Kyabo: Exploring Natural Indoor Navigation with Robots

Aadish Gupta
Department of Computer Science
University of Colorado Boulder
Boulder, Colorado 80309
Email:aadish.gupta@colorado.edu

Bryan Bo Cao
Department of Computer Science
University of Colorado Boulder
Boulder, Colorado 80309
Email:bo.cao-1@colorado.edu

Kyle Wiese
Department of Computer Science
University of Colorado Boulder
Boulder, Colorado 80309
Email:kyle.wiese@colorado.edu

***Abstract*—Guides of many different forms, ranging from a simple map to people who will personally guide others to their destination, are present nearly everywhere in daily life; however, these forms of guides have been updated little since their conception. As a result and due to the increase in capability of robotic technology, the promise of embodied agents acting as a new form of guide is growing. To become an effective guide, these agents must contain the ability to communicate with a user, have an intuitive interface, and relay information in the most efficient way possible. In this paper we present a design for an embodied indoor navigation agent that can relay directions to the user through the use of verbal, visual, and mechanical cues. An experimental evaluation of our approach showed that users who received mechanical (following) and visual cues had a significant effect on reducing the difficulty involved in finding a destination in an unknown area. Our findings offer guidelines for developing embodied navigation agents for indoor environments.**

I. Introduction

Due to embodied agents' abilities to reproduce multiple social cues as well as providing help through physical means, these robots hold great promise for integrating seamlessly into everyday service tasks. To know when best to utilize these abilities, service robots must be able to communicate with users in an efficient manner and be able to complete the service tasks in possibly dynamic environments. For instance, one of the most important aspects of navigation and guidance robots is to have intuitive interfaces that allow users of any background to be able to enter their destinations. Once these destinations are provided, the robot then needs to effectively communicate to the user what route will be taken and how the robot will be assisting them in reaching their destinations. In the case of a guidance robot that has the physical ability to move around a designated area, the robot must be able to know the best way to reach a desired destination while also being able to move around unpredictable obstacles. When guiding a user around a crowded area, the robot must be able to detect other people crossing its path and make the necessary adjustments to avoid running into them. These behaviors involve a number of verbal and visual communication methods as well as effective sensing and processing capabilities. For instance, verbally describing the user's directions step by step and displaying the route via a virtual map require a screen and speakers while being able to localize itself in the building requires a camera and a decent processor.

Fig. 1. Guiding Robot – TurtleBot

Popular methods of assisting people in navigating unknown indoor spaces involves utilizing indoor maps, intuitive numbering systems, directories, and signs placed in strategic locations throughout a building, or if all else fails, asking another person if they know how to reach a destination. Due to their simplicity and readability, these methods have been proven to be effective in relating directions to people of nearly any age and of any background; however, these methods also require the person to be able to recall the exact turns, numbers, and relative location several minutes after receiving the directions for the first time. As a result, there is much room for improvement as these methods don't take short term memory into account and can often cause people to backtrack and re-read the same directions multiple times.

Although other methods, such as Google Maps' indoor directions, might help with the retention and access needed for finding an unknown location in a building, they don't work in every building. More importantly, these applications don't provide any social interaction. Social interaction has been shown to increase the confidence and rapport a user feels when engaging with a piece of technology and as a result, could be an important aspect to consider when designing new guide robots[1].

As these navigation and guidance robots continue to integrate themselves into multiple facets of everyday life, such as assisting users in finding the correct isle in a store, directing users to the correct office in a complex building layout, or leading users to the person they need to find, we sought to provide an answer of how best these robots should operate. To this end, we developed a system of tasks and conditions that allowed us to explore the differences in participants' abilities in finding unknown destinations. We investigate how an embodied agent, through providing different types of directions, affected the overall user completion success (time to completion and perceived difficulty of the task) for each scenario.

II. Related Work

In last few years Human Robot Interaction has emerged as a field of its own. Humans and robots in the near future will co-exist in the same eco-system and will complement each other in day to day tasks. These embodied robots provide the social and physical abilities necessary to make this eco-system work. Examples include autonomous robots acting as way to guide people throughout museums. Navigation and guidance tasks not only benefit from physical presence, but social aspects as well. One such example of a social robot is the robot receptionist HANK, which researchers at Carnegie Mellon University have designed to help visitors through providing them with audio directions[2]. Thibault Kruse argued that for navigation, the presence of humans requires novel approaches that take into account the constraints of human comfort as well as social rules[3]. Saeed did research on human teaching objects about places through hand gestures and linguistic description[4]. V. Alvarez-Santos provided an interaction that using voice messages as the feedback for humans[5].

Nikoaloas Mavridis proposed ten desiderata in a review, providing an insight into the efficient communication between human and robot by verbal and nonverbal commands[6]. Wolfram Burgard argued that indoor tour-guide robots should be able to navigate indoor environments dynamically and take various software modules into account, including a camera and display to show the map[7]. Dan Bohus, Chit W Saw and Eric Horvitz proposed an open-world human-robot interaction in dynamic, multi party environments, where people come and go, and interleave their interactions with the system and each other[8].

However, all the efforts for providing assistance with indoor navigation through robots have been unidirectional, either they guide humans to destination without regard to social cues or provide them with non-physical directions that take social aspects into account. Few efforts has been made to seek an intuitive navigation interaction between humans and robots regarding path complexity with respects to a multi-model system with a visual map, audio, physical, and social cues. We want to develop a social indoor navigation system which, based on complexity of direction, will either give verbal and visual directions or will guide the user to the destination.

We investigated under what conditions people would be most comfortable to receive direction and when they would like to follow the robot. The results from this study will help us build the computational model for the robot to decide which mode of direction (voice directions vs guiding the human) based on destination requested by user. In addition, the dataset recorded during this study, including video from participants and primitives and movements, will be used for robot navigation training.

III. Hypotheses

Three hypotheses were developed for our experiment based on findings in previous HRI and navigation studies. We measured completion success to be the combination of time for participants to complete the task and their perceived difficulty in reaching the destination.

A. Hypothesis 1

Participants who follow the robot with a map will have the highest completion success whereas participants who only receive directions via audio with a map will have the lowest completion success.

B. Hypothesis 2

When presented with complex directions, participants who follow the robot with a map will have a more beneficial experience compared to audio directions with a map.

C. Hypothesis 3

When presented with simple directions, participants who only receive audio directions with a map will have a more beneficial experience compared to following a robot with a map.

IV. Evaluation

To investigate the effect of "mode of directions" and "complexity of the path" we conducted a Field Experiment in which participants interacted and received directions from guiding robot. Below we describe the design of our experiment, procedure, measurements, and population.

A. Experimental Design

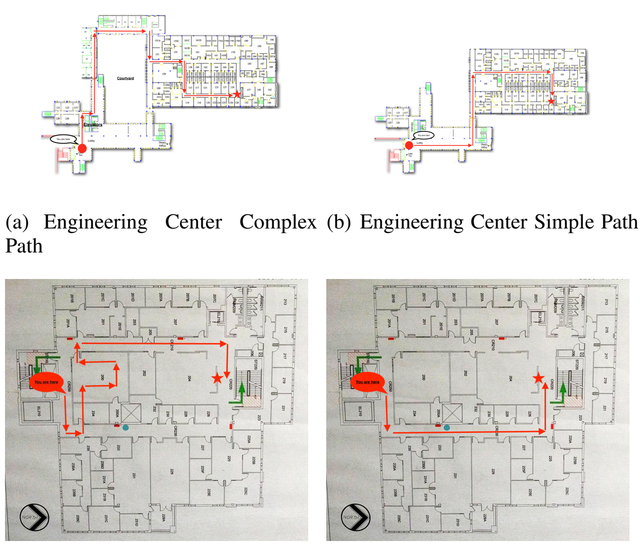
To test our hypothesis, we conducted a 2 X 2 between-participants study in which we manipulated "mode of directions" and "complexity of the path". We had two independent variables. The first independent variable was "mode of directions" which had two levels: (1) Audio directions with a map and (2) Follow the robot with a map. The second independent variable was "complexity of path" which had two levels: (1) Simple path with less than 3 turns and (2) Complex path with more than 7 turns. The dependent variables included completion time, difficulty of the task, quality of information disseminated and interaction with the robot.

TABLE I
STUDY DESIGN

Study Design	Follow Robot with Map	Audio Directions with Map
Simple Path	X	X
Complex Path	X	X

B. Experimental Procedure

In the study, each participant was presented with a task to reach a specified destination inside the building with the help of the robot. Participants were chosen such that they were not aware of the destination inside the building. Participants were asked to meet in the Engineering Center or ATLAS Building. Before starting the experiment, researcher asked them to fill a pre-experiment survey which was intended to gauge their prior interaction with robots, innate ability to navigate, and their current usage of navigation services.



(a) Engineering Center Complex Path

(b) Engineering Center Simple Path

(c) ATLAS Complex Path

(d) ATLAS Simple Path

Fig. 2. Paths Used in Different Conditions

A researcher briefed the participant about the task, which was to reach the destination inside the building. Finer details of the task were left to be explained by the robot. This helped in maintaining consistency across the participants group. Before participants started interacting with the robot, their verbal consent was taken to record the experiment.

The Wizard of Oz approach was used to simulate the experiment. Behind the scenes, a researcher was controlling the robot as well as playing the pre-recorded messages. This gave the participants the illusion that robot was autonomous. The robot configuration we chose to use was a TurtleBot speakers attached to it. For all the participants the building map with the path marked was presented on the laptop screen.

In the Engineering Center participants were provided with Audio Directions and the complexity of the path was chosen randomly. Each participant in the engineering center was greeted by the robot using a pre-recorded message followed by explanation of the task and audio directions. A map of the building was always visible on the screen. Participants could come back anytime and receive the directions again if they were lost. The final destination inside the Engineering Center was the room number which was conveyed by the robot through audio directions.

Fig. 3. Participant Interacting With the Robot

In the ATLAS Building, participants were asked to follow the Robot and the complexity of the path was chosen randomly. Similar to engineering center, each participant was greeted by the robot, followed by explanation of task and prompt to follow them to reach the destination. The final destination inside the ATLAS Building was also a room number but instead of robot conveying it, the researcher provided the participant with the printout of the room number in braille, which was used to avoid the situation when participants could reach the destination by merely knowing the room number.

At the end of the task, each participant was met by the researcher, who then asked the participants to fill out post-experiment survey. This survey was aimed to assess the participants' experiences and to gather feedback.

C. Participants

A total of 19 participants (13 Males, 6 Females) took part in this experiment. Each condition had 5 participants except for the Complex Path and Follow Robot condition which had 4 participants. Each participant could understand and communicate in English. Participants had modest prior experience with Robots (M=3.05, SD=1.96), while their innate navigation ability was bit higher (M=4.32, SD=1.34) on the scale of 1 to 7. Figure 3 shows participant interacting with the robot.

D. Measurement

Objective measurements included the time to complete the task. Additionally, subjective measures were taken by means of a seven point rating scale used to measure participant's responses on a post-experiment questionnaire. Questions were designed to get participants perception of task difficulty, information received from the robot, and interaction with the robot.

To verify our manipulations we built 2 scales from the subjective measures. We built a four item scale to measure quality of information disseminated by the robot under different conditions. The items for this scale asked participants about useful the information was, how effective was the robot's response to participants request, how natural was the robot's behavior, and the robot's understanding of the request (Cronbach's $\alpha = .86$). Our second scale was two item scale to measure the interaction between robot and participant. The items for this scale asked how effective the interactions were and how comfortable the participants felt (Cronbach's $\alpha = .8$).

Besides these scales to verify our manipulations we used completion time, which is a objective measure, and task difficulty, which is a subjective measure.

V. Results

After performing Pearsons product-moment correlation coefficient on our two provisional scales (information dissemination and interaction experience) and two independent factors (time to complete task and task difficulty), we found evidence of a significant correlation between interaction experience and information dissemination (p=.0078, r=.65) as well as a correlation between difficulty and time to complete the task (p=.0028, r=.59) (Figure 4). The results of further analyses are described in the subsections below.

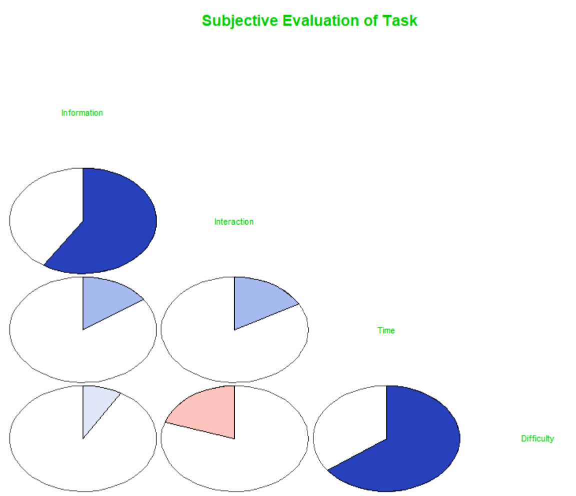


Fig. 4. Correlation Between Scales

A. Information Dissemination

We conducted a two-way analysis of variance (ANOVA) to test whether "mode of directions" and "path complexity" affected information disseminated by the robot and found no significant main effect. But there was a significant interaction effect of the conditions on information dissemination, $F(1, 15) = 6.56, p = 0.022$ in Figure 5 for interaction plot.

B. Success

We conducted a two-way analysis of variance (ANOVA) to test whether "mode of directions" and "path complexity" affected task duration and found two significant main effects but there was no significant interaction effect. "Path complexity" had a significant main effect on the task duration,

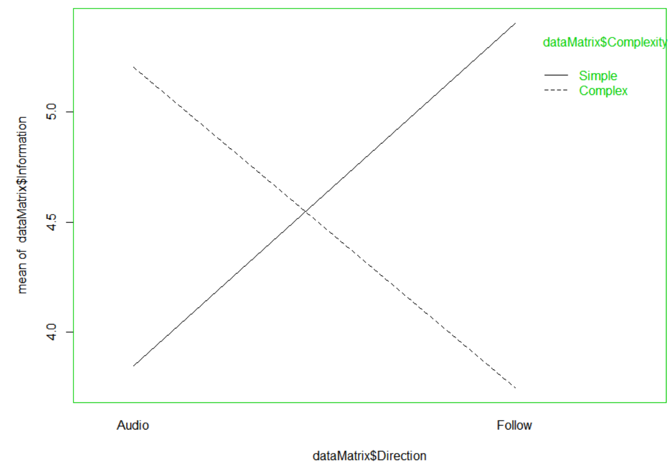
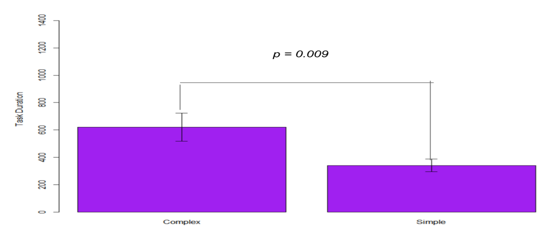
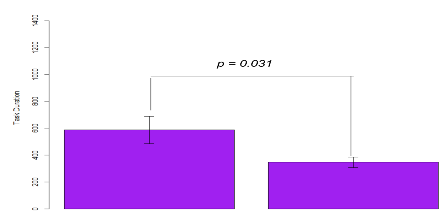


Fig. 5. Interaction Effect on Information Dissemination

$F(1, 15) = 8.72, p = 0.009$. "Mode of directions" as well had a significant main effect, $F(1, 15) = 5.65, p = 0.031$. Refer to Figure 6 for results.



(a) Effect of Path Complexity on Task Duration



(b) Effect of Mode of Direction on Task Duration

Fig. 6. Main Effect on Task Duration

We conducted a two-way analysis of variance (ANOVA) to test whether "mode of directions" and "path complexity" affected task difficulty and found two significant main effects as well as an interaction effect. "Path complexity" had a significant main effect on the task difficulty, $F(1, 15) = 7.12, p = 0.017$. "Mode of directions" also had a significant main effect on the task difficulty, $F(1, 15) = 19.64 p = 0.0004$. Both the conditions also had significant interaction effect on the task difficulty $F(1, 15) = 7.78, p = 0.014$. Refer to Figure 7 and Figure 8 for results.

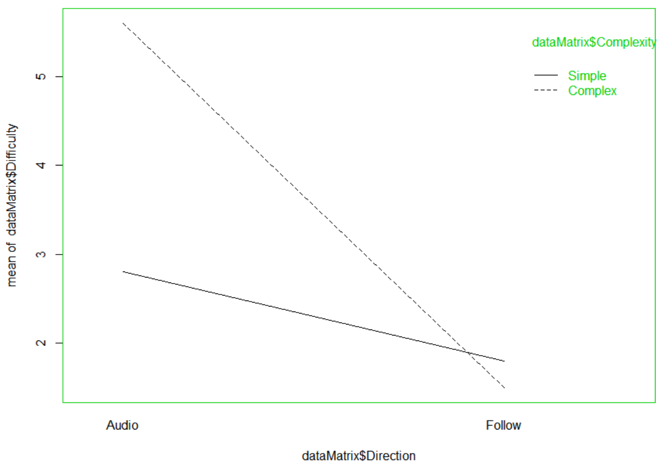
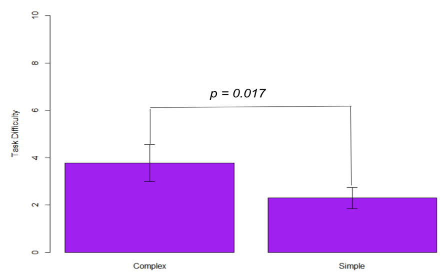
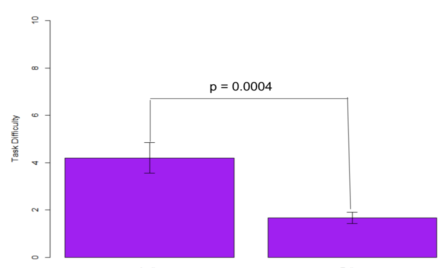


Fig. 7. Interaction Effect on Difficulty of Task



(a) Effect of Path Complexity on Task Difficulty



(b) Effect of Mode of Direction on Task Difficulty

Fig. 8. Main Effect on Task Difficulty

C. Interaction Experience

There were no significant main effects regarding the participants' interaction experiences from ”path complexity” nor ”mode of directions” received.

VI. Discussion

Our first hypothesis predicted that the completion success will be highest for participants who follow the robot with a map, while receiving directions via only audio with a map will be lowest. Our results support this hypothesis; the complexity of a task was taken into consideration when defining the mode in which directions are communicated. When the path complexity was high, in which the number of turns was high, participants tended to come back to the robot more to receive information than low path complexity.

Our second hypothesis predicted that participants will have a more beneficial experience in complex paths when following the robot compared to audio directions with a map. Experience here includes task difficulty and the participants feelings about the interaction with the robot. Our results do not support nor reject this hypothesis. On one hand task difficulty was reduced but on the other hand participants did not have a engaging interaction with the robot. More factors that take speed and social interaction into account should be taken into consideration when designing robot interaction. For instance, one participant mentioned that they ”never interacted with the bot. It was one way communication.”, which indicated a desired feature of social interaction when following the robot. In addition, the design of displaying the map also affected the comfort, one participant said that the map should face towards him/her when the robot was moving. This provides us pointers for modifying robot behavior for future studies.

Our third hypothesis predicted that when receiving only audio directions with a map, participants will have a more beneficial experience in simpler paths. Our results support this hypothesis; participants felt confused with longer complex paths when only receiving audio information. When following the robot participants found it to be too slow as one participant put it ”the robot was very slow.” One interesting finding was that one participant in ATLAS went straight to the destination when recognizing the location of the room on the map without following the robot.

The psychological and social aspects were important factors for participants to interact with robot. Participants felt that the process difficulty was lower to follow the robot with a map than to receive only audio directions. The process difficulty was significantly high in the condition of both only receiving the audio information and complex map due to the fact that this task required more cognitive memory to understand and remember the paths. Comparatively, following the robot consumed little short-term memory. 3 participants liked the fact that the TurtleBot was able to prevent obstacles.

VII. Limitations and Future Work

While following the robot did in fact reduce the difficulty perceived by participants, the viable locations, methods, and tools available to us did not extend themselves well to a current implementation for the ”wild”. For instance, the building in which we tested the audio and visual directions lacked a good enough Wi-Fi signal (in hallways the Wi-Fi would cut out, causing us to lose control over the robot) for us to test following the robot in the Engineering Center and had to limit the ”following robot” tasks to a smaller location with a stronger Wi-Fi signal (ATLAS Building). Along with not having a consistent floor plan, we were limited to a single floor due to the physical constraints of our robot (i.e. it couldn't go up stairs). Although the results of this study are limited to two different floor layouts, we feel that any future work in this field will agree with our results and demonstrate the effectiveness that an embodied robot can provide with respects to indoor navigation.

Due to the time constraints involved in running this study, we were limited in the autonomous functionality we could implement with our robot. For instance, we didn't have time to include autonomous movement to the destination and had to control it manually (unknown to the user) as a result. This autonomy feature should have the ability to perceive the environment, locate itself, avoid obstacles and calculate the optimal path. Along similar lines, we were not able to implement automatic speech recognition and had to control the responses manually. This lack of autonomy restricted the social capabilities of the robot in that it was not context-aware, meaning it could not engage in two-way communication. In addition, the lack of autonomous movement limited natural movements such that controlling the robot manually sometimes caused over correction and strange movements during the task. As a result of these limitations, future work in this field should attempt to increase the autonomy as this would increase the chances of receiving results more representative of what would be seen in the "wild". In addition to being more viable for public adoption, greater autonomy in contextual and movement could allow for the robot to be able to detect the complexity of a path to a given destination and assist the user in the most optimal way. For instance, if the destination was a single turn away, following the robot might not be necessary and audio or visual directions may be satisfactory.

Since this study was completed in a short experimental time frame (participant studies happened over the course of 3 days) we we were only able to gather 19 participants, majority of which were in the age range of 18 to 25 and attended collage. Due to the lack of diversity and the small number of participants who took part, we are not able to say with extreme confidence as to whether our results are viable at larger scales; however, we believe that our work provides a good point for larger and more diverse future studies.

VIII. Conclusion

To get the best results from embodied agents in service based positions, they must be socially capable, intuitive to use, and effective at the job they are completing. In this study, we designed a system in which a robotic agent that was able to interact with participants in a social manner attempted to assist with finding an unknown location in an indoor environment. We found a significant effect in which participants who were presented with audio and visual directions reported the tasks to be more difficult with respects to the simple path as well as the complex path in comparison to following the robot. One participant explained their experience receiving audio directions as "difficult to remember directions to a long distance - too many left rights can be confusing". With respect to task completion time, we found that participants who followed the robot had a quicker completion time than those who simply received audio and visual directions. Another result we found was that people were disappointed with the limited social interaction. These results demonstrate that receiving directions via audio and visual tend to be more difficult no matter the complexity of route whereas following the robot was perceived as a straight forward and simple task. We hope, while limited in its capabilities, our robot can be a starting point into further investigation on how communication with embodied robots can not only improve the experiences involved in finding an unfamiliar indoor destination, but acting in any type of service role as well.

Acknowledgments

A special thanks to the IronLab and Correll Lab for allowing us to use their equipment and lab areas.

References

[1] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, and Paul Rybski. Ripple effects of an embedded social agent: A field study of a social robot in the workplace. *Association for Computing Machinery*.

[2] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. Receptionist or information kiosk: How do people talk with a robot? In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 31–40. ACM, 2010.

[3] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Robot. Auton. Syst.*, 61(12):1726–1743, December 2013.

[4] Saeed Shiry Ghidary, Yasushi Nakata, Hiroshi Saito, Motofumi Hattori, and Toshi Takamori. Multi-modal interaction of human and home robot in the context of room map generation. *Auton. Robots*, 13(2):169–184, 2002.

[5] V. Alvarez-Santos, R. Iglesias, X. M. Pardo, C. V. Regueiro, and A. Canedo-Rodriguez. Gesture-based interaction with voice feedback for a tour-guide robot. *J. Vis. Comun. Image Represent.*, 25(2):499–509, February 2014.

[6] Nikolaos Mavridis. A review of verbal and non-verbal human-robot interactive communication. *Robot. Auton. Syst.*, 63(P1):22–35, January 2015.

[7] Wolfram Burgard, Armin B Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. Experiences with an interactive museum tour-guide robot. *Artificial intelligence*, 114(1-2):3–55, 1999.

[8] Dan Bohus, Chit W Saw, and Eric Horvitz. Directions robot: In-the-wild experiences and lessons learned. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 637–644. International Foundation for Autonomous Agents and Multiagent Systems, 2014.