Human-Robot Interaction in a Robotic Guide for the Visually Impaired

Vladimir Kulyukin * **Chaitanya Gharpure** **Nathan De Graw**
Computer Science Department
Utah State University
Logan, UT 84322
vladimir.kulyukin@usu.edu

Abstract

We present an assisted indoor navigation system for the visually impaired. The system consists of a mobile robotic guide and small sensors embedded in the environment. We describe the hardware and software components of the system and discuss several aspects of human-robot interaction that we observed in the initial stages of a pilot study with several visually impaired participants.

Introduction

Since the adoption of the Americans with Disabilities Act of 1990 that provided legal incentives for improvement in universal access, most of the research and development (R&D) has focused on removing structural barriers to universal access, e.g., retrofitting vehicles for wheelchair access, building ramps and bus lifts, improving wheelchair controls, and providing access to various devices through specialized interfaces, e.g., sip and puff, haptic, and Braille.

For the 11.4 million visually impaired people in the United States(LaPlante & Carlson 2000), this R&D has done little to remove the main functional barrier: the inability to navigate. This inability denies the visually impaired equal access to many private and public buildings, limits their use of public transportation, and makes the visually impaired a group with one of the highest unemployment rates (74%)(LaPlante & Carlson 2000). Thus, there is a clear need for systems that improve the wayfinding abilities of the visually impaired, especially in unfamiliar indoor environments, where conventional aids, such as white canes and guide dogs, are of limited use.

Related Work

Over the past three decades, some R&D has been dedicated to navigation devices for the visually impaired. Benjamin, Ali, and Schepis built the C-5 Laser Cane(Benjamin, Ali, & Schepis 1973). The cane uses optical triangulation with three laser diodes and three photo-diodes as receivers. Obstacles are detected at head height in a range of up to 3 meters in front of the user. Bissit and Heyes developed the Nottingham Obstacle Detector (NOD)(Bissit & Heyes 1980). NOD is a hand-held sonar device that gives the user auditory feedback with eight discrete levels. Each level distinguishes a discrete distance value and plays different musical tones.

More recently, Shoval et al. developed the NavBelt, an obstacle avoidance wearable device equipped with ultrasonic sensors and a wearable computer(Shoval, Borenstein, & Koren 1994). The NavBelt produces a 120-degree wide view ahead of the user. The view is translated into stereophonic audio directions that allow the user to determine which directions are blocked. Borenstein and Ulrich built GuideCane(Borenstein & Ulrich 1994), a mobile obstacle avoidance device for the visually impaired. GuideCane consists of a long handle and a sensor unit mounted on a steerable two-wheel axle. The sensor unit consists of ultrasonic sensors that detect obstacles and help the user to steer the device around them. The Haptica Corporation has developed Guido©, a robotic walking frame for people with impaired vision and reduced mobility(http://www.haptica.com/walker.html). Guido© uses the onboard sonars to scan the immediate environment for obstacles and communicates detected obstacles to the user via speech synthesis.

Limitations of Prior Work

While the existing approaches to assisted navigation have shown promise, they have had limited success due to their inability to reduce the user's navigation-related physical and cognitive loads.

Many existing systems increase the user's navigation-related physical load because they require that the user wear additional and, oftentimes substantial, body gear (Shoval, Borenstein, & Koren 1994), which contributes to physical fatigue. The solutions that attempt to minimize body gear, e.g., the C-5 Laser Cane(Benjamin, Ali, & Schepis 1973) and the GuideCane(Borenstein & Ulrich 1994), require that the user abandon her conventional navigation aid, e.g., a white cane or a guide dog, which is not acceptable to many

*This research has been supported, in part, through a Community University Research Initiative (CURI) grant from the State of Utah and through a New Faculty Research grant from Utah State University.

visually impaired individuals.

The user's navigation-related cognitive load remains high, because the user makes all final navigation decisions. Device-assisted navigation focuses on obstacle avoidance but provides little cognitive improvement over its conventional counterparts, i.e., canes and guide dogs.

Robot-Assisted Navigation

Robot-assisted navigation can help the visually impaired overcome these limitations. First, the amount of body gear carried by the user is significantly minimized, because most of it can be mounted on the robot and powered from onboard batteries. Consequently, the navigation-related physical load is significantly reduced. Second, the user can interact with the robot in ways unimaginable with guide dogs and white canes, i.e., speech, wearable keyboard, audio, etc. These interaction modes make the user feel more at ease and reduce her navigation-related cognitive load. Third, the robot can interact with other people in the environment, e.g., ask them to yield or receive instructions. Fourth, robotic guides can carry useful payloads, e.g., suitcases and grocery bags. Finally, the user can use robotic guides in conjunction with her conventional navigation aids.

Are all environments suitable for robotic guides? No. There is little need for such guides in familiar environments where conventional navigation aids are adequate. However, unfamiliar environments, e.g., airports, conference centers, and office spaces, are a perfect niche for robotic guides. Guide dogs, white canes, and other navigation devices are of limited use in such environments because they cannot help their users find paths to useful destinations.

The idea of robotic guides is not new. Horswill(Horswill 1993) used the situated activity theory to build Polly, a mobile robot guide for the MIT AI Lab. Polly used lightweight vision routines that depended on textures specific to the lab. Thrun et al.(Thrun *et al.* 1999) built Minerva, a completely autonomous tour-guide robot that was deployed in the National Museum of American History in Washington, D.C.

Unfortunately, neither project addresses the needs of the visually impaired. Both depend on the users' ability to maintain visual contact with the guides, which cannot be assumed for the visually impaired. Polly has very limited interaction capabilities: the only way users can interact with the system is by tapping their feet. In addition, the approach on which Polly is based requires that a robot be evolved by its designer to fit its environment not only in terms of software but also in terms of hardware. This makes it difficult to develop robotic guides that can be deployed in new environments, e.g., conference halls, in a matter of hours by technicians who know little about robotics. Completely autonomous solutions like Minerva that attempt to do everything on their own are expensive and require substantial investments in customized engineering to become operational, which makes them hard to reproduce.

A Robotic Guide

We have built a prototype of a robotic guide for the visually impaired. Its name is RG, which stands for "robotic guide."

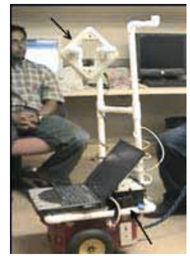

Figure 1: RG.

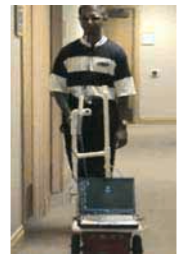

Figure 2: RG guiding a visually impaired person.

We refer to the approach behind RG as *non-intrusive instrumentation of man-made environments*. The idea is to instrument the environment with inexpensive and reliable sensors in such a way that no activities indigenous to that environment are disrupted. Additional requirements are that the instrumentation be fast and require only commercial-off-the-shelf (COTS) hardware components. Effectively, the environment becomes a distributed tracking and guidance system (Kulyukin & Blair 2003) that consists of stationary nodes, i.e., computers and sensors, and mobile nodes, i.e., robotic guides.

Hardware

RG is built on top of the Pioneer 2DX commercial robotic platform (http://www.activmedia.com) (See Figure 1). The platform has three wheels, 16 ultrasonic sonars, 8 in front and 8 in the back, and is equipped with three rechargeable Power Sonic PS-1270 onboard batteries that can operate for up to 10 hours at a time.

Figure 3: An RFID tag attached to a wall.

What turns the platform into a robotic guide is a Wearable Wayfinding Toolkit (WWT) mounted on top of the platform and powered from the on-board batteries. As can be seen in Figure 1, the WWT resides in a PCV pipe structure attached to the top of the platform. The WWT's core component is a Dell Inspiron I820 laptop connected to the platform's microcontroller. The laptop has a Pentium 4 mobile 1.6 GHz processor with 512 MB of RAM. Communication between the laptop and the microcontroller is done through a TrippLite USB to Serial cable. The laptop has an Orinoco Gold PC Card 802.11b wireless card that allows for remote wireless connectivity.

The laptop interfaces to a radio-frequency identification (RFID) reader through another USB to Serial cable. The TI Series 2000 RFID reader is connected to a square 200mm by 200mm RFID RI-ANT-GO2E antenna that detects RFID tags placed in the environment. In Figure 1, the arrow in the top left corner points to the RFID antenna; the arrow in the bottom right corner points to the RFID reader behind the laptop's screen. The arrow in Figure 3 points to a TI RFID Slim Disk tag attached to a wall.

These tags can be attached to any objects in the environment or worn on clothing. They do not require any external power source or direct line of sight to be detected by the RFID reader. They are activated by the spherical electromagnetic field generated by the RFID antenna with a radius of approximately 1.5 meters. Each tag is programmatically assigned a unique ID. A dog leash is attached to the battery bay handle on the back of the platform. The upper end of the leash is hung on a PCV pole next to the RFID antenna's pole. Visually impaired individuals follow RG by holding onto that leash. Figure 2 shows a visually impaired person following RG.

Software

RG is implemented as a distributed system. The WWT laptop runs the following low-level robotic routines: follow-wall, avoid-obstacles, go-thru-doorway, pass-doorway, and make-hallway-uturn. These routines are written in the behavior programming language of the ActivMedia Robotics Interface for Applications (ARIA) system from ActivMedia Robotics, Inc. The laptop also runs a speech recognition and synthesis engine that enables RG to receive and synthesize speech, the Map Server, and the Path Planner. The advantages and disadvantages of speech-based interaction are discussed in the next section.

The Map Server is implemented as an OpenCyc server (http://www.opencyc.org). The server's knowledge base represents an aerial view of the environment in which RG operates. Currently, the knowledge base consists of floor maps, tag to destination mappings, and low-level action scripts associated with specific tags. The base also registers the latest position of RG, which is sent there as soon as a tag is detected in the environment.

The environment is represented as a graph where nodes represent the RFID tags and the edges represent the actions required to travel from one tag to another. Since RG is designed to follow the right wall, the graph is directed. For example, the following assertion in the OpenCyc's knowledge representation language CYCL represents a graph node:

```
(#$rfidTag 5
  (#$TheList
    (#$TheList 82
      (#$sMakeHallwayUturn)
      (#$sFollowWall 1))
    (#$TheList 6
      (#$sGoThruDoorway)
      (#$sFollowWall 1)))
  (#$TheList 4))
```

This assertion states that this node is represented by the RFID tag whose ID is 5. Second argument to the predicate `#$rfidTag` is a list of nodes that can be reached from node 5. In this example, from node 5 one can reach nodes 82 and 6. Each reachable node has a sequence of actions associated with it. A single action is represented by a predicate, such as `#$sFollowWall`. The only argument to the `$sFollowWall` predicate is 0 or 1 with 0 standing for the left wall and 1 standing for the right wall. In the above example, tag 82 can be reached from tag 5 by first making a hallway u-turn and then following the right wall until tag 82 is detected. Similarly, tag 6 can be reached from tag 5 by first going through a doorway and then following the right wall.

Of course, action specifications, such as `#$sFollowWall`, are robot-specific. In selecting actions, we tried to find a minimal action set that can be used in many standard indoor environments. As new actions are developed for new environments, they can be easily specified in this knowledge formalism. Other platforms can use this formalism to describe their platform-specific actions.

The tag to destination mappings are represented by assertions such as the following:

```
(#$tag-destination
  #$John.Doe 23
  (#$TheList (#sPassDoorway)))
```

In this assertion, tag 23 corresponds to John Doe's office. The third argument to the predicate `#$tag-destination` represents a possibly empty sequence of actions to be taken before stopping, once the destination tag has been detected. For example, in the above example, once RG has detected tag 23, it continues to move until the doorway is passed regardless of whether the door is open or closed. Such actions allow RG to position itself at a destination in such a way that the destination is easily accessible to the visually impaired user. For example, RG always moves past the door of a desired office so that when it stops the visually impaired user is at the door.

The Path Planner uses the standard breadth first search algorithm to find a path from one location to the other. The Planner uses the Map Server for the graph connectivity information and generates a path plan in the form of a sequence of tag numbers and action sequences at each tag.

Human-Robot Interaction

Humans can interact with RG either through speech or GUIs. Currently, speech-based interaction is intended for visually impaired individuals. Speech is received by RG through a wireless microphone placed on the user's clothing. Speech is recognized and synthesized with Microsoft Speech API (SAPI) 5.1, which is freely available from www.microsoft.com/speech. SAPI includes the Microsoft English SR Engine Version 5, a state-of-the-art Hidden Markov Model speech recognition engine. The engine includes 60,000 English words, which we found adequate for our purposes.

SAPI couples the Hidden Markov Model speech recognition with a system for constraining speech inputs with context-free command and control grammars. The grammars constrain speech recognition sufficiently to eliminate user training and provide speaker-independent speech recognition. This ability to constrain speech input was an important consideration for our system, because visually impaired people need to be able to interact with robotic guides in situations that offer no training opportunities.

Grammars are defined with XML Data Type Definitions (DTDs). Below is a truncated rule from the context-free grammar used in the system.

```
<RULE NAME="RGActions"
   <L> <P>wake up R G</P>
       <P>what can i say</P>
       <P>where am i</P>
       <P>stop guide</P>
   </L>
</RULE>
```

GUI-based interactions are reserved for system administrators. The notion of a system administrator is construed rather broadly. It can be a technician installing the system or an administrative assistant telling the system that a specific region is blocked for two hours due to a special event. For example, a system administrator can place an RFID tag on a new object in the environment, i.e., a soda machine, and add a new tag-object pair to the OpenCyc knowledge base. Such updates prompt administrators for brief written English descriptions of the tagged objects. These descriptions are used to dynamically add rules to RG's command and control grammar. The administrator can block access to part of the environment through the GUI displaying the environment's map. Until the block is removed, the Path Planner will build paths avoiding the blocked region.

RG interacts with its users and people in the environment through speech and audio synthesis. For example, when RG is passing a water cooler, it can either say "water cooler" or play an audio file with sounds of water bubbles. We added non-speech audio messages to the system because, as recent research findings indicate(Tran, Letowski, & Abouchacra 2000), speech perception can be slow and prone to block ambient sounds from the environment. On the other hand, associating objects and events with non-speech audio messages requires training or the presence of a universally accepted mapping between events and objects

Figure 4: Microsoft's Merlin personifying RG.

and sounds. Since no such mapping is currently available, our assumption is that the user can quickly create such a mapping. The motivation is to reduce the number of annoying interactions by allowing users to specify their own audio preferences. Once such a mapping is created, the user can upload it to robotic guides she will use in different environments. Such an upload can be web-based and can be easily accomplished with standard screen readers like Jaws (http://www.freedomscientific.com) available on portable and static computing devices.

We have built a tool that allows visually impaired users to create their own audio associations. The tool associates a set of 10 standard events and objects, e.g., water cooler to the right, about to turn left, bathroom on the right, etc., with three audio messages: one speech message and two non-speech messages. This small number was chosen because we wanted to eliminate steep learning curves.

To other people in the environment, RG is personified as Merlin, a Microsoft software character, always present on the WWT laptop's screen (see Figure 4). When RG encounters an obstacle, it assumes that it is a person and, through Merlin, politely asks the person to yield the way. If, after a brief waiting period, the obstacle is gone, RG continues on its route. If the obstacle is still present, RG attempts to avoid it. If the obstacle cannot be avoided, i.e., a hallway is completely blocked, RG informs the user about it and asks the Path Planner to build a new path, thus responding to changes in the environment on the fly.

Discussion

The system has been deployed for hours at a time at the Computer Science Department of Utah State University (the target environment). The department occupies an entire floor in a multi-floor building. The floor's area is 21,600 square feet. The floor contains 23 offices, 7 laboratories, a conference room, a student lounge, a tutor room, two elevators, several bathrooms, and two staircases.

One hundred RFID tags were deployed to cover the desired destinations. Once the destinations are known, it takes one person 30 minutes to deploy the tags and about 20 minutes to remove them. The tags are attached to objects with regular scotch tape. The creation of the OpenCyc knowledge base takes about 2 hours: the administrator walks around the area with a laptop and records tag-destination associations. The administrator can also associate specific robotic actions with tags.

We have recruited five visually impaired participants for a pilot study of robot-assisted navigation and have conducted a few trial runs with the participants in the target environment. However, since we have not completed all of the planned experiments and do not have complete data sets, our observations below are confined to anecdotes. A systematic analysis of our pilot study and its findings will be published elsewhere.

We have encountered several problems with speech recognition. When done in noisy environments, simple commands are often not understood or understood incorrectly. For example, two throat clearing sounds are sometimes recognized by SAPI as the phrase "men's room." This caused problems in several experiments with live participants, because RG suddenly changed its route. Another problem with speech recognition occurs when the person guided by RG stops and engages in conversation with someone. Since speech recognition runs continuously, some phrases said by the person are erroneously recognized as route directives, which causes RG to start moving. In one experiment, RG erroneously recognized a directive and started pulling its user away from his interlocutor until the user's stop command pacified it. In another experiment, RG managed to run a few meters away from its user, because the user hung the leash on the PCV pole when he stopped to talk to a friend of his in a hallway. Thus, after saying "Stop," the user had to grope his way along a wall to RG, standing a few meters away.

As was argued elsewhere (Kulyukin 2003; 2004), it is unlikely that these problems can be solved on the software level until there is a substantial improvement in the state-of-the-art speech recognition. Of course, one could add yes-no route change confirmation interactions. However, since unintended speech recognition is frequent, such interactions could become annoying to the user. Therefore, we intend to seek a wearable hardware solution. Specifically, we will investigate human-robot interaction through a wearable keyboard. Many wearable keyboards now fit in the palm of one's hand or can be worn as badges. The keyboard will directly interface to the WWT laptop. When a guided person stops to talk to someone, one button push can disable the speech recognition process for the duration of the conversation. Similarly, when the guided person clears her throat and RG misinterprets it as a command, one button push can tell RG to ignore the command and stay on the route.

Potentially, the wearable keyboard may replace speech recognition altogether. The obvious advantage is that keyboard-based interaction eliminates the input ambiguity problems of speech recognition. One potential disadvantage is the learning curve required of a human subject to master the necessary key combinations. Another potential disadvantage is restrictions on the quality and quantity of interactions due to the small number of keys. Additional experiments with human participants will determine the validity of these speculations.

Conclusion

We presented an assisted indoor navigation system for the visually impaired. The system consists of a mobile robotic guide and small RFID sensors, i.e., passive RFID transponders, embedded in the environment. The system assists visually impaired individuals in navigating unfamiliar indoor environments.

Acknowledgments

We would like to thank Martin Blair, Director of the Utah Assistive Technology Program, for his administrative help. We would like to thank Sachin Pavithran, a visually impaired training and development specialist at the Center for Persons with Disabilities at Utah State University, for design suggestions and for his assistance with recruiting subjects for our experiments.

References

Benjamin, J. M.; Ali, N. A.; and Schepis, A. F. 1973. A Laser Cane for the Blind. In *San Diego Medical Symposium*.

Bissit, D., and Heyes, A. 1980. An Application of Biofeedback in the Rehabilitation of the Blind. *Applied Ergonomics* 11(1):31–33.

Borenstein, J., and Ulrich, I. 1994. The GuideCane - A Computerized Travel Guide for the Active Guidance of Blind Pedestrians. In *IEEE International Conference on Robotics and Automation*.

Horswill, I. 1993. Polly: A Vision-Based Artificial Agent. In *Proceedings of the 11th Conference of the American Association for Artificial Intelligence (AAAI-93)*.

http://www.activmedia.com. *ActivMedia Robotic Platforms*. ActivMedia Robotics, Inc.

http://www.freedomscientific.com. *Blind and Low Vision Devices*. Freedom Scientific, Inc.

http://www.haptica.com/walker.html. *Guido*. Haptica Corporation.

http://www.opencyc.org. *The OpenCyc Project: Formalized Common Knowledge*. Cycorp, Inc.

Kulyukin, V., and Blair, M. 2003. Distributed Tracking and Guidance in Indoor Environments. In *Conference of the Rehabilitation Engineering and Assistive Technology Society of North America (RESNA-2003)*.

Kulyukin, V. 2003. Towards Hands-Free Human-Robot Interaction through Spoken Dialog. In *AAAI Spring Symposium on Human Interaction with Autonomous Systems in Complex Environments*.

Kulyukin, V. 2004. Human-Robot Interaction through Gesture-Free Spoken Dialogue. *Autonomous Robots,* to appear.

LaPlante, M. P., and Carlson, D. 2000. *Disability in the United States: Prevalence and Causes*. Washington, DC: U.S. Department of Education, National Institute of Disability and Rehabilitation Research.

Shoval, S.; Borenstein, J.; and Koren, Y. 1994. Mobile Robot Obstacle Avoidance in a Computerized Travel for the Blind. In *IEEE International Conference on Robotics and Automation*.

Thrun, S.; Bennewitz, M.; Burgard, W.; Cremers, A. B.; Dellaert, F.; Fox, D.; Hahnel, D.; Rosenberg, C.; Roby, N.; Schutle, J.; and Schultz, D. 1999. Minerva: A Second Generation Mobile Tour-Guide Robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA-99)*.

Tran, T. V.; Letowski, T.; and Abouchacra, K. S. 2000. Evaluation of Acoustic Beacon Characteristics for Navigation Tasks. *Ergonomics* 43(6):807–827.