Mobile Robots for an E-Mail Interface for People Who Are Blind

V. Estivill-Castro and S. Seymon

School of ICT, Griffith University,
Nathan QLD 4111, Brisbane, Australia

Abstract. The availability of inexpensive robotic hardware has brought to realization the dream of having autonomous mobile robots around us. As such, the research community has recently manifested more interest in assisting robotic technology (see proceedings of the last two IEEE RO-MAN conferences, the emergence of the RoboCup@Home challenge at RoboCup and the first annual Human Computer Interaction Conference jointly sponsored by IEEE and ACM). Robots provide to the blind what was lost as textual interfaces were replaced by GUIs. This paper describes the design, implementation and testing of a first prototype of a multi-modal Human-Robot Interface for people with Vision Impairment. The robot used is the commercially available four legged SONY Aibo.

1 Introduction

It is now accepted that domestic robots will provide entertainment, care, and perform household chores. Inexpensive hardware has brought the realization that robots are moving out of industrial settings and are sharing the environments once considered only for humans. While robots may still be a long way from having generic capabilities, they can perform many useful, socially meaningful and productive tasks (like rescue robots or land-mine finding robots). Among the contributions that mobile robotics can make to the well being of humans are the entertainment, assistive and supportive roles for elderly humans, peoples with disabilities [19] and tutors for human [17]. However, for robots to constitute a useful interface between humans and intelligent systems or ubiquitous computing, many usability and Robot-Human interaction issues must be resolved.

For access by the blind, the conversion for textual output was considered trivial because of Text-To-Speech Technology or Braille output devices [8] while input could be handled by training on a keyboard. The introduction of GUIs and the mouse requires spatial display and more visual interaction resulting in challenging barriers for computer usage by the blind [18]. Multi-modal interaction can significantly facilitate computer usage by the blind [2]. While there has been significant interest in Robot-Human Interaction [3] and in assistive robotics [16], the suggestion of using robots as the multi-modal interface between the blind and ubiquitous computing has received very little attention [1,14]. *Ubiquitous Computing* is considered the next phase of ICT [10]. In this *Ambient Intelligence* [12],

G. Lakemeyer et al. (Eds.): RoboCup 2006, LNAI 4434, pp. 338–346, 2007.

buildings, domestic appliances, cars and many other devices are to provide hardware for intelligent agents that jointly provide Intelligent Environments. Such environments will be homes and offices where domestic robots would be part of the human-environment interface. Most approaches of robotic applications focus on assistive technologies [4] and orientation and navigation [6,7]. However, our approach is to consider the robot as a multi-modal replacement for the computer mouse and visual display. Robots as multi-modal interfaces have recently attracted interest [5]. We show that while general principles of user interface design are applicable to robots as interfaces, some new considerations emerge. In particular, users are not so prepared to accept robots that appear spontaneous or surprising while the user is attempting to complete a task.

The amalgamation of capabilities by computer technology in TVs, desktops, digital cameras and mobile devices makes possible what seemed impossible. Mobiles phones seemed useless to the deaf, but now, they are a flexible tool trough SMS and actions (like vibration) for incoming calls. These devices are becoming multi-modal interfaces and robotic capabilities would enable even more channels of communication through gestures and embodiment. Our thesis is that robots are to become an effective multi-modal interface for people who are blind. This thesis is illustrated with a prototype that uses the Sony AIBO legged robot as a multi-modal interface to applications on a PC with a 802.11b wireless card and Internet access. The Sony AIBO ERS-210 has a wireless card for an ad-hoc wireless connection to the PC. The PC provides the connectivity node and configuration facilities to enable tasks such as browsing the Internet, receiving and sending e-mail, and playing and recording audio files. The modes of communication for input include, physical manipulation, speech recognition, gesture recognition, and even posture. Output are also gestures of the robot, but the main output is sound since the intended users are people who are blind. Physical manipulation in this context is the touching of buttons, and movement of extremities, including the head and tail. A camera on the "nose" of the robot enables gesture recognition. An on-board speaker allows the system (through synthesised speech) to communicate available menus and options and request input. Sensors on all the joints allows for the monitoring of their positions. The system tracks when a joint is moved beyond a threshold from its current target position (because of human user shifting the limb), it triggers a signal that we can marshal and send a specific message to the PC. The system implements a state machine in the PC with state transitions that are the result of the joint moved and the current state. For instance, the moving of the tail to the left sets the system to select the next track, if it was in the audio player state.

We found [1] the ERS-210 provided enough spatial information and its design provided sufficient physical clues that made the project feasible. Fundamental operational issues were resolved. Namely, the robot could have its battery replaced and charged, and it would be strong and reliable enough to be operated manually by a blind person (for the ERS-7 the' outcome was unsatisfactory). Although there were some significant shortcoming in relation to tactile recognition of buttons, we added different textures to the three buttons [1].

Using the User Centred Design approach, we had several iterations to revise the specification of some aspects of the interface and the protocol. Two blind University students interacted 3 times each to evaluate and provide feedback on the usability of the system. Their input influenced the interface design.

2 Evaluation

We conducted two types of evaluations to draw conclusions regarding human-robot interaction. These were both usability assessments. Firstly, we video two full sessions with two blind university students. This process not only validated the User Centred Design that lead to a deployment where usability issues had been resolved, it also confirmed people who were blind from birth could operate it. The second technique was required due to the low numbers of subjects of the first evaluation. Sessions were now conducted with 15 full sighted adults. This subjects were not familiar with the Sony AIBO at all and wore a blindfold before the robot was presented and throughout the entire duration of the test.

In both case, each session lasted for 20 to 25 minutes. The actual exercise lasted between 15 and 20 minutes whilst a questionnaire was completed in the remaining time. The first part of the experiment involves the participant becoming familiar with the robot (mostly touching it, and experiencing the feedback of joints as well as its sounds). Participants were requested to pick the robot up, and explore it themselves. Once having completed this, the instructor would then direct their fingers to points of interest such as the tail, the textured buttons on the back and head, and the on/off button on the chest. Participant were asked to turn on the Sony AIBO and three applications were used to demonstrate the capability of the robotic interface. We refer to them as the *Audio Player*, the *Voice Recorder*, and the *E-mailer*. Participants were guided through the steps of each application once. Then, they were required to accomplish a simple task on each. For example, with the *Voice Recorder*, they were asked to to record a short message, and then review it. They are also given the opportunity to hear the messages in the out-box. The participant were taken through the process of composing an e-mail, send it and then verify that they could retrieve an e-mail.

Table 1. (a) Best default setting for mnemonic commands for the *Audio Player*. (b) Best default settings for mnemonic universal commands.

Audio Player mnemonic commands

Command	User's Action
Play	Head down
Stop	Head up
Pause	Tail up
Next Track	Tail left
Previous Track	Tail right

(a)

Universal mnemonic commands

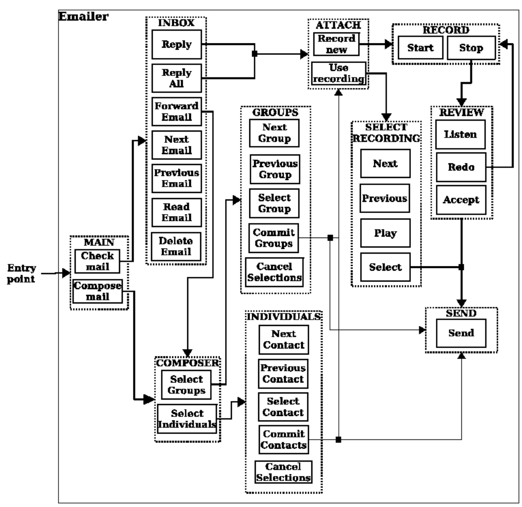
Command	User's Action
Back	Button on the back
Stop Application	Close mouth

(b)

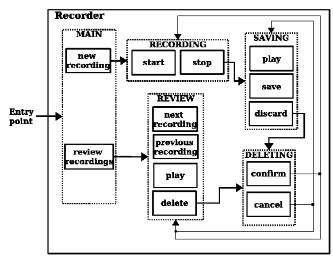
The Sony AIBO takes some time to initialise. The *Audio Player* is easiest and has basic controls of a media player. These are PLAY, STOP, PAUSE, NEXT TRACK, and PREVIOUS TRACK. Our usual mapping of these controls to the actual actions on the robot is listed in Table 1. While this mapping between actions (from the user on the robot) to effects of the *Audio Player* can be re-configured, we have found that the 4-legged shape of the robot allows for the user actions to be mnemonic. For example, the head movement down is a gesture of confirmation, acceptance and even obedience in dogs and many cultures.

The architecture uses the PC to manage the actual media. For example, a CD with musical content exist on the local disk of the PC. Typically, one can configure the track numbering by copying the audio-files to a directory. When music is selected, the sound is directed to the speakers of the PC (rather than the Sony AIBO) or another set of speakers in the environment. This achieves two issues. At a minimum, the music (or other audio, like iPod radio programs) is played through the speakers on the PC with higher clarity, and avoids high volume traffic on wireless networks. But more importantly, the intention of the *Audio Player* is not to reproduce media content on the Sony AIBO but in the environment in the sense of *Ambient Intelligence*. Thus, reproducing audio on the environment allows us to use the Sony AIBO clearly as an interface between the human and the surrounding environment.

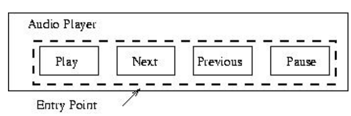
Within each application there can be several different contexts [13]. A context can be seen as a step in a process. For instance, in the *E-mailer*, when users first enters the application, they need to specify whether they want to send an



(a) The contexts for the *E-mailer* application.



(b) The contexts for the *Voice Recorder*



(c) The diagram of contexts for the *Audio Player*.

Fig. 1. The context diagrams reflect the complexity of the application

email or review the in-box. This decision is the context. As there are only two things to do here, there will be only two actions mapped. Each context has its own button configuration, complementing the universal commands (refer to Table 1 (b)). Moving between contexts is the result of actions on the robot. For example, by selecting to review the in-box, the user moves to a different context where different options are possible (among the options in the new context we find listening to an e-mail in the in-box). We can measure the complexity of an application. The *Audio Player* is simple because it comprises of just one context. The universal commands listed in Table 1 (b) apply to every context. Touching the back of the Sony AIBO results in users moved to the previous context. That is, they are taken *back* to the previous menu (provided this makes sense). In some instances it is not possible to go back a step in the process as the action is permanent. For example, sending an email can not be undone. The back button effectively behaves as an an *undo* function. Whichever application is running, closing the mouth will stop the application and return the user to the root context of the system. The mnemonic for this is based upon the recommendation that to break and stop a dog fight, one shall grab the jaw of the dog.

Voice Recorder has two functions, recording and reviewing voice messages. The participants voice is captured on the stereo microphones on the Sony AIBO's head. The sound is streamed to the PC as an audio file. Logically, this constitutes an out-box of pre-recorded messages. It is not possible to use the on-board storage of the Sony AIBO as memory sticks are unsuitable for recording large amounts of audio.

We performed the developed process with an understanding of user interface design, but not adhering to any specific methodology, in order to allow more freedom that the prescriptive approaches and guidelines that the literature recommends for human computer interaction [11,13,15]. Once we completed the deployment, we compared with the guidelines suggested in the literature. These guidelines are in agreement with our final product. For illustration, we present a discussion regarding Shneiderman's "Eight Golden Rules of Interface Design" [13]. Rule 1 says "Strive for consistency". We found this rule necessary and applicable as we consistently trough contexts we used the tail to allow users to scroll forwards or backwards through lists. Rule 3 indicates "Offer informative feedback", we found necessary for the applications to respond to some commands by synthesised phrases. Like when users entered the *E-mailer* the robot would respond with the phrase "Emailer enabled". We also provided visual feedback trough the leds in the face of the Sony AIBO. While these are clearly of no value to blind users, they are helpful to trainers as well as for maintenance, and even set up by sighting people. Rule 4 mandates to "Design dialog to yield closure"; that is, to organise actions into groups with a beginning, a middle and an end. This is precisely what contexts achieve. Our *Back* button achieves Rule 5, namely, "Offer simple error handling" and Rule 6 "Permit easy reversal of actions". However, to stress the point in Rule 5 and prevent serious errors we believe it is intuitive to require, before an irreversible action is performed, confirmation where it is required to move the head of the robot down. We found that users were rapidly in control of the robot. They also felt in control. Thus, we believe we concur with Rule 7 that dictates one should “Support internal locus of control”. The applications are simple; however, we found it necessary to provide users with a universal command that could repeat the options available and in this way comply with Rule 8 “reduce short-term memory load.”

3 Discussion

Assistive technologies overcome some problems people who are blind face when using a GUI. These include Screen Readers and zoom displays. However, these are merely add-ons onto a system that has been designed for people whose main channel of interaction is sight. One would expect that commercial products would have been available to assist the blind in using e-mail and to interact with robots. But, we were rather surprised that we could only find very limited options. For example, we evaluated the Sony AIBO messenger software. While promoted as a complete e-mail application, it is closer to an e-mail reader. It has no functionality to compose and send e-mails. Moreover, it has not been designed for people who are blind. Therefore, our system is the first system to

1. provide a mobile robotic interface for *Ambient Intelligence*,
2. enable mnemonic commands because of the embodiment in a 4-legged dog looking robot, and
3. allow rapid learning and use by blind adults.

There were several lessons learned. With the blind participants, any kind of unexpected movement by the robot is distressing. Thus, in our application, the robot remains essentially motionless when in use. This is a very distinct aspect from current trends in GUIs and the World Wide Web. Current GUIs allow or enable push content on Internet navigation. Applications allow enabling the sudden appearance of wizards while most applications now have update wizards. These unexpected appearances on the visual display are in general far more easily tolerated by users than sudden movement, gestures, actions or sounds by the robotic interface. While it seems reasonable to prompt suddenly with a wizard on a visual display in order for some software update, robots as interfaces need more delicate consideration of this issue. At least, use of sudden actions should be restricted to limited situations (the ones we foresee are alarms or regular warnings, like updating a virus signature file or battery in need of charge).

Two other lessons learned are as follows. First, gaining familiarity with the use of a robot as an interface can be a rather quick process. It must certainly be an incremental process that builds and introduces functionality on top of already familiar functionality. If this principle is followed, then the learning of the interface and the modules is fast. The incremental process can be regulated in terms of the complexity of the functionality as follows. The depth of menus or the size of the sets of options per menu item, or the number of context should all be small for the first applications and then increase as the user progresses to more complex contexts. Note that our design found very useful reuse modules and reproduce sequences in the *Audio Player* for subtasks in the *Voice Recorder*. The recording and playing of messages in the *Voice Recorder* is part of the *E-mailer*. Second, to preserve the context and the familiarity gained, it is also important that the robot starts in a default state the user is significantly familiar with (thus, the simplest is the first application, and is always the starting point).

We use anthropomorphic robots to design mnemonic commands. For example, pulling the head back is a common practice to halt a dog or other quadrupeds. Since the user will have the robot standing laterally with the head close to the right hand (to operate the head with this hand) and the tail on the left hand, a push of the tail away from the user constitutes a move to go forward to the next track, while pulling the tail back for the user brings the application one track back (as we indicated, a right-handed versus a left handed user can easily reverse the position of the robot and the mapping of commands).

A multimodal system allows confirmation of a command by simultaneous production in two channels. However, the system should be configurable to allow unimodal use. The literature suggests that 95% to 100% of users preferred to interact in a multimodal manner when given the choice to use either speech, a pen or both [9]. We do see speech as becoming the core in the system. However, speech recognition demands that the context be limited in options and still seems very useful to preserve the "push the back" action to withdraw a command misunderstood or incorrectly indicated. Also, touching commands are a strong resource when speech is not viable, as in noisy environments.

The questionnaires revealed that the majority found tasks easy to complete (65%). The hardest aspect was remembering how to accomplish a task. However, observation with the blind adults shows that in fact, after one very short session, the steps are easy to remember and they can even describe the process to others. All participants described their experience as very enjoyable and only one person did not find the Sony AIBO easy to manipulate. Unfortunately, in two occasions the tail fell off. This caused some concern to the users at the time. We also notice that a few participants commented on being "worried about breaking it". However, once they grasped the amount of force required to generate a command, they found the setting satisfactory. All participants felt that this software was very useful in assisting vision impaired people. Even 8 sighted people thought that it would assist them in day to day computing. Moreover, these sighting people indicated they would use it if available at home. Note that sighting people would not be using it at home blindfolded. All of the participants were regular users of computers. Only 2 people would prefer it over traditional interfaces, because they were of the opinion that performing tasks would be more time consuming that on a GUI interface with a desk computer.

Currently the robot does very little computation. It indeed behaves almost as a "dumb" interface. However, in the same way as computing power increased to transform "dumb terminals" into machines capable of run significant graphics layouts and client software, we expect the power on inexpensive domestic robots to increase. This power needs to be put to use for good, reliable and comfortable human-computer interfaces. Naturally, one would expect that the improvements would be along the lines of speech recognition and gesture recognition. In particular, we expect these tasks will migrate from the desktop PC across to the robot. It also remains to study how configurable the interface and its settings shall be. The availability of configurable menus allows far more flexibility and users may tailor menus and settings to their needs. However, rapidly the user can be in a position that there is too much inconsistency between sequences to achieve tasks and the tool is too personalised. It seems that, in the same way any user can access a computer in a public library, robotic interfaces may need to develop standard interfaces, so some of them can be deployed for generic, rather than personalised use.

In summary, robots as interfaces to Ambient Intelligence allow anthropomorphic mnemonics. Multi-modality reduces the situations where a command is misunderstood. But, the willingness to accept a robot moving by itself in unexpected ways is low.

References

1. Bartlett, B., Estivill-Castro, V., Seymon, S., Tourky, A.: Robots for pre-orientation and interaction of toddlers and preschoolers who are blind. In: Roberts, J., Wyeth, G., (eds.) Proceedings of the 2003 Australasian Conference on Robotics and Automation, ACRA. CSIRO's (QCAT), Brisbane, Australia (December 1-3, 2003)
2. Blenkhorn, P., Evans, D.: Using speech and touch to enable blind people to access schematic diagrams. Journal of Network and Computer Applications 21, 17–29 (1988)
3. Ghidary, S.S., Nakata, Y., Saito, H., Hattori, M., Takamori, T.: Multi-modal interaction of human and home robot in the context of room map generation. Autonomous Robots Journal 13(2), 169–184 (2002)
4. Graf, B., Hans, M., Schraft, R.D.: Care-O-bot II — development of a next generation robotic home assistant. Auton. Robots 16(2), 193–205 (2004)
5. Hoshino, A., Kato, K., Takeuchi, J.: A chat information service system using a humanoid robot. In: Hoshino, A., Kato, K., Takeuchi, J. (eds.) 14th IEEE International Workshop on Robot and Human Interactive Communication, IEEE Computer Society Press, Los Alamitos (2005)
6. Kulyukin, V., Gharpure, C.: Ergonomics-for-one in a robotic shopping cart for the blind. In: Proceedings of the 2006 ACM Conference on Human-Robot Interaction (HRI 2006), Salt Lake City, Utah, ACM Press (to appear)
7. Kulyukin, V., Gharpure, C., Nicholson, J., Pavithran, S.: RFID in robot-assisted indoor navigation for the visually impaired. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), Sendai, Japan (2004)
8. Leventhal, J., Uslan, M., Schreier, E.: A review of Braille printers. Journal of Visual Impairment and Blindness 85, 364–350 (1991)
9. Oviatt, S.: Ten myths of multimodal interaction. Communications of the ACM 42(11), 74–81 (1999)
10. Raisinghan, M.S., Benoit, A., Ding, J., Gomez, M., Gupta, K., Gusila, V., Power, D., Schmedding, O.: Ambient Intelligence: Changing forms of human-computer interaction and their social implications. Journal of Digital Information 5(4), 8–24 (2004)

11. Raskin, J.: The Humane Interface. New Directions for Designing Interactive Systems. Addison-Wesley Publishing Co, Reading, MA (2000)
12. Remagnino, P., Foresti, G.L., Ellis, T. (eds.): Ambient Intelligence. Springer, Berlin (2005)
13. Shneiderman, B.: Designing the User Interface, 3rd edn. Addison-Wesley Publishing Co, Reading, MA (1998)
14. Steinfeld, A.: Interface lessons for fully and semi-autonomous mobile robots (2004)
15. Sutcliffe, A.G.: Human-Computer Interface Design, 2nd edn. MacMillan, London (1998)
16. Takahashi, Y., Jones, J., Koyama, H., Komeda, T.: Development of the assistive mobile robot system for coexisting and cooperating with human beings. Advanced Robotics 18, 473–496 (2004)
17. Tanaka, F., Fortenberry, B., Aisaka, K., Movellan, J.R.: Developing dance interaction between QRIO and toddlers in a classroom environment: Plans for the first steps. In: 2005 IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005), Nashville USA, pp. 223–228. IEEE Computer Society Press, Los Alamitos (2005)
18. Vanderheiden, G.C.: Nonvisual alternative displays techniques for output from graphical-based computers. Journal of Visual Impairment and Blindness 83(8), 383–390 (1989)
19. Werry, I., Dautenhahn, K., Odgen, B., Harwin, W.: Can social interactionskills be taught by a social agent? the role of a robotic mediator in autism therapy. In: Beynon, M., Nehaniv, C.L., Dautenhahn, K. (eds.) CT 2001. LNCS (LNAI), vol. 2117, pp. 57–74. Springer, Heidelberg (2001)