Noname manuscript No.
(will be inserted by the editor)

Chaitanya Gharpure
Vladimir Kulyukin

Robot-Assisted Shopping for the Blind: Issues in Spatial Cognition and Product Selection

Received: date / Accepted: date

Abstract Research on spatial cognition and blind navigation suggests that a device aimed at helping blind people to shop independently should provide the shopper with effective interfaces to the locomotor and haptic spaces of the supermarket. In this article, we argue that robots can act as effective interfaces to haptic and locomotor spaces in modern supermarkets. We also present the design and evaluation of three product selection modalities - browsing, typing and speech, which allow the blind shopper to select the desired product from a repository of thousands of products.

Keywords Assistive Robotics · Service Robotics · Human-Robot Interaction · Blind Navigation · Spatial Cognition · Haptic and Locomotor Interfaces · Independent Shopping for the Visually Impaired

1 Introduction

We present several results from our ongoing research on independent shopping for the visually impaired. This particular thread of research evolved from a more generic area of robot-assisted wayfinding for the visually impaired (Kulyukin et al, 2006, 2005; Gharpure, 2004). Our motivation is fourfold. First, several visually impaired participants in our previous wayfinding experiments expressed the need for a device that would help them to do grocery shopping independently. Second, grocery shopping is an activity that presents a barrier to independence for many visually impaired people who either do not go grocery shopping at all or depend on sighted guides, e.g., store staffers, spouses, and friends (Kulyukin et al, 2005). Third, shopping

Computer Science Assistive Technology Laboratory (CSATL)
Department of Computer Science
Utah State University
Logan, UT 84322-4205
cpg@cc.usu.edu, vladimir.kulyukin@usu.edu

complexes are the most functionally difficult environments for the visually impaired (R. Passini and G. Proulx, 1988). Consequently, accessibility breakthroughs in shopping complexes will likely carry over to other environments. Fourth, to the best of our knowledge, there is no existing system that provides a complete integrated solution to this problem.

1.1 Designing Devices for Assisted Shopping

As a task, grocery shopping can be decomposed into three main subtasks: product selection, navigation and product retrieval. Our research plan was to focus on the last two tasks and then work on the first one. Research on spatial cognition and navigation of the visually impaired distinguishes two spatial categories: locomotor and haptic (Golledge et al, 1998; Millar, 1995, 1997, 1982). The haptic space is defined as the immediate space around the individual that can be sensed by touch or limb motion without any bodily translation. The locomotor space is defined as a space whose exploration requires locomotion.

Vision is the primary sensory modality that enables humans to align the egocentric and allocentric frames of reference, which is the key to reliable navigation. In the absence of vision, the frames align best in the haptic space. In the locomotor space, as the haptic space translates with the body, lack of vision causes the frames to misalign, which negatively affects action reliability. Giving the visually impaired equal access to environments that the sighted take for granted entails designing interfaces to the haptic and locomotor spaces in those environments that either eliminate the necessity of alignment or enable the visually impaired to align the frames when necessary (Kulyukin, V., Gharpure, C., and Pentico, C., 2007).

A visually impaired shopper (the shopper henceforth) can accomplish the second subtask - navigation - when the shopper has 1) a means to accurately access the pose (location and orientation) and to obtain adequate topological knowledge or 2) a means to reliably maneuver the haptic space in the locomotor space. The first choice ensures accurate frame alignment and leaves to the shopper the maneuvering of the haptic space in the locomotor space. The second choice guarantees reliable maneuvering but takes no position on frame alignment.

Either solution requires an effective interface to the locomotor space. The shopper can accomplish the last subtask - product retrieval - if the shopper has a means to maneuver the haptic space in the vicinity of a target product until the product is within the haptic space. To guarantee independence, any assistive shopping device for the visually impaired must necessarily address both subtasks and, consequently, provide the shopper with effective interfaces to the haptic and locomotor spaces in supermarkets.

Several research projects are related to our research. In (Brent and Modi, 2000), the authors have developed a shopping aid consisting of a barcode scanner and a laptop secured to a shopping cart. After the user scans the product's barcode, the laptop accesses the corresponding data files and displays them on the laptop in an enlarged font. Although the device helps the visually impaired to make an informed product choice, it does not help them shop independently. This is because it does not provide the shopper with an interface to the locomotor space to navigate within the store. It does not appear to be possible for white cane users and guide dog handlers to navigate safely with a large shopping cart in the front. In addition, the device provides only the product specific information. It does not help the shopper find the product on the shelf.

In (Lalatendu et al, 2006), the authors propose a wearable device consisting of an earpiece, display visor, bluetooth connectivity, and RFID scanner, all mounted on an eyewear. However, the proposed device is limited to product identification only. Another system, Trinetra (Lanigan et al, 2006), has its functionality limited to product identification using a text-to-speech enabled cellphone and a pen-like barcode reader. Such systems are not, in and of themselves, sufficient for independent shopping for the visually impaired. However, they can be incorporated as a part of the interface to the haptic space in systems like RoboCart.

To accomplish the subtask of product selection, the shopper should be able to - speedily, accurately and comfortably - convey her intent to the shopping device. An example of the shopper intent would be: “Take me to <product>”, where <product> can be any product from a repository of thousands of products. Thus, the interface to the product repository (catalogue) should be designed such that 1) the shopper can rapidly and comfortably select the desired product and 2) the shopper should be able to maintain an adequate level of discretion and comfort in a supermarket.

1.2 Robot-Assisted Shopping

Can robots function as effective interfaces to the haptic and locomotor spaces in the supermarket? We believe that this question can be answered in the affirmative. Several reasons justify our belief. Traditional navigation aids, such as guide dogs and white canes, can act as interfaces to the haptic space in the environment by enhancing the blind individuals perception around the body. However, neither guide dogs nor white canes can effectively interface to locomotor spaces, because they cannot help their users with macro-navigation, which requires functional topological knowledge of the environment. It is true that sighted guides ensure the reliable maneuvering of the haptic space, but only at the expense of independence. Loss of independence translates into loss of privacy. Robot-assisted shopping experiments as reported in (Kulyukin, V. and Gharpure, C., 2006; Kulyukin et al, 2005), indicate that the visually impaired shoppers are usually not willing to use store staffers when shopping for personal hygiene items, medicine, and other products that require discretion.

Our central research hypothesis is that, in order to function as an effective interface to the haptic and locomotor spaces in the supermarket, the robot must satisfy a two-fold objective: in the locomotor space, the robot must eliminate the necessity of frame alignment and, in or near the haptic space, the robot must cue the shopper to the salient features of the environment sufficient for product retrieval.

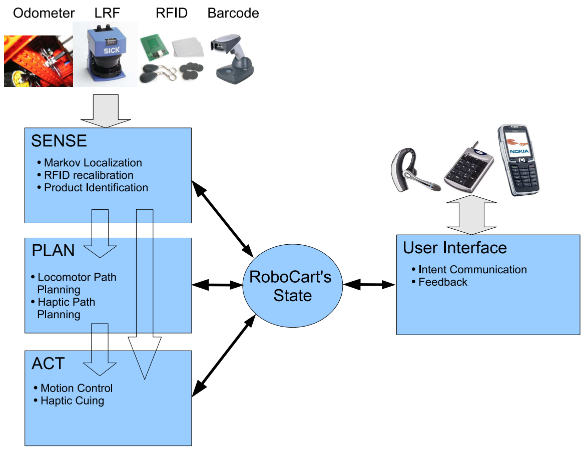


Fig. 1 RoboCart's Component Connectivity

1.2.1 RoboCart: A Robotic Shopping Assistant

The overall design of RoboCart is based on the principles of ergonomics-for-one (Kulyukin, V. and Gharpure, C., 2006) and reflects the dual interface functionality through two modules: locomotor and haptic. The locomotor module consists of a Pioneer 2DX mobile robotic base from ActivMedia, Inc. upon which a wayfinding toolkit is fitted in a polyvinyl chloride (PVC) pipe structure. A shopping basket is mounted upon the PVC structure, as shown in Figure 2. The current Robocart setup is slightly unstable due to a small robotic base and possibly heavy top, when the basket is full with products. However we feel that, as a proof-of-concept prototype, the setup is sufficient to conduct feasibility studies. Using a numeric keypad, the shopper can select a product through an appropriate product selection interface. Once the shopper confirms the selection, RoboCart guides the shopper to the vicinity of the product.

The haptic module consists of a wireless omni-directional barcode reader shown in Figure 2. The reader is ergonomically modified with a plastic structure that helps the blind shopper align the barcode reader with the shelf. After RoboCart brings the shopper in the vicinity of the product, RoboCart uses the shopper's egocentric frame of reference to instruct the shopper through synthetic speech on how to find the product, e.g. Honey Nut Cheerios is on the top shelf to your right. The shopper finds the shelf and uses the barcode to scan the barcodes on that shelf. The product name of each scanned barcode is read to the shopper.

RoboCart software architecture is the popular three-tier architecture consisting of the sense, plan, and act modules. RoboCart's shared state is updated and checked by these modules and the user interface module.The connectivity of different components in RoboCart's architecture is shown in figure 1.

Fig. 2 a. RoboCart hardware, b. Shopper scanning a barcode, c. Modified barcode reader.

The remainder of our article is organized as follows. In Section 2, we talk about spaces in a supermarket. In Section 3, we present the design and implementation of RoboCart. In Section 4, we discuss the product selection problem and present our solution to it. In Section 5, we present the results of longitudinal robot-assisted shopping experiments with ten visually impaired participants in a real supermarket. In Section 5, we present the results of the experiments with the three different product selection interfaces: browsing, typing, and speech. In Section 6, we discuss our findings from field trials and user feedback. In Section 7, we present and discuss several user comments. In section 8, we present our conclusions.

2 Spaces in the Supermarket

2.1 Spaces and Interfaces

Spatial ontologies come about when we attempt to categorize space according to the ways we interact with it (Tversky, B., Morrison, J., Franklin, N., and Bryant, D., 1999). Freundschuh and Egenhofer (Freundschuh and Egenhofer., 1997) give a comprehensive review of previous work on categorization of space and distinguish six categories based on manipulability, locomotion, and size and use their ontology to describe previous ontologies of space in the geography literature. We contribute to this line of research a trichotomous ontology of spaces in a supermarket. This trichotomy is an extension of the dichotomous ontology (haptic vs. locomotor) currently used by many researchers on spatial cognition of the visually impaired. Our trichotomy is certainly incomplete. We developed it solely for the purpose of describing and analyzing the interactions between visually impaired shoppers and RoboCart. We believe that the trichotomous ontology can be used to evaluate any assisted shopping device for the visually impaired.

RoboCart assists the shopper in two stages. It first guides the shopper into the vicinity of the desired product and then instructs the shopper on how to maneuver the haptic space within that vicinity. In the first stage, RoboCart interfaces the shopper to the locomotor space, guiding the shopper to the required aisle section and interacting with other shoppers by asking them to yield the way when a passage is blocked. In the second stage, RoboCart cues the shopper to some salient features of the environment near the haptic interface through voice instructions grounded in the shopper's egocentric frame of reference.

To describe how the visually impaired shopper interacts with the supermarket space using RoboCart, we introduce the category of *target space*. The target space is the shopper-centric subspace of the locomotor space in which the shopper perceives the target product to be. The target space is always defined with respect to a specific target product. Haptic cues in the target space act as external reference points during the shopper's maneuvering of the haptic space in the target space until the haptic space contains the product.

2.2 Inside the Target Space

RoboCart eliminates the necessity of frame alignment by restricting the target space to the searchable vicinity of the product. In the target space, the shopper is left with the task of retrieving the product from the shelf and placing it into RoboCart's basket. It seems reasonable to conjecture that the shopper's performance in the target space depends on the shopper's knowledge of the target space, the shopper's sensory, cognitive, and physical abilities, and the complexity of the target space.

The knowledge of the target space includes knowledge of distances, product shapes, neighboring products, the spatial arrangement of shelves, and other haptic cues that may be unique to the shopper. If the robot is consistent overtime in how it sets up the target space with respect to a given product and verbally orients the shopper in the target space, then, as the shopper gets more exposure to the target space, the shopper's performance should improve.

The target space instructions given by the robot to the shopper need not be exhaustive. It is unreasonable to expect the robot designer to equip the robot with the knowledge of every possible haptic cue that may assist the shopper with retrieving the product. Instead, our assumption is that with time the shopper learns many haptic cues, such as level differences among shelf levels, bad barcodes, and protruding labels, that may be valid only for that shopper. Thus, the robot ensures the consistency of the target space and orients the shopper inside that target space. These haptic cues are cognitive indices into the target space used by the shopper in product retrieval. The robot also emits audio beacons to help the shopper with placing the found product into the basket. In the case of RoboCart, the audio beacon is the sonar clicking sound. RoboCart does not use sonars but turns them to further orient the shopper in the target space.

The shopper's performance in the target space may also be conditional on the shopper's sensory and physical abilities. The literature on the haptic space performance of the visually impaired (Kay, 1974; Lahav and Mioduser, 2003; Ungar, 2000) suggests that the degree of vision may affect the efficiency of product retrieval. In addition, the shopper's abilities to bend and scan products on bottom shelves, to hold products, and to find barcodes on the shelves are likely to contribute to the shopper's performance. The shopper's performance may also depend on the intrinsic complexity of the target space. The complexity of the target space can be characterized by the number of products, the product density, the homogeneity of product types, the number of shelves, package sizes, product layouts, presence of other shoppers, etc.

3 Design of RoboCart

3.1 Locomotor Module

Autonomous indoor navigation was necessary to realize the interface to the locomotor space. Since we had already developed an orientation free RFID based navigation algorithm, we decided we use that in the supermarket. However, our approach failed in unstructured open spaces. To deal with unstructured open spaces, we decided to use laser-based Monte Carlo Markov localization (MCL) (Fox, 1998), as it was already implemented in ActivMedias Laser Mapping and Navigation software.

Navigation in RoboCart is a partial realization of Kupiers Spatial Semantic Hierarchy (SSH) (Kupiers, 2000). The SSH is a model to represent spatial knowledge. According to SSH, spatial knowledge can be represented in five levels: sensory, control, causal, topological and metric. Sensory level is the interface to the robot's sensory system. RoboCarts primary sensors are a laser range finder, a camera, and an RFID reader. The control level represents the environment in terms of control laws which have trigger and termination conditions associated with them. The causal level describes the environment in terms of views and actions. Views specify triggers; actions specify control laws. For example, follow-hall can be a control law triggered by start-of-hall and terminated by end-of-hall. The topological level of the SSH is a higher level of abstraction, consisting of places, paths and regions, and their connectivity, order and containment relationships. The metrical level describes a global metric map of the environment within a single frame of reference.

Several problems with MCL localization were discovered (Gharpure et al, 2006). First, the robot's ability to accurately localize rapidly deteriorated in the presence of heavy shopper traffic. Second, MCL sometimes failed due to wheel slippage on a wet floor or due to the blind shopper inadvertently pulling on the handle. Third, since MCL relies exclusively on odometery to localize itself along a long uniform hallway that lacks unique laser range signatures, it would frequently get lost in an aisle. Fourth, MCL localization frequently failed in the store lobby, because the lobby constantly changed its layout due to promotion displays, flower stands, product boxes. Finally, once MCL fails, it either never recovers, or recovers after a long drift.

To allow for periodic and reliable MCL recalibration, we decided to turn the floor of the store into an RFID-enabled surface, where each RFID tag had its 2D coordinates. We use the low frequency RFID tags with 134.2

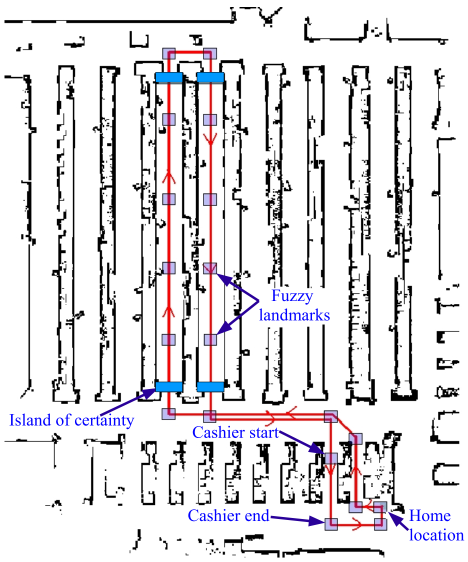


Fig. 3 Fuzzy areas in the grocery store environment.

khz frequency. Some ubiquitous computing researchers had started thinking along the same lines (Scooter and Helal, 2004). The concept of the RFID-enabled surface was refined into the concept of recalibration areas, i.e., areas of the floor with embedded RFID tags. In our current implementation, recalibration areas are RFID mats which are small carpets with embedded RFID tags. The mats are placed at specified locations in the store without causing any disruption to the indigenous business processes. One advantage of recalibration areas is deterministic localization: when the robot reaches a recalibration area, its location is known with certainty. We built several RFID mats with RFID tags embedded in a hexagonal fashion. Every recalibration area is mapped to a corresponding rectangular region in the stores metric global map constructed using ActivMedias metric map building software, Mapper3 ©.

Several research efforts in mobile robotics are similar to our approach. Kantor and Singh (Kantor and Singh, 2002) use RFID tags for robot localization and mapping. They utilize time-of-arrival signals from known RFID tags to estimate distance from detected tags and localize the robot. Hahnel et. al. (Hahnel et al, 2003) propose a probabilistic measurement model for using RFID signals to improve the performance of laser based localization of mobile robots in office environments.

3.2 Haptic Module

RoboCart's haptic module consists of a wireless omnidirection barcode reader using which the shopper can scan barcodes on the edges of the shelves. The barcode reader was ergonomically modified with plastic structures that help the shopper to align the barcode reader with the shelf (Figure 2 c.). The plastic strips attached to both sides of the reader are placed on small protrusions from metallic casings that contain barcodes. After placing the plastic strips on these protrusions, the shopper slides the reader along the shelf and scans barcodes. Whenever a barcode is scanned, the shopper can hear the product name through a pair of wireless bluetooth headphones. When RoboCart brings the shopper in the vicinity of the product, the shopper is also given instructions about finding the product in the target space. For example, "The product on the right hand side, on the third shelf from the bottom. Search to the left." The shopper can then feel the third shelf from the bottom on the right hand side and start scanning the barcodes toward the left until the target product is found.

Our approach has two drawbacks: First, the shopper has to scan all barcodes until the target barcode is found. Second, the shopper can completely miss the target barcode and never find the product. To remedy this situation, we plan to maintain a spatial connectivity of barcodes in the future. Using this, the shopper can be instructed to skip a certain number of barcodes, or alerted if she misses to scan a barcode. However, we feel that our current design is sufficient to evaluate our claims and test the hypotheses we put forth later in the paper.

It seems reasonable to conjecture that the shopper's performance in the target space depends on the shopper's knowledge of the target space, the shopper's sensory, cognitive, and physical abilities, and the target space complexity. The knowledge of the target space includes knowledge of distances, product shapes, neighboring products, the spatial arrangement of shelves, and other haptic cues that may be unique to the shopper. If the robot is consistent over time in how it sets up the target space with respect to a given product and verbally orients the shopper in the target space, then, as the shopper gets more exposure to the target space, the shopper's performance should improve.

It should be noted that the egocentric target space instructions are not exhaustive. It is unreasonable to expect the robot designer to equip the robot with the knowledge of every possible haptic cue that may assist the shopper with product retrieval. Instead, our assumption is that with time the shopper learns many haptic cues, such as height differences among shelf section levels, bad barcodes, and protruding labels, that may be uniquely valid only for that shopper. Once learned, these cues function as cognitive indices into the target space used by the shopper during product retrieval.

The robot ensures the consistency of the target space and of the verbal egocentric orientation of the shopper in that space. To help the shopper with placing the retrieved product into the basket, the robot emits a continuous audio beacon - a low-pitch clicking sound emitted by its sonars.

4 Product Selection

In this section, we address the problem of product selection during unplanned shopping. This problem can either be posed as a problem of searching a specific product rapidly, or a problem of browsing the product hierarchy to select a desired product. The problem of accessing electronic information by blind users is not new. The Web Accessibility Initiative has published Web Content Accessibility Guidelines (WCAG) (W3C, 2000), which provides guidelines for developing accessible websites. However, since these guidelines are geared toward websites, there are several implicit assumptions made. First, task at hand (browsing a website) is not time critical. Second, the user is sitting in the comfort of her home or office. Third, the user will have a regular keyboard at her disposal.

Several ideas for implementing auditory interfaces for navigating menus and object hierarchies have been proposed (Walker et al, 2006; Raman, 1997; Smith et al, 2004). In (Brewster, 1998), the author investigates the possibility of using non-speech audio messages called *earcons* to navigate a menu hierarchy. In (Walker et al, 2006), the authors propose a new auditory representation called *spearcons*. Spearcons are created by speeding up a phrase until it is not recognized as speech. Both approaches are however suitable for navigating menus, and not large object hierarchies. Another approach for browsing object hierarchies (Raman, 1997) uses conversational gestures like *open-object*, *parent*, which are associated with specific navigation actions. Work done on auditory interaction objects (Klante, 2004; Gaver, 1989), outlines the requirements for an auditory interaction object that supports navigation of hierarchies.

In our implementation, we follow the functional requirements for *accessible tree navigation tool* given in (Smith et al, 2004). In that, the participants are required to find six objects from a large object hierarchy. However, the evaluation is done only to check for successful completion of the task, and not the speed of completion.

In (Divi et al, 2004), the authors present a spoken user interface in which the task of invoking responses from the system is treated one of retrieval from the set of all possible responses. In our case, the responses are the product names. They use the SpokenQuery system (Wolf et al, 2004) which is effective for searching spoken queries in large databases, is robust to environment noise and is effective as an user interface.

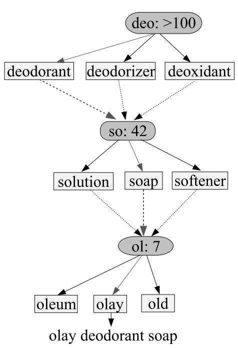
In (Sidner and Forlines, 2002), the authors propose the use of subset languages for interacting with collaborative agents. The advantage of using subset language is that it can easily be characterized in a grammar for a speech recognition system. In our case, the subset language is nothing more that just the names of the products. To make the problem easier, we developed a grammar where each rule consisted of a single word. The user is thus limited to speaking one word at a time. This greatly simplifies the grammar and reduces the misrecognitions. In addition, it obviates the need for the user having to learn a complicated subset language.

We implemented and tested three non-visual interfaces for product selection. The first one is a browsing interface, where, given the product categories, the shopper browses through the complete hierarchy to find the desired product. For the typing and speech interfaces, we employ an information retrieval based approach. In the typing based interface, the shopper is required to type the search string using a numeric keypad. In the speech based interface, speech recognition is used to form the search string.

4.1 Searching the Product Repository

Each entry in the product repository is represented by an N-dimensional vector where N is the total number of unique keywords in the repository. Thus, each vector is an N-bit vector with a bit set if the corresponding keyword exists in the product string. The search vector obtained from the search string is also an N-bit vector. The result of the search is simply all entries i, such that $P_i \& S = S$, where P_i is the N-bit vector of the i^{th} product, S is the N-bit search vector, and '&' is the *bit-wise and* operation.

This is a simple approach with two obvious problems: 1) the users will need to type complete words, which is tedious using just a numeric keypad; and 2) the search will fail if a word is spelled incorrectly. To solve the first problem, we use word prediction where the whole word is predicted by looking at the partial word entered by the user. However, instead of having the user make a choice from a list of predicted words, or waiting for the user to type the whole word, we search the repository for all prediction options. To solve the second problem, we do not use the spell checker, but instead provide the user with continuous feedback. Every time the user types a character, the number of results returned is informed to the user. The user obviously chooses not to type a character that returns zero results. At any point, the user can choose not to type the remaining word and continue to type the next word.



Possible search strings
for the Prediction tree
deodorant solution **(0)**
deodorant soap oleum **(0)**
deodorant soap olay **(2)**
deodorant soap old **(2)**
deodorant softener **(0)**
deodorizer solution **(0)**
deodorizer soap oleum **(0)**
deodorizer soap olay **(1)**
deodorizer soap old **(2)**
deodorizer softener **(0)**
deoxidant solution **(0)**
deoxidant soap **(0)**
deoxidant softener **(0)**

Fig. 4 a. Prediction tree, b. Possible search strings

The predictions of partially typed words form a tree. Figure 4 shows the prediction tree and the resultant search strings when the user types "deo so ola". The yellow (sharp-cornered) rectangles represent the keywords in the repository, also called keyword nodes. The blue (round-cornered) rectangles are the partial search words entered by the user, also called the partial nodes. Keyword nodes are nothing but all possible extensions of their (parent) partial node, as found in the keyword repository.

A search string is associated with each keyword consisting of all keywords from the root to that keyword node. The prediction subtree is terminated at the keyword node where the associated search string returns zero results. For example, in Figure 4 a., the subtree rooted at *solution*, along the path *deodorant-solution* will be terminated since the search string "deodorant solution" returns zero results. Figure 4 b shows the possible search string for the prediction tree in Figure 4 a. The numbers in the parentheses indicate the number of results returned for those search strings. The number after the ':' in the partial nodes indicates the total results returned by all search strings corresponding to its children (keyword) nodes.

At any point, the user can press the *Return* key to start browsing through the list of search results. We use Microsoft's Speech API (SAPI) for the speech interface. The user is required to speak one word at a time, which is recognized correctly is appended to the search string. Microsoft's SAPI provides a set of alternates for every word recognized. These are the alternate keywords which replace the predicted keywords in the prediction tree, and are thus incorporated in the search. However Microsoft SAPI did not provide alternates in case of the grammar used for our product repository. Our speech interface is just a simple version of the SILO (Speech In List Out) approach proposed in (Divi et al, 2004).

5 Experiments

5.1 Robot-Assisted Shopping Experiments

The robot-assisted shopping experiments conducted with 10 visually impaired participants in Lee's Market Place, a supermarket in Logan, UT. We formulated several research hypotheses to test how well RoboCart functions as haptic and locomotor interface to the supermarket. As is often the case with studies involving visually impaired participants, it is not feasible to test in a statistically significant way all contributing factors in a single study (Bradley and Dunlop, 2005). The main reason is that the visually impaired population in the U.S. is not distributed evenly, with the majority living in just a few urban areas. Therefore, our hypotheses below address only a fraction of the factors outlined in the previous section.

Hypothesis 1: *If the robot is consistent overtime in how it sets up the target space with respect to a given product and verbally orients the shopper in the target space, the shopper's efficiency of maneuvering the haptic space in the target space increases with experience in the target space where experience is measured as the number of shopping iterations.*

Hypothesis 2: *The shopper's efficiency of maneuvering the haptic space in the target space is inversely related to the area of the target space.*

Hypothesis 3: *The shopper's efficiency of maneuvering the haptic space in the target space is inversely related to the complexity of the target space.*

Hypothesis 4: *In the absence of any prior knowledge of the target space, minor differences in sensory abilities affect the target space performance.*

Hypothesis 5: *The location of the product on the shelf (top, middle, bottom levels) affects the performance.*

5.1.1 Participants

Ten visually impaired participants from various locations in Utah were recruited for the experiments through the Utah Chapter of the National Federation of the Blind (NFB) in Salt Lake City, Utah. The Utah NFB Chapter provided us with a list of visually impaired Utah residents. Each individual on the list was first contacted by e-mail. The e-mail briefly described RoboCart and the experiments and asked the addressee if he or she would be interested in participating in the experiments. A brief phone interview was conducted with all those who responded positively. The inclusion criteria were: 1) the participant must be ambulatory; 2) the participant may not have any hearing or cognitive impairments; 3) the participant must understand English; and 4)the participant must be willing to travel to Logan, Utah, for a period of two days. Ten participants were thus selected. Each of the selected participants was transported to Logan, Utah, by vehicle for a period of two days and was paid a $90 honorarium.

5.1.2 Procedure

The procedure consisted of three stages. First, the individual was given a 30 minute introduction to RoboCart in our laboratory. The participant was trained in using RoboCarts keypad and used RoboCart to navigate a short route in the office space around the laboratory. The participant was then asked to take a technology readiness survey (Gockley and Mataric, 2006; Parasuraman, 2000) which was used to calculate the participants Technology Readiness Index (TRI).

Second, the participant was driven to Lees MarketPlace, a supermarket in Logan, Utah, and asked to use RoboCart to shop for several products. Twelve products were chosen from two aisles: 4 products from bottom shelves, 4 products from middle shelves, and 4 from top shelves. In Lees MarketPlace, each aisle consists of several shelf sections. A shelf section spans 4 to 5 meters in length and consists of 5 to 8 shelves. The selected products were divided into 4 sets. Set 1 included 3 products on top shelves; Set 2 included 3 products on middle shelves; Set 3 included 3 products on bottom shelves; Set 4 included one product on a top shelf, one product on a middle shelf, and one product on a bottom shelf. A single shopping trial consisted of the shopper picking up RoboCart from the docking area near the supermarkets entrance, navigating to each of the three products from a randomly chosen set, retrieving the products from the shelves, navigating to a designated cash register, unloading the product on the belt, picking them up on the other side of the belt, and navigating back to the docking area near the entrance. Figure 3 shows the route along which the experiments were conducted.

Five shopping iterations were conducted for each product set. The following measurements were taken during each run: 1) navigation time from location to location; 2) product retrieval time (time interval that starts when RoboCart stops and instructs the shopper on where the product is in the target space and ends when the shopper puts the product into RoboCarts basket); 3) the distance between the robot and the product; and 4) the number of other shoppers encountered on route. We also recorded observations regarding specific haptic cues used by the participants to find products. Two observers accompanied the participants on every run: the first observer was monitoring RoboCart; the second observer was making and recording measurements and observations.

Third, each participant was given an exit questionnaire to assess the subjective impression of the shopping experience with RoboCart. The questionnaire consisted of five questions whose answers were points on 10 point Likert scales (See Apendix A). If the participants response was below 5, the participant was asked to comment on why he or she gave a low mark.

5.1.3 Results

To test Hypothesis 1, the two populations were product retrieval times for the first and the fifth iteration, respectively. Each participant shopped for 12 products, which gave us 12 items in each sample. The paired t-test at p = 0.01 was used to compute the t-statistics for all participants. The resulting p-values for each of the t-statistics are statistically significant at the p= 0.01 level. There appears to be sufficient evidence to reject the null hypothesis that the shopper's efficiency in the supermarket is not affected by the shopper's exposure to the target space. As Figure 5 indicates, the product retrieval time reduces with every iteration. Thus, the participant's efficiency in maneuvering the haptic space in the target space appears to improve with experience. It is reasonable to expect that the shopper's performance in the target space eventually reaches an asymptote and becomes optimal given the participant's sensory, cognitive, and physical abilities.

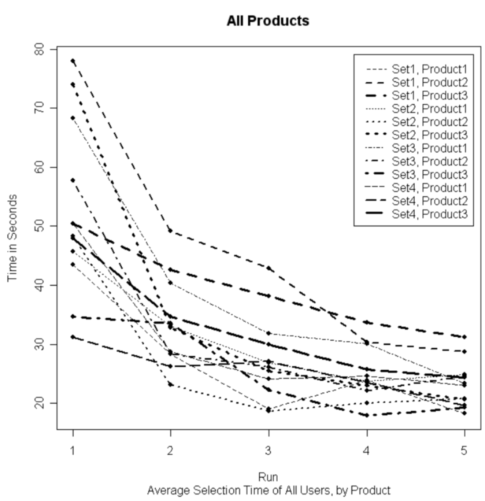


Fig. 5 Shopping Iteration Vs. Product Selection time.

To test Hypothesis 2, we measured the distance to the product from where the robot stops. Since the description that RobotCart gives to the shopper after it brings the shopper into the target space contains the direction and shelf number of the product (e.g., product X is on the third shelf on your right), the distance can be considered as an accurate indicator of the area of the target space. Surprisingly, we found a very low correlation coefficient (Pearsons product moment) of 0.37 between the target space area and the shopper's performance, which suggests that Hypothesis 2 may not hold for our sample. Our notes during the experiments and informal conversations with the participants after the experiments suggest that this outcome may be explained by the presence of various haptic cues in the target space. Certain haptic cues in the target space help the shopper retrieve the target product faster. One participant remembered the shape of the cooking oil spray can (one of the target products) and remembered that it was located near a label protruding from the shelf. Another participant remembered that a target product was located 10 barcodes to the left of a bad barcode which the barcode reader could not read.

To test Hypothesis 3, we used the product density as the measure of the target space complexity. The product density was computed as the number of products per foot between the robot and the target product on the correct shelf. The measurement was motivated by our previous ergonomic studies where it was found that the shoppers easily find the correct shelf from the robot's instructions but may have difficulties scanning the barcodes on the shelf. A general trend of decrease in efficiency with increasing complexity is observed. Product retrieval time and target space complexity have a correlation coefficient of 0.7.

To test Hypothesis 4, we compared how the participants performed with respect to each other. During the first shopping iteration, the shopper does not have any knowledge about the target space. Since the target space complexity is the same for all shoppers for any given product, it is sensible to suggest that the shopper's individual sensory and physical abilities will make the greatest difference in the absence of any knowledge of the target space. Using the data from all participants after the first shopping Figure 3: Performance Vs. Iteration: Learning in target space iteration, one-way ANOVA was used to test for a statistically significant difference in the target space performance between he participants. To make the test less biased, we removed Participant 10, because she had partial sight sufficient to detect product shapes and even read product labels at a close distance.

Among the other 9 participants, 4 had some residual vision and could see color blobs. However, none of them could read enlarged text. The other 5 participants were completely blind. No significant difference in performance was found (df = 8, F = 1.504, P = 0.17). One explanation is that RoboCart minimizes minor sensory differences of the shoppers and enables them to retrieve products in the absence of any knowledge of the target space.

One-way ANOVA was computed on the data from 9 participants after the fifth iteration, i.e., after the participants were exposed to the target spaces. It was found that there was significant difference in performance between participants after exposure (df = 8, F = 5.961, P ¡ 0.0001). Thus, minor differences in sensory abilities appear to make a difference given some knowledge of the target space. The shopper may be able to utilize his or her sensory abilities optimally after receiving some exposure to the target space, but not before.

Hypothesis 5 is based upon our conjecture that some parts of the target space are more easily accessible than others. We expected that there might be significant differences in performance of the shoppers between retrieving products from top, middle and bottom shelves. Using the data collected after the first iteration, one-way ANOVA was computed for three samples of runs, each of size 40. It was found that there was significant difference in performance (df = 2, F = 4.737, P = 0.011). We were now interested in finding out if knowledge of the target space was a factor. One-way ANOVA was computed on three samples of runs, each of size 40, obtained after the fifth iteration. No significant difference in performance was found (df = 2, F = 0.2701, P = 0.76). Some knowledge of the target space appears to make different parts of the space equally accessible.

We also tested whether the technology readiness index (TRI) (Gockley and Mataric, 2006; Parasuraman, 2000) is an indicator of how well the shopper performs with RoboCart. Each participant was given the TRI survey. The survey consists of four groups of questions to be answered on a Likert scale: Optimism, Innovativeness, Discomfort, and Insecurity. All four TRI components have low correlation coefficients with performance: 0.47, 0.29, 0.53 and 0.22, respectively. While the TRI may be a reliable predictor of a users readiness to use desktop computer technologies, it was not a reliable predictor of how the participants in our sample performed with RoboCart.

We conducted exit surveys to assess the participants' subjective impression of safety, smoothness, comfort, informativeness, and overall experience with RoboCart. The questions in our survey focused more on safety and comfort and are less generic than the NASA Task Load Index (NASA-TLX) questionnaire (Hart and Staveland, 1988) that attempts to assess the perceived level of workload using more abstract categories. We obtained the average values for each quality, in the range of 1 to 10, with 1 being the worst and 10 being the best. The averages were as follows: safety = 8.66; smoothness = 7; comfort of navigation = 8.5; informative feedback from the robot = 6.66; overall experience = 8.33. Low value for smoothness was given by one participant who thought that RoboCart made several sharp turns at the ends of aisles. Two participants gave low scores on robot feedback from the robot and indicated in their comments that they wanted more feedback from the robot during locomotion. We will address their comments in Section 5.

5.2 Product Selection Interface Experiments

It was logistically difficult to acquire a large number of blind participants for the study. Experiments were conducted with 5 blind and 5 sighted-blindfolded participants. The participants' age ranged from 17 years through 32 years and all participants were males. To avoid the discomfort of wearing a blindfold, the keypad was covered with a box to prevent the sighted participants from seeing it. The experiment was conducted in a laboratory setting. The primary purpose behind using sighted-blindfolded participants was to test whether they differed significantly from the blind participants, and thus decide whether they can be used in further experiments to replace blind participants. We propose to test the following planned research hypotheses. Let H1-0, H2-0, H3-0, H4-0 and H5-0 be the corresponding null hypotheses.

Hypothesis 1: (H1) *Sighted-blindfolded participants perform significantly faster than the blind participants, on average.*

Hypothesis 2: (H2) *The browsing interface is significantly worse than the typing interface.*

Hypothesis 3: (H3) *The browsing interface is significantly worse than the speech interface.*

Hypothesis 4: (H4) *The typing and speech interfaces are significantly different.*

5.2.1 Procedure

We used the product repository obtained from the household products database website Household-Products-Database (2004). The repository contains 11,147 products organised into a 4-level hierarchy. A grocery store typically contains over 20,000 products. We used the household products database since we were unable to obtain a sufficiently large grocery database.

The following procedure was followed for each participant. After arriving at the lab, the participant was first given an introduction about the purpose of the experiments. Each participant took on an avergae of 20 minutes to become familiar with the interfaces. In that, the participant tried to select three product using each interface, under the guidance of the experimenter. The session-1 of experiments started after the training session. Each task was to select a product using an interface. A set of 10 randomly selected products (set-1) was formed. Each participant was thus required to perform 30 tasks (10 products x 3 interfaces). Because of his schedule, one of the participants was unable to perform the ten browsing interface tasks. The product description was broken down into 4 parts: product name, brand, special description (scent/flavor/color), and the text that would appear in the result. An example is given in table **??**. During the course of a task, if the participants forgot the product description, they were allowed to revisit it by pressing a key. This would be avoided if a task rises out of the personal needs of the participants.

For session-2, another 10 products (set-2) were randomly selected. After the initial 30 tasks in session-1, , 20 more tasks were performed by each participant (10 products x 2 interfaces). Since the typing and speech interfaces were of more interest to us, we skipped the browsing interface tasks in session-2. The purpose of session-2 was to check if, and how much the participants improved on each of the two interfaces, relative to the other. Since the all the tasks were not necessarily of the same complexity, there was no way to check the learning effect. After both sessions, we conducted a subjective evaluation of the interfaces by monitoring the NASA Task Load Index

Table 1 Main Effects

Source	Main Effects (ANOVA)
Interface	F(2,243)=42.84, P<0.0001
Condition	F(1,243)=9.8, P=0.002
participant	F(8,243)=9.88, P<0.0001

(NASA-TLX) to each participant. The experiments were first conducted with 5 blind participants and then with 5 sighted, blindfolded participants.

5.2.2 Statistical Methodology

Repeated measures analysis of variance (ANOVA) models were fitted to the data using the SAS^{TM} statistical system **?**. Model factors were: interface (3 levels: browsing, typing, speech), condition (2 levels: blind, sighted-blindfolded), participant (10 levels: nested within condition, 5 participants per blind/sighted-blindfolded condition), and set (2 levels: set-1 and set-2, each containing 10 products). The 10 products within each set were replications. Since each participant selected each product in each set, the 10 product responses for each set were repeated measures for this study. Since the browsing interface was missing for all participants for set-2 products, models comparing selection time between sets included only typing and speech interfaces. The dependent variable was, in all models, the product selection time, with the exception of analyses using the NASA-TLX workload measure. The overall models and all primary effects were tested using an α-level of 0.05, whenever these effects constituted planned comparisons (see hypotheses). However, in the absence of a significant overall F-test for any given model, post-hoc comparisons among factor levels were conducted using a Bonferroni adjusted α-level of 0.05/K, where K is the number of post-hoc comparisons within any given model, to reduce the likelihood of false significance.

5.2.3 Results

For an overall repeated measures model which included the effects of interface, condition, and participant (nested within condition), and the interaction of interface with each of condition and participant, using only set-1 data, the overall model was highly significant, F(26,243) = 7.00, $P < 0.0001$. The main effects observed within this model are shown in table 1. All the main effects were significant. Interaction of interface x condition, F(2, 243)=0.05, P = 0.9558 and interface x participant, F(14, 243)=1.17, P = 0.2976 was observed. Thus, mean selection time differed significantly among interfaces, but the lack of interactions indicated that the interface differences did not vary significantly between blind and sight-blindfolded groups, nor among individual participants. In the ANOVAs, note that the DoF for the error is 243 because one of the participants did not perform the browsing tasks.

Mean selection time for the group of blind participants was 72.6 secs versus a mean of 58.8 secs for sighted-blindfolded participants, and the difference in these means was significant (t = 3.13, P = 0.0029). As might be expected, participants differed on mean selection time. However, the majority of the differences among participants arose from blind participant 5, whose mean selection time of 120.9 (s) differed significantly from the mean selection time of all others participants (whose mean times were in the 53-63 secs range) ($P < 0.0001$ for all comparisons between blind participant 5 and all other participants). When blind participant 5 was dropped from the analysis, the main effect of both condition and participant(condition) became non-significant ($F(1, 216) = 0.16$, $P = 0.6928$, and $F(6,216) = 0.44$, $P = 0.8545$, respectively). The interactions of interface with each of condition and participant remained non-significant also. It appears that on an average, when the outlier (participant 5) was removed, blind and sighted-blindfolded participants did not really differ. Thus the data leads us to confirm the null hypothesis H1-0, and thereby reject H1.

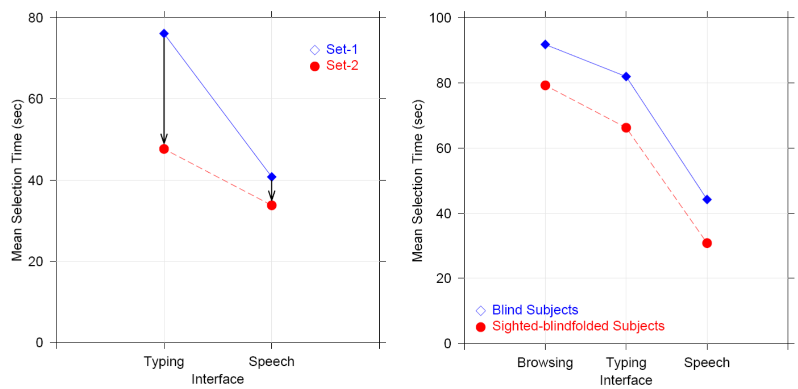


Fig. 6 a. Mean selection times for blind and sighted-blindfolded participants against all interfaces, b. Change in mean selection times for typing and speech interfaces from Session-1 to Session-2.

Further, a graph of the mean selection times of the blind and the sighted-blindfolded participants for each interface is shown in figure 6a. The almost parallel lines for the blind and sighted-blindfolded participants suggest that there is no interaction between the interface and the participant type, which is also confirmed by the ANOVA result presented earlier. In other words, the result suggests that the interface which is best for sighted-blindfolded users will also be best for blind users. We therefore take the liberty not to make any explicit distinction between the blind and sighted-blindfolded participants, during the remaining analysis in this paper.

The main effect of interface type as shown in table 1, suggests that on an average (over all participants), two or more interfaces differ significantly. Mean selection times for the 3 interfaces were, respectively: 85.5, 74.1, and 37.5 (seconds). Post-hoc pairwise t-tests showed that the typing interface was faster than the browsing interface ($t = 2.10$, $P = 0.0364$), although statistical significance is questionable if the Bonferroniadjusted is used here. We therefore were unable to reach a definite conclusion about H2. Each of the browsing and typing interfaces was significantly slower than the speech

Table 2 Post-hoc t-tests to study workload, mental demand and frustration imposed by the modalities ($\star$ indicates a signifcant test)

	Post-hoc t-tests		
	Browsing x Typing	Browsing x Speech	Typing x Speech
Mental Demand	t=1.075, P=0.2962	$\star$ t=3.822, P=0.0012	$\star$ t=4.011, P=0.0008
Frustration	$\star$ t=6.974, P<0.0001	t=1.348, P=0.1833	$\star$ t=4.428, P=0.0004
Overall Workload	$\star$ t=3.369, P=0.0034	$\star$ t=4.126, P=0.0006	t=0.9910, P=0.3348

Table 3 Mean values of mental demand, frustration and overall workload

	Browsing (0-100)	Typing (0-100)	Speech (0-100)
Mental Demand	45.6	35.9	13.4
Frustration	47.8	1.8	34
Overall Workload	64.4	41.65	35

interface (t = 8.84, P < 0.0001, and t = 6.74, P < 0.0001, respectively). This led us to reject the null hypothesess H3-0 and H4-0.

Since this difference in typing and speech interfaces concerned us the most, we decided to compare the interfaces on the measures obtained from the session-2 of the experiments. Set-2 was significantly faster than set-1, averaged over the two interfaces and all participants (t = 6.14, P < 0.0001). Since we did not have a metric for the task complexity, we were unable deduce if this result reflected the learning effect of the participants from session-1 to session-2. However, a significant interaction of interface x set, F(1, 382)=13.8, P=0.0002 was observed. The graph of the selection times during sessions 1 and 2, against the interface type is shown in figure 6b. It appears from the graph (figure 6b) that the improvement with the typing interface was much larger than that with the speech interface. The reduction in selection times from session-1 to session-2 varied significantly for the typing and speech interface (t = , P < 0.0001). This was probably because, since the participants were already much faster with the speech interface than the typing interface suring session-1, they had much less room to improve with the speech interface during session-2.

A strong Pearson's product moment correlation was found between selection time and query length for both typing and speech interfaces, with r = *0.92* and r = *0.82* respectively. To calculate the PPM correlation, we averaged the selection times over all products having the same query length. This just confirms the obvious that on an avergage, selection time increases with number of characters typed or words spoken.

We used a between-subjects design to study the data obtained from the NASA TLX questionaire. The *modality type* was the independent variable and *mental demand*, *frustration* and *overall workload* were the dependent variables.

A one-way ANOVA indicated that there was a significant difference among the three modalities in terms of the mental demand, frustration, and overall workload, . ((F(2, 27) = 16.63, P < 0.0001), (F(2, 27) = 16.63, P < 0.0001), and (F(2, 27) = 10.07, P = 0.0005) respectively). Post-hoc pair-wise t-tests for the three dependent variables with Bonferoni adjusted α-level of 0.016 are shown in Table 2. The mean values of mental demand, frustration and overall workload for the three modalities are shown in Table 3.

6 Discussion

The shopper's efficiency in maneuvering the haptic space in a given target space appears to improve with experience provided that the target space remains the same each time the shopper is exposed to it. By providing an effective interface to the locomotor space of the supermarket, the robot ensures the stability of the target space with respect to each product over time, which gives the shopper an opportunity to learn how to maneuver the haptic space inside the target space.

Ideally, the target space coincides with the haptic space. In practice, however, the best the robot can do is to bring the size of the target space as close to the size of the haptic space as possible. A low correlation coefficient was found between the target space area and the shopper's performance in our sample, which may indicate that the target space sizes ensured by RoboCart and the egocentric verbal instructions inside the target spaces contribute to faster product retrievals and minimize the effect of minor differences in visual abilities. The shopper's performance in the target space is likely to become optimal with respect to the shopper's sensory, cognitive, and physical abilities. It remains an open question what is the optimal size of the target space for a given shopper and a given product.

Our results suggest that the paucity of verbal instructions in target spaces may be desirable for some shoppers, because it does not raise the shopper's cognitive load and enables the shopper acquire individually valid haptic cues not explicitly mentioned by the system. Minor differences in sensory abilities appear to make a difference only after some knowledge of the target space has been acquired. In other words, the shopper may be able to put to personally optimal use his or her sensory abilities after receiving some exposure to the target space.

We have argued that, to guarantee independence, an assistive shopping device for the visually impaired must provide the shopper with interfaces to the haptic and locomotor spaces in the supermarket. Robotic mobility aids, such as GuideCane (Ulrich and Borenstein, 2001), Guido (Haptica, 2001), and Harunobu-6 (Mori and Kotani, 1998) focus on the haptic space and leave the handling of the locomotion space to the user. Recent proposals to address the visually impaired shopping problem with purely wearable means also stay in the haptic space (Lanigan et al, 2006). In our opinion, these solutions are unlikely to lead to the elimination of the barrier to large scale or small scale shopping, because they focus only on one space (haptic) and do not address the loss of independence. We believe that, to remain independent, the visually impaired shopper must be able not only to retrieve products from shelves but to handle the locomotor space in the supermarket. In addition, though the trichotomous ontology of spaces in the supermarket was developed to evaluate the shopper performance with RoboCart, we believe that it can be used to evaluate the performance of any assisted shopping device.

Are robots necessary to overcome this barrier for the visually impaired? For large scale independent shopping the answer appears to be affirmative because locomotion with a shopping cart in front may not be possible in the absence of vision. Whether or not wearable small scale shopping solutions will be found for the visually impaired may well depend on the ability of those solutions to provide effective interfaces to the locomotor spaces of supermarkets.

7 User Comments

User comments provide valuable insights into the limitations of our system. In this section, we will analyze several user comments that are relevant to the problem of robotic interfaces to the haptic and locomotor spaces in the supermarket, and the problem of rapid product selection interfaces.

7.1 Comments on RoboCart

Comment 1: *Instead of just following the robot, doing nothing, I would like to know what products I am passing by.*

The comment suggests that some participant may want to get more out of their shopping experience than just buying the required products. It is an open research question how much information is enough. More broadly, this question is related to the question of understanding and representing the users context to provide more useful computational services dynamically (Dey and Abowd, 2000; Jacquet et al, 2004; Tversky, B. and Lee, P., 1998). In the context of a supermarket, it is not practical to tell the user about each and every product that the robot passes by. Perhaps, informing the shopper about sales and higher product categories, e.g. cereal, canned products, etc, is what some participants would like RoboCart to inform them about.

Comment 2: *It would really help me if it can tell me in which direction to go after I scan every barcode. That way I can know if I miss a barcode.*

During the experiments it was observed that some participants would often overestimate the distance from the robot to the product. In such cases it was hard for them to realize that they should scan the next barcode in the opposite direction. What is needed is a detailed spatial topology of barcodes in target spaces. Once such a topology is available, RoboCart can give the user step-wise instructions to find the product in the target space: scan left, scan right, move one shelf up, etc.

Comment 3: *It would be really helpful if the robot could stop exactly in front of the product.*

RoboCart cannot stop exactly in front of the product. Even if the robot knew the x and y coordinates of every product in the store, it would still be unable to stop exactly in front of each product due to the presence of other shoppers in the store. When this was explained to the participant, the participant gave the following comment.

Comment 4: *Okay, I understand that. However, it is easier to stay behind the robot than go around it. So it would be nice if the robot could stop after passing the product.*

This is a fair comment that gives us a valuable insight on how we can improve the robot's interface to the target space in the future.

Comment 5: *It would be better if the robot can emit some beeps instead of the sonar clicking. The clicking sound gets irritating real soon and is difficult to trace in presence of background noise.*

This comment suggests that such seemingly minor issues like the sonar clicking can affect the shopper's experience. Since the shopper must return to RoboCart after retrieving the product from the shelf, RoboCart should emit adequate sonification beacons in the target space. What beacons are adequate is a research question that we hope to address in the future.

Comment 6: *How will this thing help me with produce?*

It will not. RoboCart can take the shopper to the target space of a given item in the produce section and instruct the shopper on how to find the target item. However, RoboCart cannot help the shopper recognize the freshness of fruits and vegetables. We consider this important problem to be outside of the scope of our project.

7.2 Comments on Product Selection Interfaces

Comment 1: *Though the speech-based interface works fast, the speech recognition errors are really frustrating.*

The comment conforms with the high frustration level indicated by the participants in the NASA TLX survey for the speech based interface. Often times some words could not be recognized and the participants would just sit there wondering what to do next. Such situations would be greatly discomforting for the shopper in a real supermarket.

Comment 2: *It would be nice to have some combination of speech and typing. That way I can speak a word that I can't spell.*

A hybrid interface which allows the user to switch between typing and speech could be advantageous. Though there is been significant research in multi-modal interfaces, none of these interfaces are targeted specifically toward rapid selection in large hierarchies.

Comment 3: *Browsing is useful when you are not sure what exactly you want to buy. It helps you to kind of browse through different kinds of products out there.*

We believe that the browsing interface is not appropriate for rapid product selection. However, the shopper can be allowed to switch to the browsing mode when the exact product is not known. Also the shopper might be able to select a product category fairly rapidly if the product hierarchy is trimmed to contain only the high level categories. The browsing efficiency with a trimmed hierarchy would however depend on the size and organization of the trimmed hierarchy.

8 Conclusions

An assistive shopping device for the visually impaired that ensures independence must provide the shopper with interfaces to the haptic and locomotor spaces in the supermarket. These interfaces should either eliminate the necessity of frame alignment or enable the visually impaired to reliably align the frames. Robots can function as interfaces to the haptic and locomotor spaces in supermarkets by eliminating the need of frame alignment in the locomotor space and cuing the shopper to the salient features of the target space sufficient for product retrieval.

The paucity of verbal instructions in target spaces may be desirable for some shoppers, because it does not raise the shopper's cognitive load and enables the shopper acquire individually valid haptic cues not explicitly mentioned by the system. Minor differences in sensory abilities appear to make a difference only after some knowledge of the target space has been acquired. In other words, the shopper may be able to put to personally optimal use his or her sensory abilities after receiving some exposure to the target space.

An interface which enables rapid selection of products is essential. Browsing interfaces which work well for smaller menu driven hierarchies, performs poorly for large object hierarchies. A hybrid interface combining speech and typing would be helpful.

To remain independent, the visually impaired shopper must be able not only to retrieve products from shelves but to handle the locomotor space in the supermarket. Thus, robot assistants may be necessary to for the visually impaired in large scale independent shopping, because locomotion with a shopping cart in front may not be possible in the absence of vision. Whether or not wearable small scale shopping solutions will be found for the visually impaired may well depend on the ability of those solutions to provide effective interfaces to the locomotor spaces of supermarkets.

Acknowledgements The second authord would like to acknowledge that this research has been supported, in part, through NSF grant (IIS-0346880), the National Eye Institute of the National Institutes of Health under Grant (1 R41 EY017516-01A1), and three Community University Research Initiative (CURI) grants (CURI-04, CURI-05, and CURI-06) from the State of Utah. We would like to thank Mr. Lee Badger, the owner of Lee's MarketPlace in Logan, Utah, for allowing us to use his supermarket as a research site. We are grateful to Mr. Ron Gardner, President of the Utah NFB Chapter, for his help with recruiting participants. We would like to thank the participants for their time and feedback.

References

Bradley N, Dunlop M (2005) An experimental investigation into wayfinding directions for visually impaired people. Ubiquitous Computing 9:395–403

Brent J, Modi N (2000) Shopping aid for the visually impaired. In: In the proceedings of conference on Rehabilitation Engineering and Assistive Technology Society of North America

Brewster SA (1998) Using nonspeech sounds to provide navigation cues. ACM Trans Comput-Hum Interact 5(3):224–259

Dey A, Abowd G (2000) Towards a better understanding of context and context-awareness. In: In Proceedings of the CHI Workshop on the What, Who, Where, and How of Context-Awareness, The Hague, Netherlands

Divi V, Forlines C, Gemert J, Raj B, Schmidt-Nielsen B, Wittenburg K, Woelfel P, Zhang F (2004) A speech-in list-out approach to spoken user interfaces. In: Proceedings of Human Language Technologies, Boston, MA

Fox D (1998) Markov localization: A probabilistic framework for mobile robot localization and navigation. PhD thesis, University of Bonn, Germany

Freundschuh S, Egenhofer (1997) Human conceptions of spaces: implications for geographic information systems. Transactions in GIS 2(4):361–365

Gaver W (1989) The sonicfinder: An interface that uses auditory icons. Human Computer Interaction 4(1):57–94

Gharpure C (2004) Orientation-free rfid based navigation in a robotic guide for the visually impaired. Master's thesis, Utah State University, USA

Gharpure C, Kulyukin V, Kutiyanawala A, Jiang M (2006) Passive radio frequency exteroception in robot assisted shopping for the blind. In: proceedings of the International Conference on Ubiquitous Intelligence and Computing (UIC), Wuhan and Three Gorges, China

Gockley R, Mataric MJ (2006) Encouraging physical therapy compliance with a hands-off mobile robot. In: In the proceedings of Human Robot Interaction (HRI) Conference, Salt Lake City, USA

Golledge R, Klatzky R, , Loomis J (1998) Cognitive mapping and wayfinding by adults without vision. In J Portugali (Ed) The Construction of Cognitive Maps pp 215–246

Hahnel D, Burgard W, Fox D, Fishkin K, Philipose M (2003) Mapping and localization with rfid technology. In: Technical Report, IRS-TR-03-014, Intel Research Institute, Seattle, Washington

Haptica C (2001) Guido (c). www.haptica.com

Hart S, Staveland L (1988) Development of nasa-tlx: results of empirical and theoretical research. In: In Hancock, P. and Meshkati, N. (Eds.) Human mental overload, North Holland, pp 139–183

Household-Products-Database (2004) www.householdproducts.nlm.nih.gov

Jacquet C, Bellik Y, , Bourda Y (2004) A context-aware locomotion assistance device for the blind. In: In Fincher, S. and Markopoulos, P. and Moore, D. and Ruddle, R. (Eds.) People and computers XVIII design for life, Springer-Verlag, London

Kantor G, Singh S (2002) Priliminary results in range-only localization and mapping. In: IEEE Conference on Robotics and Automation, Washington, D.C.

Kay L (1974) A sonar aid to enhance spatial perception of the blind: engineering design and evaluation. The Radio and Electronic Engineer 44:40–62

Klante P (2004) Auditory interaction objects for mobile applications. In: Proceedings of 7th International Conference on Work With Computing Systems, WWCS2004., Kuala Lumpur

Kulyukin V, Gharpure C, Nicholson J (2005) Robocart: Toward robot-assisted navigation of grocery stores by the visually impaired. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

Kulyukin V, Gharpure C, Nicholson J, Osborne G (2006) Robot-assisted wayfinding for the visually impaired in structured indoor environments. Autonomous Robots 21 (1):29–41, http://dx.doi.org/10.1007/s10514-006-7223-8

Kulyukin, V and Gharpure, C (2006) Ergonomics-for-One in a Robotic Shopping Cart for the Blind. In: Proceedings of the 2006 ACM Conference on Human-Robot Interaction, Salt Lake City, Utah

Kulyukin, V, Gharpure, C, and Pentico, C (2007) Robots As Interfaces to Haptic and Locomotor Spaces. In: Proceedings of the 2007 ACM Conference on Human-Robot Interaction, Arlington, Virginia

Kupiers B (2000) The spatial semantic hierarchy. Artificial Intelligence 119:191–233

Lahav O, Mioduser D (2003) A blind persons cognitive mapping of new spaces using a haptic virtual environment. Journal of Research in Special Educational Needs 3(3):172177

Lalatendu S, Pierce N, A M (2006) Cat eye: An assistance system for independent shopping. In: Proceedings of International Conference on Aging, Disability and Independence, St. Petersburg, Florida

Lanigan P, Paulos A, Williams A, Narasimhan P (2006) Trinetra: Assistive technologies for the blind. In: CyLab Technical Report CMU-CyLab-06-006, Carnegie Mellon University

Millar S (1982) The problem of imagery and spatial development in the blind. In B de Gelder (Ed) Knowledge and Representation pp 111–120

Millar S (1995) Understanding and representing spatial information. British Journal of Visual Impairment 13:8–11

Millar S (1997) Theory, experiment and practical application in research on visual impairment. European Journal of Psychology of Education 12:415–430

Mori H, Kotani S (1998) Robotic travel aid for the blind: Harunobu-6. In: Second European Conference on Disability, Virtual Reality, and Assistive Technology, Sovde, Sweden

Parasuraman A (2000) Technology reacdiness index(tri): A multiple-item scale to measure readiness to embrace new technologies. Journal of Service Research p 307

R Passini and G Proulx (1988) Wayfinding without vision. An experiment with congenitally totally blind people. Environment and Behavior 22:227–252

Raman TV (1997) Concrete implementation of an audio desktop. In: Auditory User Interfaces Toward The Speaking Computer

Scooter S, Helal S (2004) A passive rfid information grid for location and proximity sensing for the blind user. University of Florida Technical Report number TR04-009

Sidner C, Forlines C (2002) Subset language for conversing with collaborative interface agents. In: Mitsubishi Electric Research Laboratory, TR-TR2002-36., Cambridge, MA, USA

Smith A, Francioni J, Anwar M, Cook J, Hossain A, Rahman M (2004) Non-visual tool for navigating hierarchical structures. In: Proceedings of International ACM SIGACCESS Conference on Computers and Accessi-

bility (ASSETS)., Atlanta, GA
Tversky, B and Lee, P (1998) How space structures language. C Freksa, C Habel, and KF Wender (Eds), Spatial Cognition An interdisciplinary approach to representing and processing spatial knowledge pp 157–175
Tversky, B, Morrison, J, Franklin, N, and Bryant, D (1999) Three spaces of spatial cognition. Professional Geographer 51:516–524
Ulrich I, Borenstein J (2001) The guidecane: Applying mobile robot technologies to assist the visually impaired. IEEE Trans on Systems, Man, and Cybernetics, Part A: Systems and Humans 31:131–136
Ungar S (2000) Cognitive mapping without visual experience. In: In Cognitive Mapping: Past Present and Future, London: Routledge
W3C (2000) Web content accessibility guidelines 1.0. In: Web Accessibility Initiative
Walker B, Nance A, Lindsay J (2006) Spearcons: Speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the 12th International Conference on Auditory Display, London, UK
Wolf P, Woelfel J, Gemert J, Raj B, Wong D (2004) Spokenquery: An alternate approach to choosing items with speech. In: Mitsubishi Electric Research Laboratory, TR-TR2004-121., Cambridge, MA, USA