience using a rating scale of 1 (disagree) to 5 (agree) on the questionnaire. The reasons for their ratings were noted in an interview following the evaluation.

In evaluations of the impression made by the robot, GODSPEED⁽¹⁰⁾, a standardized evaluation index in the field of Human-Robot Interaction (HRI), was used. This is a questionnaire encompassing five concepts (anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety) important in the field of HRI. For example, the words “fake – natural” are paired together, and participants rate anthropomorphism on a scale of 1 to 5, choosing 1 if the robot seemed fake and 5 if the robot seemed natural. In addition, the four additional items of “apathetic – sympathetic,” “fear – secure,” “unpleasant – pleasant,” and “inaccurate – accurate” were used. These were chosen as items to evaluate the Compass to Accompany, the guide service concept in this instance. Survey participants filled out the evaluation questionnaires after completing all tasks. Their reasons for giving the ratings they did were noted in the interviews that followed.

An evaluation index called Affect Grid⁽⁸⁾ was used to evaluate users’ feelings about the reception service and their

willingness to use it. The Affect Grid represents the feelings experienced at that moment on two orthogonal axes, namely, pleasure – displeasure, and arousal - sleepiness. If a user said, “It was somewhat pleasant and fairly stimulating,” the extent of those impressions was evaluated on each of the two axes using 10 levels, and a mark was inscribed on points corresponding to a plane consisting of the two axes. Survey participants evaluated their feelings before performing the tasks and after completion of each task. This served the purpose of investigating time-series variation in the user’s feelings before and after experiencing the reception service as well as during. In the post-evaluation interview, the user’s

impression of the entire reception service and willingness to use it if encountered at an airport were noted.

This survey was conducted with approval from the Bioethics Committee of research and development at Honda R&D Co., Ltd.

7.3. Survey Results and Conclusions

Survey results for all participants relating to touch screen usability are shown in Fig. 11. The blue line shows average ratings by Japanese and the red line shows average ratings by non-Japanese. The average response rating was 3.57 for Japanese as a group, and 4.06 for non-Japanese.

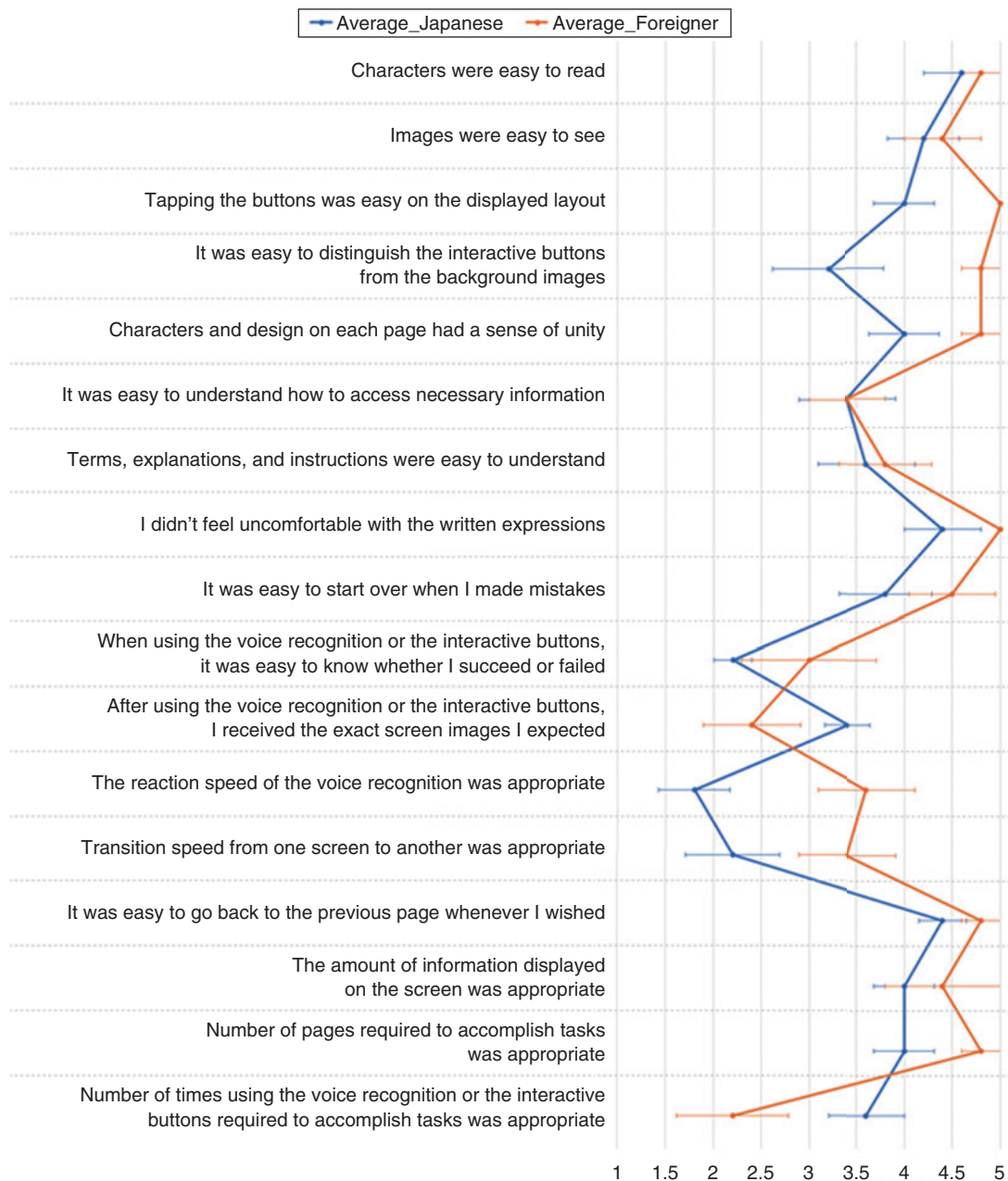


Fig. 11 Evaluation results for touch screen usability

The survey results, including the interviews, suggested five areas needing enhancement. These are: response at the time of vocal input (delay in acknowledgement, uncertainty as to whether input was understood, handling of recognition flaws); the microphone button (operation, understanding where to direct speech); list display (no coherence in items, inconvenience of behavior); map screen (smallness of map, understanding expansion button); and keyboard input (text entry). Based on these results, an allowance of up to three seconds for a response time was set for the voice recognition system described in Chapter 4. In addition, the touch-screen configuration and user interface flow laid out in Chapter 6 were configured on the basis of these results.

Survey results for all of the participants' impressions of the robots are shown in Fig. 12, using the GODSPEED index. The blue line is the average rating by Japanese and the red line is the average rating by non-Japanese. Non-Japanese tended to have a more positive response on average with respect to anthropomorphism than Japanese. Non-Japanese were apt to have a more positive response on average for animacy than Japanese. Likability tended to be rated highly by both Japanese and non-Japanese. Both Japanese and non-Japanese tended to give high ratings for sense of intelligence. Japanese and non-Japanese tended to assign similar ratings, on average, for sense of safety. On the four items that were added, both Japanese and non-Japanese gave high ratings, but on average, ratings by non-Japanese were higher.

The interviews suggested the following findings. Feedback relating to anthropomorphism and animacy ratings concerned the robot's appearance, movements, and sound, which generally had a positive impact. Ratings for likability, sense of intelligence, and the additional items "sympathetic" and "accurate" indicated that negative

impressions of the touch screen carried over into negative impressions of the robot. It became clear that negative impressions of the touch screen, which is a separate system, caused a negative assessment of the robot, and thus changes were made to address areas that required enhancement according to touch screen usability evaluations. Also, rating results for the four additional items selected for evaluation of the Compass to Accompany concept suggested that survey participants experienced the robot much in the manner that had been intended.

When the Affect Grid index was used to analyze all participants' reactions to the reception service, it was evident that changes in the impression made during the course of the reception service were strongly influenced by how smoothly tasks were executed. Participants had a sense of curiosity and anticipation prior to executing tasks, but were left feeling frustrated if they could not easily complete the tasks by means of the touch screen. Regarding willingness to use the reception service, 8 out of 10 individuals replied that they would be willing to use it if they actually encountered it at an airport, assuming that the issues with the touch screen had been addressed. When asked the reasons for this in interviews, Japanese and non-Japanese indicated differing tendencies. Japanese expressed negative opinions about their willingness to use the service, citing concerns about their companions or others present in the area. On the other hand, non-Japanese expressed largely positive responses to the idea of using the service. Thus, it was clear that willingness to use the service was strongly related to individual characteristics, and in particular that it might be influenced by experience in conversing with a robot or concerns about feeling self-conscious. Enhancements suggested by the survey results were incorporated into the design.

8. Conclusion

A guide robot system was developed to perform a guide service at Narita Airport, and as part of that service a reception robot was introduced. During development of the reception robot the concept of a Compass to Accompany in a guide robot service was broken down into the following elements: universal design, impasse-free dialogue, and natural responses, as well as functions corresponding to each of these requirements, i.e., multilingual support/height adjustment, voice recognition/dialogue engine, and movement-sound-display, were provided. User evaluations conveyed that survey participants experienced the robot much as envisioned in the Compass to Accompany concept, indicating success in the development of a robot aligned with that concept. The interviews revealed that negative impressions of the touch screen negatively impacted impressions of the robot, and that negative opinions of the touch screen, which was a separate system, negatively impacted impressions of the robot itself. Still, the fact that 80% of survey participants responded that they would like

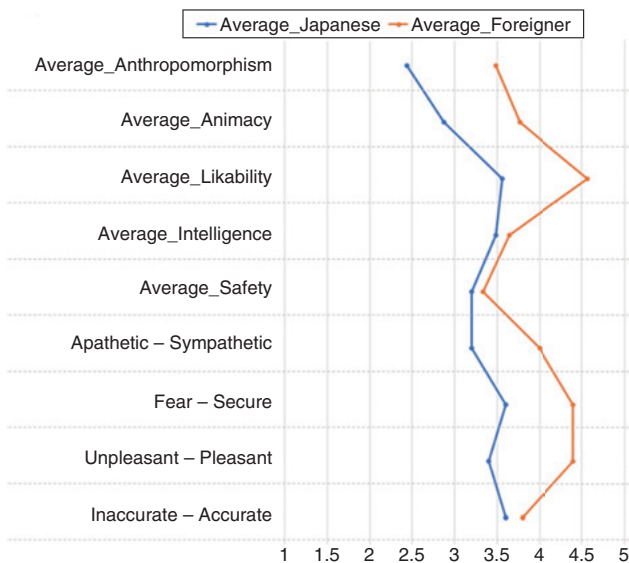


Fig. 12 Evaluation results for impressions of robot

to try the system if it were actually installed in an airport, with the proviso that the touch screen issues be resolved, indicates that this would be a useful service.

References

- (1) Shigemi, S., Kawaguchi, Y., Yoshiike, T., Kawabe, K., Ogawa, N.: Development of New ASIMO, Honda R&D Technical Review, Vol. 18, No. 1, p. 38-44
- (2) Saito, Y., Koshiishi, T.: Development and Styling Design of Airport Guide Robot System, Honda R&D Technical Review, Vol. 33, No. 1, p. 1-9
- (3) Nishimiya, K., Kaneko, S.: Guide Robot Management System, Honda R&D Technical Review, Vol. 32, No. 2, p. 51-55
- (4) Sha, Z., Natori, Y., Ariizumi, T.: Research of Robust Follower Recognition Function for Airport Guide Robot, Honda R&D Technical Review, Vol. 33, No. 1, p. 19-27
- (5) Chung, W., Kim, H., Yoo, Y., Moon, C.-B., Park, J.: The Detection and Following of Human Legs Through Inductive Approaches for a Mobile Robot With a Single Laser Range Finder, IEEE Transactions on Industrial Electronics, Vol. 59, No. 8, p. 3156-3166, (2012)
- (6) Jung, E.-J., Lee, J. H., Yi, B.-J., Park, J., Yuta, S., Noh, S.-T.: Development of a Laser-Range-Finder-Based Human Tracking and Control Algorithm for a Marathoner Service Robot, IEEE/ASME Tran. Mechatronics, Vol. 19, No. 6, p. 1963-1976, (2014)
- (7) Nakamori, Y., Hiroi, Y., Ito, A.: Multiple player detection and tracking method using a laser range finder for a robot that plays with human, Robomech Journal 5, 25, (2018)
- (8) Russell, J. A., Weiss, A., Mendelsohn, G. A.: Affect Grid: A Single-Item Scale of Pleasure and Arousal, Journal of Personality and Social Psychology, Vol. 57, No. 3, p. 493-502, (1989)
- (9) Nishimoto, K., Nambu, M., Ito, K.: Evaluation An Experimental for Effectiveness and Implementation Ease of Remote Usability Testing, Proceedings of the 31st JSSST Conference, Ippan 6-4, p. 480-485, (2014) (in Japanese)
- (10) Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots, International Journal of Social Robotics, p. 71-81, (2009)

■ Author ■



Junya ONISHI



Ayako MASUDA



Chihiro OBAYASHI



Yuichi KAWASAKI



Yuki TAKEBE