Social Tag: Finding the person with the pink hat

Marek Michalowski and **Didac Busquets** and
Carl DiSalvo and **Laura Hiatt** and
Nicholas Melchior and **Reid Simmons**
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Selma Sabanovic
Rensselaer Polytechnic Institute
110 8th Street
Troy, NY 12180

Abstract

At the AAAI 2005 Robot Exhibition, the robot GRACE played a game of "social tag" that involved human-robot social interaction, navigation, and interface design. The task was for GRACE to locate and rendezvous with one of our team members, who was wearing a pink hat. The robot found the target not primarily through the modalities of sight or sound, but rather through social interactions with strangers in the environment.

Introduction

At the AAAI 2005 Robot Exhibition, the robot GRACE (Graduate Robot Attending a ConferencE) played a game of "social tag" in which the task was to locate and rendezvous with a team member who was wearing a pink hat. In this game, our purpose was not simply to create an object localization task (such as a scavenger hunt); rather, it was our intention to create a robot that could enlist the help of humans through frequent social interactions. We designed the game of "social tag" so that the robot's primary source of information about the whereabouts of its target came not through the modalities of sight or sound, but rather through social interactions with strangers in the environment. The task explores issues in human-robot interaction that involve shared space, intuitive interface design, and the negotiation of an environment filled with dynamic, untrained humans.

In this paper, we present our motivation for developing the game of social tag and describe the system we developed to perform this task. Our goal was to build a robot that was social and interactive enough that it was equally capable of requesting and accepting assistance (i.e., either the robot or a human could initiate an interaction that advanced the robot's progress). We present results that suggest that we were successful in this respect.

Background

GRACE

GRACE grew out of a multi-institution collaboration to design a robot capable of performing the AAAI Robot Challenge, which involves autonomously registering for the conference, navigating through the conference area, interacting with people, and delivering a talk. GRACE performed most of these tasks at AAAI 2002 (Simmons *et al.* 2003). The previously developed socially interactive components of GRACE are additionally described in (Bruce, Nourbakhsh, & Simmons 2002) and (Gockley *et al.* 2004). This line of work has been continued in research on designing robots for long-term interaction by incorporating a changing story line in the interactive repertoire of Valerie, a robot receptionist (Gockley *et al.* 2005).

Figure 1: The robot GRACE.

GRACE (Fig. 1) is a RWI B21 mobile robot. For this task, she also included a SICK laser scanner, a Canon VC-C4 PTZ camera, an LCD monitor with an animated face, and an ELO 1224 LCD touch screen. The robot has two computers on board: one for controlling mobility, sensing with the laser, avoiding obstacles, and handling the touch screen interface, and the other for vision, control of the face and voice, and general task control. The software architecture follows the design used in previous years for our entries in the Robot Challenge. Independent processes communicate via IPC message passing (http://www.cs.cmu.edu/∼IPC). No remote communications were used, except for startup and shutdown.

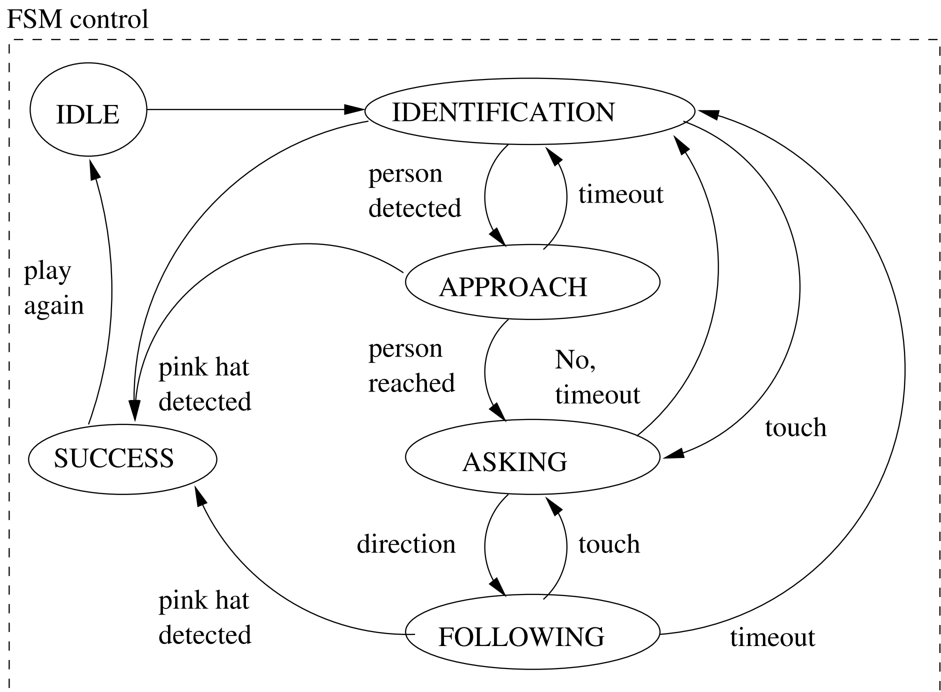

Figure 2: The finite state machine that comprises GRACE's control task for Social Tag.

"Social Tag"

The task of social tag – "finding the person with the pink hat" – was conceived as having a number of important benefits. The task was relatively simple and could be executed reliably, which allowed us to focus on the questions of human-robot interaction. Formulating the task as a fun and commonly-understood game increased the chances that people would be willing to interact with the robot. While these properties were important for designing and observing human-robot interaction, the interaction must serve a purpose; this was to locate the pink hat. The generality of object localization was another important property of our selected task, as it involves search, planning, navigation, etc., and can be applied to a wide range of useful robotic tasks.

In typical robot tasks involving detection of a visual target, distinctively colored objects such as pink hats are used to simplify the vision problem. However, in this case, the pink hat was intended to be as much for the benefit of other people as for GRACE herself. The team member was to be a highly visible individual who would be easily recognized and remembered by conference participants, so that it would be easy for them to help a wandering robot. Accordingly, it was not the completion of the goal (finding the pink hat) in which we were most interested, but rather in GRACE's journey and her social interactions along the way.

From a machine perception standpoint, the task of finding a hat in a crowded room is an extremely difficult one. While research in computer vision and speech recognition continues to improve robot perception, GRACE relies on the assistance of humans with fully developed senses of sight and hearing, and the ability to communicate. Indeed, there are cases in which no amount of sensory ability would enable successful task completion; if the hat is in a different room, even a human would need to enlist the help of other humans who are moving through the environment, communicating with each other, and collectively remembering the location of prominent objects. GRACE depends on her own sense of vision for the completion of the task (i.e. recognizing the pink hat), but her primary mode of gathering information is asking people for help in an intuitive and socially acceptable manner.

The task has five main phases:

1. *Identification* of approachable humans.
2. *Approach* toward a human with whom the robot would like to interact.
3. *Asking* for directions to the person with the pink hat.
4. *Following* those directions until a pink hat is found visually or more help is required.
5. Demonstrate *success* when the hat has been found.

As in typical human interactions, these phases may be regarded as a "script" that suggests an appropriate plan of action. However, the robot's human partner in the interaction may interfere or deviate from this script, and the robot should be able to handle such deviations in a context-dependent manner. Specifically, humans may volunteer to offer help when the robot is not actively looking for it, and GRACE can accept these offers during various phases of performing the task; this is a form of mixed-initiative interaction.

Implementation

Interaction and interface design

An LCD touch screen mediated the interaction between GRACE and conference attendees. The touch screen was mounted on the front of GRACE, below the screen that displayed her face, at approximately chest level with an average-sized adult.

The touch screen interaction consisted of a number of full-screen interfaces that both conveyed the state of the robot and prompted certain types of input. The general approach to designing the screens was to make them as bold and simple as possible while still being aesthetically appealing. To this end, we limited the amount of text on the screen and supplemented the screens with spoken information (via a text-to-speech system with lip-synchronization on the animated face). In order that GRACE's verbalizations do not become repetitive or boring, she chooses randomly from a library of appropriate phrases at each point in the task.

Music is another important component of GRACE's repertoire. For example, when GRACE begins looking for the person in the pink hat, she plays an excerpt from the theme music to "The Good, The Bad, and The Ugly." In addition to the music adding to the playful character of GRACE, it also functions as another mode of (auditory) expression to participants, emphasizing the current state of the robot.

The robot was controlled by a simple state machine (Fig. 2), where states correspond with the phases of the task as mentioned above. We now describe these phases and discuss significant implementation details.

Task phases

Identification & sensor fusion When the game is started, GRACE is in the `IDENTIFICATION` state. She is looking for the pink hat or for a person that might help her find it. The touch screen displays the *wandering* image (Fig. 3), which depicts a pink hat, with question marks, and the phrase "Touch Me. I am looking for a pink hat." The *wandering* screen serves three purposes: it informs people about what GRACE is doing; it provides an opportunity to interact with the robot; and it mitigates the limitations of the vision system by not relying on it to find people. Meanwhile, GRACE plays music and periodically says, "Where is the person in the pink hat?" Her animated face frequently changes direction, giving the appearance of actively looking around while wandering.

GRACE is equipped with a laser scanner near human knee-height and a camera near human face-height. The laser scanner clusters short adjacent range readings, labels those that appear to be human beings, and tracks those objects over time using a Kalman filter. The camera locates faces using appearance-based frontal face detectors and tracks them using skin color models. Data from these two sensors are combined to determine more reliably where there are people. This is done by registering the locations of camera-located faces and laser-located obstacles with each other in a robot-centric coordinate frame and labeling a laser obstacle as a person according to whether there is a face located above it.

Figure 3: The *wandering* screen.

A color model was obtained for the pink hat during a calibration procedure in the particular lighting environment of the conference center. If the number of pixels in the camera image that match this color model exceed a certain threshold, the hat is considered found. GRACE enters the `SUCCESS` state and the *Gotcha!* screen is displayed (Fig. 5), asking to be touched to start the game again.

If neither the pink hat nor a person is found after some time, the robot moves randomly and looks again. When a person (face) is detected, the robot enters the `APPROACH` state. However, if at any time someone touches the screen before being detected, GRACE directly enters the `ASKING` state.

Approach When a person is detected, GRACE first says, "I think I've found someone to ask," and begins to move toward the person to initiate an interaction. During this approach, the robot tries to observe societal norms such as speed, direction of approach, greeting (e.g. saying, "Excuse me!"), and personal space. These were designed from common sense understanding, and fine-tuned through testing. Without these behaviors, people who are not accustomed to interacting with robots may not realize that the robot is attempting to engage them, and they may move away from the robot rather than stand in her path.

When the approach is complete, if the person is still there (approximately four feet away), GRACE enters the `ASKING` state.

Asking In the `ASKING` state, GRACE displays a "Can you help me?" screen, with buttons "Yes" and "No." Meanwhile, she says "I am looking for the person in the pink hat. Can you help me?" In this case, the touch screen contains minimal text regarding the question, so as not to potentially burden the participant.

If the person presses "No," GRACE thanks the person, displays a *Thank you* screen, and returns to the `IDENTIFICATION` state.

If the person presses "Yes," GRACE asks the person to point her in the direction of the pink hat. To facilitate this, the *directions* image is displayed on the touch screen (Fig. 4). This image depicts GRACE in the center and eight arrows pointing outwards from the robot along with the line "Which way is the person in the pink hat?" The arrows are foreshortened to provide directional perspective from the participant's point of view.

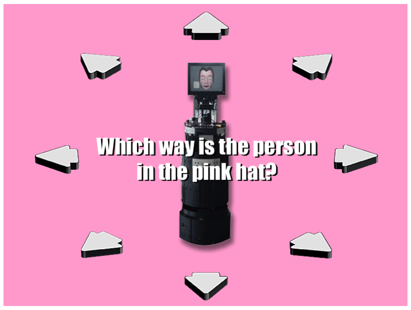

Figure 4: The *directions* interface.

Participants then touch the arrow that most accurately points in the direction of the person in the pink hat. For example, pressing on the top-most arrow causes GRACE to turn $180°$ and travel away from the participant. Once an arrow is pressed, GRACE briefly turns her animated face in the indicated direction (to help convey that the information was properly received), displays the *Thank you* screen, verbally thanks them, and enters the `FOLLOWING` state.

Following GRACE turns and follows the suggested direction, avoiding obstacles found by the laser scanner. At the same time, GRACE looks for the pink hat. If it is not found after traversing some minimum distance, GRACE returns to the `IDENTIFICATION` state and looks for another person to approach.

During this phase of the task, the *wandering* image is displayed on the touch screen. If the image is pressed at any point in this phase, GRACE stops and enters the `ASKING` state.

Whenever the robot needs to wait for human input, there is a timeout in case no input is given.

Success When GRACE finally locates the person in the pink hat, she displays the *Gotcha!* screen (Fig. 5). This may happen at any point during the task, except when she is interacting with someone in the `ASKING` state.

GRACE performs a dance (spinning two times in a circle) and plays the Aerosmith song "Pink." Given the distributed social nature of the task, it is not enough for the robot simply to finish: since she has distributed her perceptual burdens on others, it is important to inform them when the person in the pink hat is found. The success is not just the robot's, but is attributable to everyone who interacted with the robot. Thus the little dance provides "fun" visual and auditory feedback to all.

Results

Challenges

There were many challenges involving sensing: poor lighting, mirrors, and bright sunlight through the windows made it difficult to obtain a pink-color model that was effective yet discriminating; a satisfactory model was obtained only after coating the already pink hats with a more saturated pink

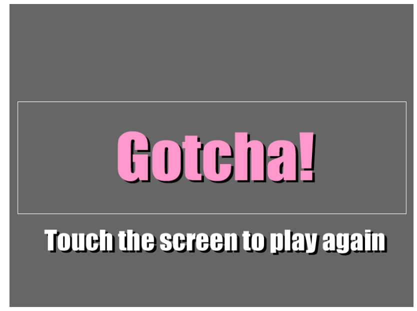

Figure 5: The *Gotcha!* screen.

paint. Still, pink or red shirts occasionally registered as the hat.

Additionally, computational constraints limited the resolution of our camera images and therefore the distance at which frontal faces would be detected. However, these challenges were fully anticipated in our formulation of the task and in our development of the robot; in fact, these challenges were the very motivation for this endeavor.

There were some instances of participants being confused about the orientation of the arrows in the interface relative to the robot in space; perhaps the interface was not completely self-evident. In particular, some participants provided the opposite direction, thinking they had to flip the frame of reference.

An additional challenge, exacerbated by the high density of conference attendees, was that participants would often stand in GRACE's way after giving her directions (despite GRACE's visible but unsuccessful attempts to find a path to her new goal). In retrospect, it would have been beneficial to implement a set of increasingly aggressive behaviors (e.g. facially, verbally, and in terms of motion). This would have been an appropriate context-dependent behavior: when a human helps the robot by giving a direction, the robot should resist immediate human interference until it is necessary once again. The museum tour-guide robot Minerva successfully employed such methods after her designers encountered similar difficulties (Thrun, Schulte, & Rosenberg 2000).

Successes

GRACE's ability to get to the hat was based on her capabilities as a socially interactive system rather than a mobile perceptual robot. Therefore, the humans interacting with GRACE can be seen as partners in GRACE's search for the pink hat rather than mere "users." The robot's success was not judged primarily by metrics such as time or turns to completion (although these would indicate how helpful people were). Rather, our evaluation of GRACE's performance looked at how interactions occurred and whether the robot was able to use interaction effectively in order to seek and obtain human help in achieving its goals.

As expected, many people were eager to help GRACE. We observed a great deal of full-body interaction, as people positioned themselves in front of the robot and gestured fruitlessly (by pointing and waving). Many people became invested in the task themselves, wanting to see the task through to completion. Several participants followed the robot around and repeatedly gave GRACE directions until she completed the task. On the other hand, many conference attendees actively ignored the robot even when it addressed them, which is perhaps not surprising at a conference on artifical intelligence. For this reason, the mixed-initiative strategy was effective and seemed natural. Another interesting phenomenon we observed was that GRACE was a catalyst for a large number of interpersonal interactions; conference attendees ended up communicating with each other regarding the location of the pink hat and how to help the robot.

Over the course of the three days of the conference, GRACE operated for approximately 15 hours, mostly during coffee breaks, receptions, and poster sessions. We logged the robot's activities for approximately half of this time. As we were interested in promoting and observing human-robot interaction, we considered a "good" run to be one in which GRACE interacted with at least four people before locating the hat. 53% of our logged interactions met this criterion (Fig. 6). Of the logged interactions, 85% were "helpful;" that is, a human interactor said that they could help and provided a direction. We did not record whether these directions were accurate, although it would have been interesting to compare people's honesty and usefulness when the robot asked for help versus when a human offered to help.

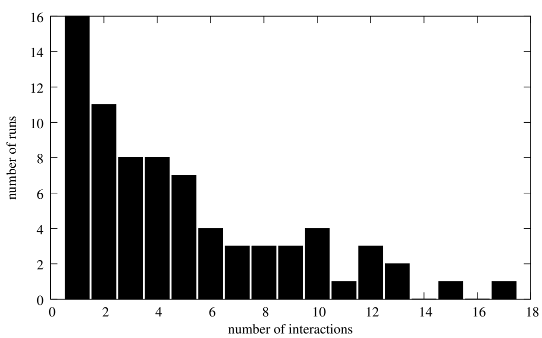

Figure 6: The number of logged runs for each interaction length.

When the robot initiated a logged interaction, humans provided help 185 out of 224 times (or 83%). When humans initiated the interaction, they provided help 152 out of 167 times (or 91%). We consider it a success that the robot initiated requests for help about as frequently as humans initiated offers of help. It is also a significant result that humans were more willing to help the robot if they initiated the interaction than if the robot initiated it; this result should be investigated further.

Fortunately, GRACE caused no injuries, flattened AIBOs, or broken mirrors.

Conclusion

We have introduced social tag as a way of developing and demonstrating a robot's ability to augment its sensory capabilities through social interaction. We have discussed the structure of the task, and we have described the design of GRACE's control system that allowed her to perform this task over the course of AAAI 2005. Our observations and results suggest that the robot was able to successfully use social interaction to request and accept assistance from conference participants with the aim of finding the person in the pink hat.

The game of social tag is a useful construct for testing a robot's ability to use people as a resource for enhancing its perceptual understanding of the environment. Furthermore, this ability is a potentially important design principle for mobile robots operating in human environments. Much of the work in human-robot social interaction is targeted toward a particular application, e.g. in the areas of entertainment, service robotics, or psychological research. We propose that *any* robot intended to perform tasks in human environments should be able not only to safely negotiate such environments, but also to actively engage the environment and its inhabitants in order to improve performance.

Acknowledgements

Thanks to Greg Armstrong and Brennan Sellner for technical contributions.

References

Bruce, A.; Nourbakhsh, I.; and Simmons, R. 2002. The role of expressiveness and attention in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '02)*.

Gockley, R.; Simmons, R.; Wang, J.; Busquets, D.; DiSalvo, C.; Caffrey, K.; Rosenthal, S.; Mink, J.; Thomas, S.; Adams, W.; Lauducci, T.; Bugajska, M.; Perzanowski, D.; and Schultz, A. 2004. GRACE and George: Social robots at AAAI (technical report ws-04-11). In *Proceedings of the National Conference on Artificial Intelligence (AAAI '04) Mobile Robot Competition Workshop*, 15–20. AAAI Press.

Gockley, R.; Bruce, A.; Forlizzi, J.; Michalowski, M.; Mundell, A.; Rosenthal, S.; Sellner, B.; Simmons, R.; Snipes, K.; Schultz, A. C.; and Wang, J. 2005. Designing robots for long-term social interaction. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS '05)*.

Simmons, R.; Goldberg, D.; Goode, A.; Montemerlo, M.; Roy, N.; Sellner, B.; Urmson, C.; Shultz, A.; Abramson, M.; Adams, W.; Atrash, A.; Bugajska, M.; Coblenz, M.; MacMahon, M.; Perzanowski, D.; Horswill, I.; Zubek, R.; Kortenkamp, D.; Wolfe, B.; Milam, T.; and Maxwell, B. 2003. GRACE: An autonomous robot for the AAAI Robot Challenge. *AAAI Magazine* 24(2):51–72.

Thrun, S.; Schulte, J.; and Rosenberg, C. 2000. Interaction with mobile robots in public places. *IEEE Intelligent Systems* 7–11.