IMPROVING THE REALITY PERCEPTION OF VISUALLY IMPAIRED THROUGH PERVASIVE COMPUTING

Vlad Coroamă,* Tarik Kapić, Felix Röthenbacher

Abstract

The visually impaired experience serious difficulties in leading an independent life, due to their reduced perception of the environment. However, we believe that ubiquitous computing can significantly improve the perception of the surrounding reality for the blind and visually impaired. In this paper we describe the Chatty Environment, a system that addresses this problem and has been developed after a series of interviews with potential users. The system, which reveals the surroundings to the user by speech output, is usable in both indoor and outdoor contexts.

1. Everyday Problems of the Visually Impaired

Most blind and visually impaired people confront serious difficulties when finding themselves in new, unknown environments. A cane or a guidance dog won't be enough to enable them to find their way through an unknown city, not even in a less complex environment, like an airport terminal or a university building. Many other problems encountered by the visually impaired don't seem obvious to sighted people. In a supermarket, for example, the blind person has great trouble finding the needed items, since all packed food feels similar. Without external help, he or she will only go to the known local supermarket and only buy a few items in learned locations. Another problem most sighted people are unaware of, is that the visually impaired will often not be able to catch a bus because of its brief stop at the station, which is too short to allow him/her to find the bus' door and the button to be pushed for opening it. Here again, blind people have to rely on external help.

Why do visually impaired confront such difficulties in leading an independent life? The cause probably lies in the nature of how humans use their senses to perceive the world – most people, when asked, will identify sight as the most important sense. This subjective impression is supported by anatomical facts. The brain region processing the visual input is with about 10 billion neurons more than five times larger than the brain regions handling any other sensorial input. Sight being the most important among the human senses, the modern world is tailored to meet this fact, which worsens the problem for the visually impaired. When constructing busses with buttons for opening the doors, it is likely that nobody thought of blind people and the trouble they will have finding those buttons.

Certainly, with the ever increasing miniaturization of electronic devices, the rapidly increasing understanding of human genetics and brain functioning, and the possible emergence of hybrid neuronal-

*Swiss Federal Institute of Technology (ETH) Zurich, 8092 Zurich, Switzerland, coroama@inf.ethz.ch

electronic circuits, blindness could be eradicated in a foreseeable but distant future. Miniature cameras, installed in the eyeballs, would then transmit their images directly to the brain. Until medicine ultimately reaches this goal, however, we believe that pervasive and ubiquitous computing technology can be used to help the visually impaired gain an increased quality of life and a higher degree of independence.

2. The Chatty Environment

Bearing the difficulties encountered by blind people in mind, we proposed the paradigm of a *chatty environment* [7, 8], a system that enhances the visual information by other means of sensorial input that can be experienced by the visually impaired, i.e. spoken information. While moving through the chatty environment, this spoken information is continuously presented to the visually impaired user. Thus, he finds out how his surroundings are shaped and which entities exist around him, e.g., where the incoming bus goes to and where its nearest door is located, which package of food he is holding in his hand in the supermarket, or where the next fast-food-restaurant is located. The visually impaired is also able to get more in-depth information on selected parts of the environment and may even perform actions on some of these entities.

Figure 1. The virtual aura of tagged real-world objects.

The chatty environment is realized by using pervasive computing technologies to enhance the environment's real-world objects with a virtual component, that holds information about the corresponding object. In the real-world, each object possesses a beacon, which creates a *virtual aura* around the object (see figure 1). When the user moves into the aura of such an enhanced real-world entity (or when the entity moves towards the person, as in the case of a bus), the device carried by the user – the *world explorer* – tells her about the object's existence and offers her a standardized interface for interacting with it. This feature – the environment endlessly speaking to the user, telling her about the surroundings – might seem annoying to most sighted people. However, during our interviews with blind and visually impaired people, we learned that there can almost never be too much spoken input for the visually impaired.

To better understand the specific needs of visually impaired people, we conducted a series of interviews with blind and visually impaired persons. The next section summarizes the most relevant results of these interviews.

3. Visually Impaired User Study

Blind and visually impaired people have a different perception of the world than sighted people. Relevant differences also exist between the perception of completely blind and of those with a certain level of visual impairment. Any system designed for the visually impaired has to be aware of these differences in order to provide a user interface adapted to the limitations and special needs of its users.

To this extent, a series of nine interviews (5 women, 4 men) with blind and visually impaired allowed us to get valuable information for the system design. The medium age of the questioned people was 54 years, in the range from 30 to 81 years. They live in different regions of Switzerland and their educational level varies from high-school level to university level. The impairments of the interviewed range from total blindness to 40% of sight. Interviews were conducted in two steps: All interviewed first answered a questionnaire comprising 20 questions, ranging from general information about their age, profession or impairment grade to precise questions about use of handheld devices, preferred in- and output methods and particular requirements for object descriptions. The interview was also based on the "open-end" principle, each participant being able to add any information or suggestion considered to be relevant. The interviews were about one hour long.

After conducting these interviews, we derived a list of requirements for an assistance pervasive computing system aimed for the blind and visually impaired. According to the survey, valuable would be a system that:

1. increases the user's perception of the surroundings by telling her which entities she is passing by. This seems to be the most important user requirement – to have an extension of their own world perception by having the environmental entities in their immediate neighborhood announced to them,

2. helps in an environment with many small items (e.g., products in the supermarket) as well, by answering questions like: "which item am I holding in my hand?" or "where is the huckleberry jelly?",

3. does not require the user to pinpoint to a certain location to get the desired information (this being especially difficult for completely blind people),

4. announces points of interest located further away,

5. helps them navigate to these points of interest, outdoor as well as indoor (especially relevant for complex buildings like airport terminals or large office buildings),

6. lets them filter objects (like fast-food-restaurants or restrooms) according to a classification and then presents a list of such entities situated in the neighborhood, so that the user may subsequently choose to be guided to either the nearest instance or another one from the list,

7. enables communities to emerge, by allowing the user to leave marks or remainders for herself and/or other users, (e.g., a message on a traffic light "large crossroad ahead, must be passed quickly").

There was more valuable data gathered from the interviews. We learned that most interviewed would not be disturbed by objects that “speak”. Basically, they imagine that these objects would help them find their way more easily, even without an explicit guidance aid. Speech is the preferred output medium for almost all interviewed people. Some of them could find utility in additional signalling techniques, such as vibration or non-speech audio signals (i.e., beep). All rejected the idea of force feedback on the blind cane, as the normal use of the cane would be altered too much through this technique. All participants would like to carry the device in the pocket (handbag for women) or hanged around the neck, in order to keep the hands free for other functions. Hence, the acoustic output needs to be transmitted to the user via a headset-device. Nevertheless, most important about speech output (and audio output in general) is the fact that the normal hearing of the user must not be altered by this system. Blind need stereometric hearing in order to determine for example the direction of moving obstacles. Therefore, any kind of headphones or earphones used has to comply with this requirement. This excludes stereo headphones; mono headphones are suitable if they let environmental sounds muddle through.

4. The System

Over the past months, we have developed a first prototypical implementation of the chatty environment. In this section, we give an overview of its components.[1] The main system components are:

Tagged Entities All objects in the chatty environment are tagged with electronic beacons. These beacons generate the real-world object’s auras, thus making them detectable.

In this first prototype, we use the Berkeley Motes [1] as beacon devices. They offer the advantage of adjustable emitting power. Hence, the tagged objects aura can vary between a few centimeters and almost 100 meters. This property of adjustable range enables us to create large auras around large or important objects, and smaller auras around less important or small objects, recreating the way sighted people would gather information about those objects. Aside from Berkeley Motes, other passive or active tagging method could be used, like active or passive RFID tags, or BTnodes [5]. Using other beacons is a relatively easy task, the system being component-based.

World Explorer The World Explorer device (see figure 2), carried by the user, is both beacon reader and mobile computing platform. It constantly sends identification requests. If a beacon receives this request, it sends a reply with a unique identification. Upon detection of a beacon, the system creates a beacon object and sets it to an active state. Beacons have to store a unique identification and a small human-understandable textual description.

As World Explorer device, we use an HP iPaq 5450, with integrated WLAN 802.11 and Bluetooth connectivity. At its serial interface, a mote is connected, which sends out the inquiries for motes in the environment and transmits the answers to the software running on the mobile platform. Aside from ID and textual description, further information may be stored on the beacon, but also on its virtual counterpart (see below).

Virtual Counterparts Generally, beacons are small devices with limited resources. The amount of information for the user may be considerable so that only a small part can be stored on the beacon device itself. Further information has to be collected from remote data sources. A

[1] A detailed description of the system can be found in [9].

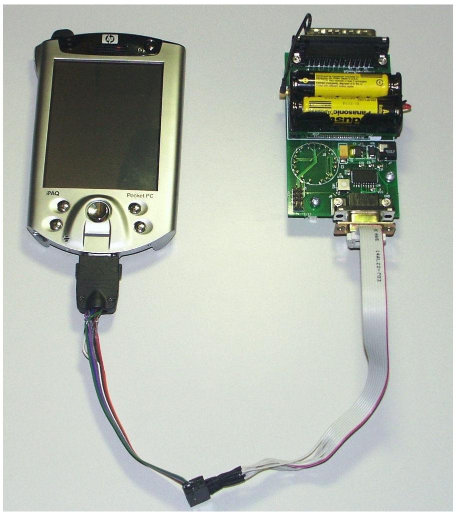

Figure 2. View of the World Explorer device.

key piece of information stored on the beacon device is the identification number, which is transformed into an URL where the other information is stored – the *virtual counterpart*.

Communication All communication between the World Explorer and the data sources is text-based (i.e., XML files). A lightweight text-to-speech-engine[2] on the user's mobile device generates the spoken output needed by the user.

Since text files are much smaller than audio-files containing the same information, two advantages are achieved by this approach. First, more information can be stored on the beacon itself. With devices like the Berkeley motes, having a rather large memory, it is even possible to store all the needed information on the beacon. Saving the effort to access virtual counterparts improves the flexibility and robustness of the system. Second, the data exchange is done more rapidly, thus giving the user an up-to-date view of the environment.

The information from the beacons is gathered via the communication protocol of the Berkeley motes. With the virtual counterparts, the mobile device communicates via WLAN, using HTTP as communication protocol.

User Interface The user interface is designed to allow the user to operate the device by reacting to speech output. A five-way cursor serves as input device. The keys are conceived to resemble the possible interactions with a web browser, namely, *select, back, forward, up,* and *cancel*. Another key called *home* is used to get an overview of the current system state.

The communication with the user is based on natural language. Each menu item consists either of an introductory text (e.g., "Please choose:"), followed by an enumeration of the subtopics, or of an information text (e.g., "The traffic light is red."). A menu item may be associated with an action; the action is executed when the user selects this item. The World Explorer supports two message queues with different priority levels to inform the user. The high-priority queue is used to announce warnings against obstacles, red traffic lights, etc. The normal priority queue is used to present ordinary information to the user. If a high-priority message arrives, the currently processed message is interrupted and the high-priority message is read.

[2] Elan Tempo™ PocketSpeech. A multilingual text-to-speech engine for PDA. http:\\www.elanspeech.com.

5. Discussion

There is basically two ways in which a ubiquitous computing system may recognize entities in the environment. First, as in our approach, by tagging the objects. Second, through localization. If the user's location is known, it is possible to derive the known objects in her surroundings.

By using GPS location correlated with a central database, the task becomes relatively straightforward for outdoor environments. However, to deploy the system in both indoor and outdoor contexts, both kinds of location systems need to be implemented on the user mobile device. Another problem is mobility. Since somewhere a database has to store the location of objects, changes of an object location are hard to reflect in the virtual world and the mechanism for doing so is inherently error-prone. That applies also for any other change in the object's state, like the object's temperature. You could think of the mobile objects having a location system themselves and a wireless communication device to continuously report their state changes, but that would obviously be more of a headache than just tagging the objects.

On the other hand, when using the tagging approach, the main issue is the reliability of the tagging devices. They can run out of autonomous power or fail completely. And, of course, the objects have to be tagged in the first place, which seems a huge overhead. However, we believe that these two drawbacks will significantly decrease in the near future, since more and more real-world entities will be tagged anyway – be it through cheap RFID tags (not depending on energy supply), or through more complex beacons, like Berkeley motes or BTnodes. By using cheap tagging devices, there can also be a built-in redundancy – an object having multiple tags identifying it – so that failures could be tolerated to a certain extent. Therefore, extending the system to the use of RFID tags as beacon devices, is a high priority for us.

Both paradigms have their strengths and drawbacks. However, for the specific application domain of supporting the visually impaired, the tagging approach seems to be more adequate, mobility being an especially relevant issue in this domain.

5.1. Related Work

An early system to implement similar functionality is Talking Signs [4], which started back in 1993. Infrared beacons mark the environment's objects, the user's mobile device has an infrared receiver and a small computing platform. The infrared technique used, however, has several drawbacks. The user has to actively scan the environment, pointing the device to all possible directions until an answer comes, which may seem annoying to blind people. While scanning the environment, one of the user's hands is occupied. Most important, however, since only short audio messages are transmitted, the user can not navigate through the object to gain more information about it.

The recently introduced system "Navigational Assistance for the Visually Impaired" (NAVI) [3] also uses a similar approach. The user's portable device combines a CD player with a mobile RFID tag reader. The tags mark objects in the environment and trigger the corresponding track on the CD. The approach is similar – the device explains the surroundings to the user by reading messages such as "front of Rush Rhees library". NAVI's weaknesses seem to be threefold. First, the system does not scale well. The user has to know a priori where she is heading and insert the correct CD (which she must have obtained in advance). When the number of tagged entities in a given environment varies, the CD must be updated, too, and supplied to all users. Second, using passive RFID tags only constraints the perception of all tagged objects to a radius of about one meter (depending on the used tags and reader). It is often advantageous to be able to define virtual auras with different extensions for different class of objects (a railway station should have a larger aura than a package of baked beans!). Third, NAVI's approach does not allow objects to change their state, since the information is stored statically on the CD.

5.2. Meeting User Requirements and Future Work

The main user requirement according to our survey – having a device that increases their perception of the surroundings – has been the main project focus so far. Our system allows this perception extension in a way suitable for both large items and small supermarket-like items as well as for both *indoor* and *outdoor* environments, by tagging the environment's objects with Berkeley motes, with variable emitting power. However, tagging cheap supermarket items with expensive motes will work for a research project only. We are currently in the process of extending the chatty environment with the Hitachi μ-Chip [2] as beacon device. Ultimately, both systems should run on the world explorer device, accounting for the foreseeable spreading of RFID technology and allowing both cheap short-range tagging and the more expensive large-range tagging, that also provides other sensorial input.

Chatty environment's communication is radio-based and does not need line-of-sight or the user to pinpoint at locations she is not able to see. Thereby, the mobile device could be left in the user's pocket, backpack, or handbag. Nevertheless, since the buttons allowing the user to navigate through the surrounding entities are currently located on the world explorer, the user cannot let the mobile device disappear in his pockets yet. We intend to overcome this drawback by moving the selection buttons on the user's cane and let them communicate with the world explorer via Bluetooth.

Other future research issues include: Letting the user choose objects according to a classification (e.g., "ticket counter" or "fast-food") and guiding him to these remote objects of interest. For the corresponding user input, a Braille-PDA should be suitable, but we also want to explore voice recognition. The needed indoor navigation will be based on a previously developed probabilistic indoor positioning system [6]. One last issue would be to let the user to drop spoken notes in the environment for later collection by herself or others, thus allowing communities to emerge.

While many problems remain to be solved, we are confident that the system we propose is a first step towards an assistance for the visually impaired by means of pervasive computing.

6. Acknowledgements

This work has been supported by the Gottlieb Daimler- and Karl Benz-foundation, Ladenburg, Germany, as part of the interdisciplinary research project "Living in a Smart Environment – Implications of Ubiquitous Computing".

References

[1] Berkeley Motes. http://webs.cs.berkeley.edu/tos/.

[2] Mu Solutions. Hitachi μ-Chip – The World's smallest RFID IC. http://www.hitachi.co.jp/Prod/mu-chip/.

[3] Navigational Assistance for the Visually Impaired. http://www.rochester.edu/pr/Currents/V31/V31N15/story07.html.

[4] Talking Signs Project. http://www.talkingsigns.com.

[5] The BTnodes Project. http://btnode.ethz.ch/.

[6] BOHN, J., AND VOGT, H. Robust Probabilistic Positioning based on High-Level Sensor-Fusion and Map Knowledge. Tech. Rep. 421, Institute for Pervasive Computing, Distributed Systems Group, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, Apr. 2003.

[7] COROAMĂ, V. The Chatty Environment – A World Explorer for the Visually Impaired. In *Adjunct Proceedings of UbiComp 2003* (Oct. 2003), J. McCarthy and J. Scott, Eds.

[8] COROAMĂ, V., AND RÖTHENBACHER, F. The Chatty Environment – Providing Everyday Independence to the Visually Impaired. *UbiHealth Workshop, Seattle* (Oct. 2003).

[9] RÖTHENBACHER, F. World Explorer for Visually Impaired People. Master's thesis, Institute for Pervasive Computing, Distributed Systems Group, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, Oct. 2003.