Decision-Theoretic Planning Meets User Requirements: Enhancements and Studies of an Intelligent Shopping Guide

Thorsten Bohnenberger[1], Oliver Jacobs[2], Anthony Jameson[2], and Ilhan Aslan[1,*]

[1] Department of Computer Science, Saarland University, Saarbrücken, Germany
[2] DFKI, German Research Center for Artificial Intelligence, Saarbrücken, Germany

Abstract. This paper reports on extensions to a decision-theoretic location-aware shopping guide and on the results of user studies that have accompanied its development. On the basis of the results of an earlier user study in a mock-up of a shopping mall, we implemented an improved version of the shopping guide. A new user study with the improved system in a real shopping mall confirms in a much more realistic setting the generally positive user attitudes found in the earlier study. The new study also sheds further light on the usability issues raised by the system, some of which can also arise with other mobile guides and recommenders. One such issue concerns desire of users to be able to understand and second-guess the system's recommendations. This requirement led to the development of an explanation component for the decision-theoretic guide, which was evaluated in a smaller follow-up study in the shopping mall.

Keywords: Mobile commerce, navigation support, decision-theoretic planning, user studies, recommender systems, explanation.

1 Introduction

A natural and popular application domain for pervasive computing is assistance for shoppers who are shopping in physical environments such as grocery stores, shopping centers, or larger areas such as entire towns.

This paper describes the development of a mobile shopping guide that focuses on one of the many types of assistance that have been explored to date: Given a shopper who wants to find a particular set of products within a limited time, how can the shopper be guided to the possible locations of these products in an order that tends to maximize the likelihood of finding the products while minimizing the time required

* This research was supported by the German Science Foundation (DFG) in its Collaborative Research Center on Resource-Adaptive Cognitive Processes, SFB 378, Project EM 5, READY; and by the German Ministry of Education and Research (BMB+F) under grant 524-40001-01 IW C03 (project SPECTER). We thank the six reviewers for their valuable comments.

H.W. Gellersen et al. (Eds.): PERVASIVE 2005, LNCS 3468, pp. 279–296, 2005.

to do so? What makes this problem more than just a navigation problem is the uncertainty that pervades many aspects of it: In particular, there is considerable uncertainty about whether the shopper will find a desired product at a given location and how long she will have to spend looking for it. Despite these uncertainties, the shopper may be under pressure to complete the trip quickly, perhaps even before a fixed deadline.

In the rest of this section, we summarize our earlier work on a decision-theoretic shopping guide that addresses this problem and place it in the larger context of mobile shopping assistants. In the remainder of the paper, we discuss recent significant enhancements to the system and how they have been tested in a real shopping environment.

1.1 A Decision-Theoretic Shopping Guide

Bohnenberger and Jameson [1] introduced the basic idea of a decision-theoretic shopping guide: Given a set of products that a user wants to buy, a set of stores (e.g., within a shopping mall) where she may be able to find them, and estimated times for traveling among the stores, such a system generates a *policy* for guiding the user among the stores: At any given point in time, given knowledge of where the user is and what products she has already found, the system recommends the next store to visit and gives directions for getting to that store. Decision-theoretic planning (cf. Section 2.2) is generally well suited to this recommendation task in that it can take into account not only the costs of visiting particular stores but also the uncertainty about whether the desired product will be found in a given store.

To obtain some initial feedback about the usability and acceptance of this type of guide, we created an initial prototype for a PDA[1] and conducted a study in an artificial mockup of a shopping mall (Bohnenberger et al. [2]): A number of stores were crudely simulated on two floors of a computer science building. Localization was realized with infrared beacons attached to the walls at a number of points, which transmitted identifying signals to the subject's PDA. Each of 20 subjects performed two assigned "shopping" tasks, with the role of the shopkeeper being played in each case by the experimenter, who followed the subject around. Each subject performed one shopping task with the PDA-based guide and the other one with a conventional paper map of the shopping mall that they were allowed to study in advance. Shopping performance with the shopping guide was somewhat better than performance with the paper map, illustrating the basic viability of the method. Even the artificial setting enabled the subjects to suggest a number of improvements to the user interface, which were taken into account in the next version (see Section 2). On a more general level, although the study did not itself create a realistic shopping experience, it did make the functionality clear to the subjects, so that they could offer some speculative comments about their willingness to use a system of this type in a real shopping situation. Their overall attitudes were almost uniformly positive, and they showed appreciation for the system's ability to save them time, cognitive effort, and the frustration of having to replan when the store did not have the product that they had expected to find there. On the negative side,

[1] More precisely: a Compaq iPAQ Pocket PC.

a number of subjects felt that they did not have the "big picture", almost as if they were being led blindfolded through a shopping mall.

On the basis of the user feedback from the study, we developed a version of the shopping guide that included a number of improvements (see Section 2), ranging from minor interface improvements to significant enhancements that address fundamental user requirements. This improved prototype was then tested in a real shopping mall (see Section 3).

1.2 Related Work

Shopping Assistant Prototypes. A large number of systems have been developed since the early 1990s that offer various types of assistance to a shopper. Asthana et al. [3] presented an early portable shopping assistant which, among other functions, helped the shopper to locate particular products within a large store and alerted the shopper to potentially interesting special offers.

iSHOPPER (Fano [4]), like our own guide, assisted the user within a larger geographical area, such as a shopping mall. It alerted a shopper to desired products that were available near his or her current location. This approach is the opposite of the one taken by our guide: Instead of taking the shopper's location as given and recommending products that can be conveniently bought nearby, our guide starts with a list of desired products and guides the user to the locations at which she is most likely to find the products in accordance with an efficient policy.

The more recently developed shopping assistant iGrocer (Shekar et al. [5]) offers a variety of services to a shopper in a grocery store via a mobile phone. In particular, it shows the "quick shopper" the fastest route within the grocery store for picking up the products on the shopping list. This function is similar to that of our shopping guide, except that it does not take into account uncertainty about whether a given product will be found at a given location—a capability that requires the probabilistic reasoning that is characteristic of decision-theoretic planning.

Recently, Cumby et al. [6] have shown how a shopping assistant can predict (and therefore suggest) a shopper's current shopping list on the basis of information about past purchases. Again, this method is complementary to our approach, which presupposes that a shopping list already exists (whether specified entirely by the shopper or with external assistance).

On the whole, existing shopping assistants like these offer functionality that is complementary to that of our decision-theoretic shopping guide. In the long run, it should be possible to integrate various types of functionality offered by different systems into a more comprehensive shopping assistant.

Shopper-Centered Studies. Newcomb et al. [7] have carefully examined the requirements of grocery shoppers for mobile shopping assistants. Some of their results confirm the importance of the goals of our shopping guide: Their survey of 46 diverse shoppers showed that two of the three most widely requested features of a grocery shopping assistant concerned help in arranging the items on the list and in locating the items within the grocery store. The authors also found that shoppers view waiting in checkout lines as a major nuisance in the shopping process; our guide addresses this problem in that it takes into account the expected length of time that a shopper will have to spend in a given store. In a test of their prototype with five shoppers in a grocery store, the authors noted a tendency that was commented on by some of the subjects in our first study: the tendency to focus on getting the products on their list quickly, as opposed to exploring the shopping area.

Taken together with our early results, these results indicate that (a) a decision-theoretic shopping guide can fulfill important requirements that shoppers often have; but (b) it tends to discourage the kind of recreational and exploratory shopping that is in many cases desired by shoppers and store owners. Since this latter fact is sometimes claimed to limit the practical deployability of our approach severely, we should point out that the recreational and convenience orientations to shopping are not as contradictory as they may seem (cf. Kim and LaRose [8]). For example, a shopper who intends to do both some convenience shopping and some recreational shopping in a given visit to a shopping mall may have more time and energy for the recreational shopping if she can finish the convenience shopping quickly and with little effort. And the fact that a given shopping mall has infrastructure that supports efficient convenience shopping (as well as recreational shopping) may constitute a reason for visiting that mall in the first place rather than another one.

Decision-Theoretic Planning for Recommendation. In addition to our own work cited in Section 1.1, other researchers have recently applied decision-theoretic methods to recommendation problems in shopping contexts. Plutowski [9] addresses shopping scenarios very similar to those addressed by our guide, offering new solutions to the complexity issues raised by decision-theoretic planning, which tend to limit the scalability of decision-theoretic methods. New approaches to complexity issues are also presented in the work of Brafman et al. [10] (first introduced by Shani et al. [11]), whose shopping domain is an on-line bookstore, in which the authors were able to test their decision-theoretic recommender in real use.

2 Description of the New Prototype

On the basis of the results of our first user study (Section 1.1) and the related research just summarized, we developed an improved version of the decision-theoretic shopping guide.

2.1 User Interface

Before beginning with her actual shopping, the user must specify to the system in one way or another which products she would like to buy in the shopping mall. In the relatively simple solution that we adopted for this prototype, the shopping list is specified on a larger stationary computing device, on which the computations necessary for the decision-theoretic planning are also carried out.[2] The usability and acceptance of this aspect of the prototype were not evaluated in our study.

[2] In a practical deployment, this device might be a computer with touch screens mounted in the walls of the shopping mall.

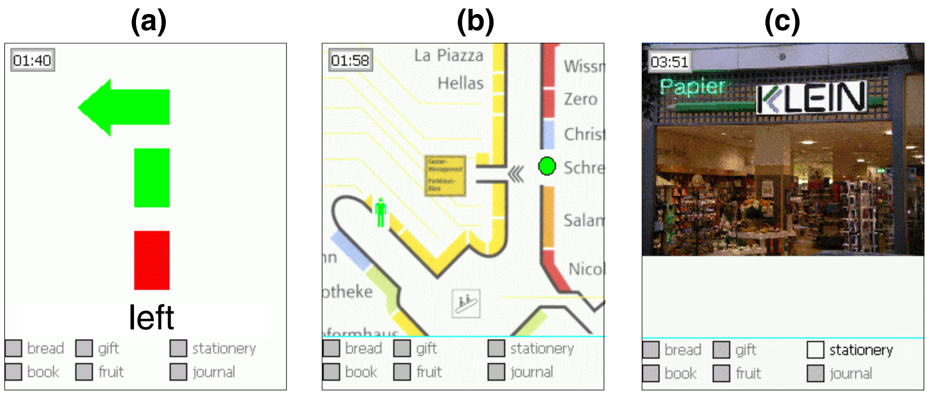

Fig. 1. Three screen shots of the improved shopping guide for the user study in the shopping mall. (a: Animated arrows are used for user-centered navigation recommendations. b: Overview maps show the user's location, the expected next store, and further information about the environment. c: Shortly before the user reaches the expected next store, a picture of the store is shown on the display of the PDA to help the user recognize the destination.)

The main display on the PDA has two parts (see Figure 1): On the bottom of the display, the user can always see the shopping list (which included 6 items in the current user study) with check boxes for indicating to the system which items have been purchased so far. As is shown in screen shot (a), a large animated arrow is used to indicate the direction in which the user should walk whenever she reaches a point at which a choice is available. The animation shows movement from the beginning of the arrow to the end, so as to emphasize that the recommended motion is relative to the user's current orientation. Between one such instruction and the next, an overview map (b) is presented that shows a part of the shopping mall that includes both the user's current location and the next store to be visited. These overview maps were introduced in response to comments of subjects in the first study that they lacked a "big picture" of the shopping mall and wanted to be able to prepare mentally for the next store that they were to visit. When the user approaches the next recommended store, the system displays a photograph (c) of the store, so as to reassure the user that the intended store has been reached.

The user is alerted to each change in the display by an acoustic signal—a feature suggested by users in the first study who wanted to minimize the extent to which they had to attend to the display.[3]

When the user has reached a store, the items available in that store become active in the shopping list, so that the user can check them off if she finds them. Inside a store, the user alone is responsible for finding the right places to look for the desired items. The next navigation recommendation is given to the user when she leaves the store.

[3] In the user study, users wore an ear set, which ensured that they heard the signals regardless of the level of ambient noise. It would also be possible to give complete navigation instructions via speech, as several subjects in the earlier study requested.

2.2 Decision-Theoretic Planning

Data Required. For the purpose of this paper, the most important questions concerning the decision-theoretic approach concern (a) the data required by the system and how they can be obtained and (b) the nature of the result of the planning process and the computational resources required.[4]

Table 1. Overview of the types of data about the shopping environment required by the decision-theoretic planner

A topography of the shopping mall that specifies:	For each shop, for each of a set of *product characterizations*:
1. the location of each shop	1. the probability of finding a product fitting that characterization
2. the location of each of a set of beacons, each of which sends an infrared signal to the PDA of the user as she approaches the beacon	2. the typical time spent in the store searching for a product
3. for each beacon, the typical time required to walk between it and each of the neighboring beacons	3. the typical time spent waiting in line after a product has been found

Table 1 gives an overview of the required data. A *product characterization* can have different degrees of specificity. For example, for a bookstore, it might specify the name of a particular magazine; in which case the probability refers to the likelihood that a shopper who enters the store looking for that particular magazine will find it. Or the specification might be more general, such as “sports magazine”, in which case the probability refers to the likelihood that a shopper who is looking for either a particular sports magazine or simply some sports magazine that she considers worth buying will end up finding it in this store.

The question of where all of these quantitative estimates might come from in real-life applications of the technology is open-ended, because there are many possibilities (see Bohnenberger [12], chap. 6). Some solutions presuppose a certain amount of cooperation by the vendors in the shopping mall (e.g., some sort of access to their databases). Other solutions would make use of data collected during the actual use of the system to update parameters such as the probability of finding a product fitting a given characterization in a given store, provided that at least some users allowed some data concerning their shopping trips to be transferred to a central computer. A simple, low-tech method was applied for our user study: To determine the time cost of walking from one store to another, the experimenter did the necessary walking and counted the steps required. To estimate the probability of finding a particular type of item in a given store and the time required to find and buy it, the experimenter looked at the number and variety of products offered by the store.

[4] For information on technical aspects of the planning process, we refer the interested reader to the brief description given by Bohnenberger et al. [2] and the detailed account given by Bohnenberger [12]. General introductions to decision-theoretic planning are given by Boutilier et al. [13] and Russell and Norvig [14], among others.

Result of the Planning Process and Computational Resources. The result of the planning process is not a single route through the shopping mall but rather a recommendation *policy*: For each possible *state* of the shopping process, the policy specifies a recommended user action, which may involve either walking in a given direction between stores or entering a particular store. A state in the shopping process is characterized by the user's current location, the set of products that the user has bought so far, and (if a time limit has been specified), the amount of time left until the deadline. Note that, if the system generated a fixed plan, it would be determined in advance, for example, how many bookstores the user would visit. With a policy, the user will visit as many bookstores as is necessary to obtain the desired book(s)—unless time runs out first to the point where it is no longer worthwhile to look for books.

Although the resulting recommendation policy covers a large number of possible states, it can be represented and applied on a resource-limited PDA, because the mere application of an existing policy is not computationally demanding. By contrast, the planning process that generates the policy can be highly resource-intensive; for this reason, it may be necessary for this computation to be performed on a larger computing device, as was done for our user study.

3 User Study in a Real Shopping Mall

By testing the use of the prototype in a real shopping mall, we aimed to answer the following questions:

1. Does the shopping guide work as expected in effectively enabling users to find the desired products in an unfamiliar shopping mall and with limited time available?
2. How do potential users evaluate the shopping guide overall, and in what situations would they be inclined to use such a system for their own shopping?
3. What features of the system are well received, and where do users see a need for improvement?

3.1 Method

Realization of Localization. A major practical obstacle to a study in a real shopping mall was the infeasibility of installing localization infrastructure just for the purpose of the study. Therefore, in our study, the experimenter simulated the localization infrastructure of the earlier user study with a second PDA that communicated with the device of the subject. Whereas the subject's PDA would normally receive infrared signals from beacons mounted at various locations in the shopping mall, in our study these signals were transmitted from the experimenter's device via wireless LAN (it is convenient to speak of these signals as coming from *virtual beacons*). The experimenter's device ran a specially written program that automated as much as possible of the task of simulating the beacons.[5]

[5] The program exploited the fact that, given knowledge of the mall topology and the current shopping policy, the movements of a user were largely predictable—although on a few brief occasions subjects moved too unpredictably for the experimenter to keep up.

Shopping Mall. The study was conducted in the *Saarpark-Center* shopping mall[6]in the city of Neunkirchen—the largest shopping mall in the German province of the Saarland. The mall hosts about 120 stores on 33,500 m^2 spread over two floors.

Shopping Task. Instead of conventional payment for participation in the study, each subject was given a fixed budget of 25 Euros (about $30) to spend. With this money, the subject could buy (and keep) 1 item from each of 6 categories (some bread, a book, a gift item, some fruit, a magazine, and some stationery). In order to recreate the normal real-life situation in which a shopper has some particular ideas in mind about the product being sought (as opposed to looking for, say, just any magazine) the subject was asked to specify in advance, for each category, some specific characteristics of the product sought (e.g., a particular magazine or a particular type of bread). This measure also introduced some (realistic) uncertainty as to whether the desired product would in fact be found in any given store of the relevant type. Each subject was allowed to shop for at most 20 minutes. This restriction was explained to the subjects with reference to the real-life situations of having to finish shopping before the closing time of the mall or before leaving the mall to go to an important appointment.[7]

Subjects. We recruited 21 subjects (10 female, 11 male),[8] aiming mainly to find subjects who were not familiar with the shopping mall in question (because subjects who are thoroughly familiar with a shopping mall can be expected to benefit relatively little from a shopping guide). Of the 21 subjects, 14 had not been to the shopping mall within the previous 12 months, and none had been there more than 3 times during that period. Only 6 of the subjects had experience with navigation systems, in most cases with car navigation systems. Only 2 subjects had ever used a PDA. The subjects' ages ranged from 21 to 73 years, with a median of 27.

Procedure. Each subject performed the shopping task individually, accompanied by the experimenter. First, the experimenter familiarized the subject with the use of the shopping guide. He then explained the shopping task described above and asked the subject to characterize the specific product to be sought in each category. The subject began at the mall's main entrance and was followed around by the experimenter. Each time the subject's PDA received a signal from a virtual beacon, the shopping guide displayed the navigation recommendation specified by the previously computed policy. When a subject found one of the desired products in a store, he or she checked it off of the list, and the experimenter paid for it. When the 20-minute shopping period was over, the subject filled in a questionnaire and was subsequently asked in a verbal debriefing to elaborate on her answers.

[6] http://www.ece.de/de/shopping/center/spn/spn2.jsp

[7] We did not include a control condition in which the subjects shopped without the shopping guide, because (a) we wanted to devote all available resources to the study of the shopping guide in a natural setting and (b) we assumed that subjects already had a great deal of experience in shopping in shopping malls in the normal way.

[8] Since no reliable gender differences were found in the results, the gender variable will not be mentioned again.

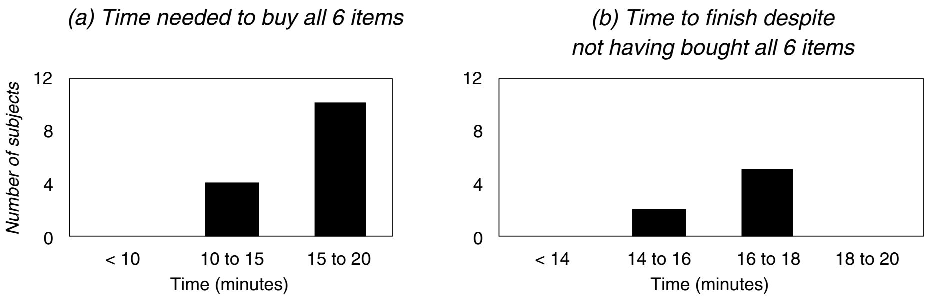

Fig. 2. Time required to finish shopping for subjects who (a) did and (b) did not find all 6 items

3.2 Results

The study proceeded without problems, and the subjects supplied detailed comments on the questionnaire and in the debriefing.

Subjects' Performance with the Shopping Guide. The recommendation policy used by the shopping guide, which took into account the time limit of 20 minutes, did in fact ensure that all subjects reached the mall exit before the deadline had elapsed. This result is not trivial, because of course not all subjects proceeded with the exact speed assumed in the underlying model. Even when a subject walked more slowly than predicted, spent a relatively long time in stores, or simply experienced bad luck in finding the particular products that she had specified, the recommendation policy was able to bring the subject to the exit in time by in effect "giving up" on finding one or more products.

Six of the subjects were guided to the exit although they had managed to find only 5 of their 6 products (or only 4, in the case of one of these subjects). In all of these cases, the shopping guide recommended skipping the (relatively expensive) gift item, which would have been bought at a gift store near the mall exit. When they arrived at the gift store, these subjects had between 2 and 6 minutes left—less than the typical time assumed in the system's model for finding and paying for a gift item. Therefore, the system acted appropriately in advising them to bypass the gift store, given the higher-priority goal of reaching the mall exit in time. But why didn't the policy recommend skipping a less valuable item (such as the magazine) at an earlier stage of the shopping trip, freeing enough time to look for a gift later? In fact, a thorough analysis of the policy reveals that the system exhibits exactly this behavior in some time windows. But if a subject performs well in the early stages of the shopping, the system in effect does not consider it necessary to skip a less important item, expecting that the subject will manage to complete the entire shopping task. If the subject shops more slowly than expected later, the system has no choice but to recommend skipping the last item, even if it happens to be the least desirable one to skip.

Having to skip the gift store was a potentially disappointing result for those subjects who had almost enough time left when they reached this store. If these subjects had understood exactly what was going on, instead of going straight to the exit they could have taken an action that was not represented in the system's model: They could have entered the gift store, seen if they could find a suitable gift especially quickly (keeping an eye on the remaining time) and if necessary left the store without a gift just before the time ran out. It would be possible to enrich the system's model so that it could actually recommend this type of action; but generally speaking situations can often arise in which it is desirable for the user to be able to second-guess the system's recommendation. We will return to this point in Section 3.3.

Subjects' Assessment of the Shopping Guide. Figure 3 summarizes answers that subjects gave to questions concerning their acceptance of the shopping guide. It is generally agreed (see, e.g., Landauer [15]) that subject's expressions of positive attitudes toward a novel system should not be taken at face value, given that the subjects in a study of this type are often biased in favor of a positive evaluation. Still, when we consider all of the frustrations that subjects might conceivably have experienced in this situation, the small overall number of problems and complaints does offer some encouragement.

When asked about conditions in which they thought that the shopping guide would be particularly useful, 16 subjects mentioned unfamiliar environments, 13 referred to situations involving time pressure, and 5 spoke of particularly large shopping malls.

Asked about particularly positive aspects of the shopping guide, 9 subjects mentioned the speed with which they were able to find relevant stores; 4 subjects com-

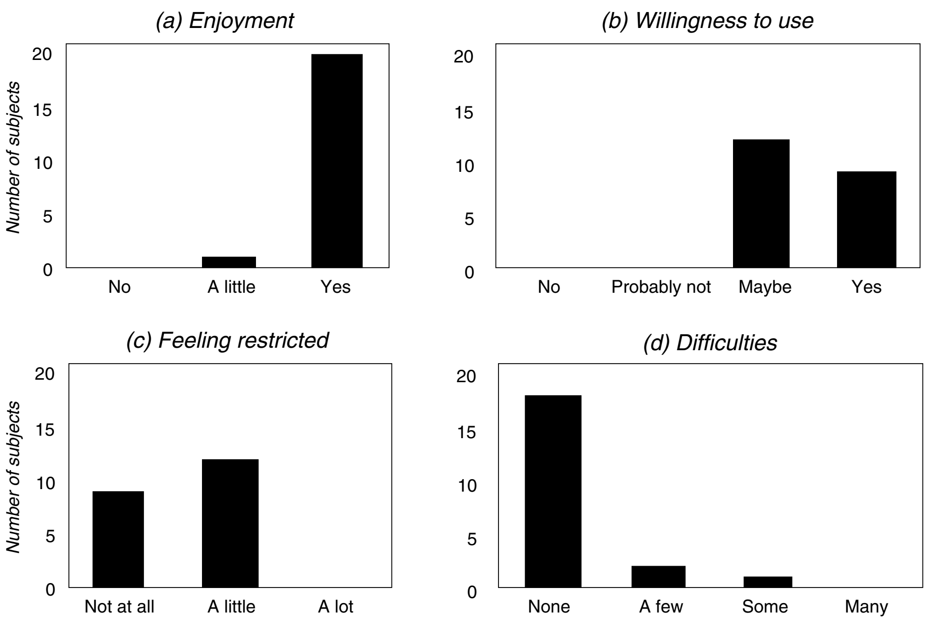

Fig. 3. Subjective acceptance of the shopping guide. The questions were: (a) "Did you enjoy using the shopping guide?" (b) "Would you use such a system in a suitably equipped environment?" (c) "Did you feel restricted by the system? (d) "Did you experience difficulties in using the shopping guide?"

mended the reliability with which they received the recommendations at the right locations; 4 subjects mentioned the photos of the stores that are displayed when a store is reached; 3 subjects singled out the form in which the recommendations were presented via the arrows on the PDA's display; and 2 subjects mentioned the overview maps.

Of those who stated that they would "maybe" use a shopping guide of this type in real shopping situations, 3 subjects urged that for each recommended store, one or more similar, alternative stores should be presented. On a related note, 2 subjects expressed a desire to be able to walk in a different direction than the one recommended by the system. This point will be discussed in Section 3.3.

Whereas the overview maps had been introduced to help subjects to prepare mentally for the next store, 3 subjects pointed out that it would be even better to be reminded which particular product(s) they would be looking for at the next store.

As Figure 3(d) shows, 3 subjects stated that they had experienced difficulties in using the shopping guide. One of the specific difficulties mentioned illustrates a more general issue. It concerned an inadequacy in the system's modeling of products: In order to look for a gift, one subject would have liked to enter a bookstore; but the system's modeling had assumed that gifts can be bought only in gift stores. Although a specific problem of this sort might be corrected post hoc through an improvement in the modeling, there will presumably often arise situations in which the user's conception of the relevant products differs from the system's conception. In such cases, it would be desirable for the user to be able to second-guess the system, choosing an action that deviates from the system's recommendation.

In the questionnaire, the subjects were also asked if they felt restricted by the use of the shopping guide. Twelve subjects responded "A little". Most of the complaints concerned physical restrictions: the need to carry the device in one hand while shopping (8 subjects) and the wire that connected the earphone to the PDA (3 subjects). But 4 subjects mentioned cognitive restrictions, complaining that the shopping guide took their attention away from the environment or diminished their freedom of decision about where to go next. It is noteworthy that another subject considered it to be an advantage that the shopping guide took his attention away from the environment because as a result he was not tempted to buy any items that he did not really need.

3.3 Discussion

Taken together, these results indicate three directions in which the tested prototype still calls for improvement, despite the generally positive feedback:

Further Reduction of Distraction from the Environment. One way in which all mobile and wearable systems can hamper the natural performance of a user's task is simply by unnecessarily consuming physical or mental resources that could otherwise be devoted to the task. Although our efforts to minimize this type of distraction (e.g., easy-to-perceive animated arrows; acoustic alerts to changes in the display) were partly successful, hardware improvements that minimize the need to hold the PDA in one's hand and to use a wire with the earphone are still called for.

Allowing the User to Second-Guess the System. As we have seen, there are various situations in which a user may have good reason to deviate from the system's recommendation, and several users expressed a desire to do so. The system's modeling of the user's shopping needs, preferences, and capabilities can never be more than a serviceable approximation. In addition to the observed examples mentioned above, the user may, for example, have a strong liking or disliking for a particular chain of stores, which the system is unlikely to know about. In such cases, the user should have the option of deviating from the system's recommendation without sacrificing the benefits of using the system; but she should also be able to judge whether deviation from the recommendation will in fact lead to an improvement or not. The results of the study suggest several specific enhancements along these lines:

1. One reason why subjects felt that they could not second-guess the system is that they had little idea of the consequences of doing so. In reality, the system is quite robust in this regard, much like a car navigation system that computes a new route if the user for some reason leaves the recommended route: Even if the user were to disregard the system entirely for a while and wander around at will, as soon as she began consulting the recommendations again, they would once again constitute an optimal policy in view of her current situation, including the amount of time remaining.[9] In such a case the user's sequence of actions before and including the deviation may no longer be optimal in any sense. On the other hand, if the user's deviation is minor and well-founded (e.g., going into a nonrecommended store because the user sees a desired product displayed in the store window), it is likely that the entire sequence of actions will be more successful than it would have been if the user had followed the recommendations strictly.

Since these basic facts about the system were not clear to the subjects in the study, it is understandable that they were hesitant to deviate from the recommendations even when they saw the desirability of doing so. One strategy for improvement, therefore, is to convey to the users in some appropriate way the necessary understanding of the system's capabilities.

2. Even if users know that they can deviate from the recommendations, it will be difficult for them to do so if they have little information on which to base such a decision. The users in our study could see the stores around them, and the overview maps also showed stores that were not immediately visible. But it was not made explicit what alternative options the user had for finding the desired products. A second strategy for improvement, then, is to give users some sort of preview of what will happen if they take some action that was not recommended by the system.

One possible way of realizing both of the solutions just proposed is to follow a trend that is becoming increasingly popular in the area of recommender systems (see, e.g., Herlocker et al. [16]): providing a mechanism for explaining the system's recommendations. The next section will describe how we designed and implemented a simple explanation component and how users responded to it in a small follow-up study.

[9] The system does not even need to recompute its policy in this case, because the originally computed policy already takes into account all possible states that the user might reach, as long as they involve the modeled locations and the relevant time interval.

4 Introduction of an Explanation Component

A well-crafted explanation can convey understanding not only of how a recommendation was arrived at by the system but also of the basic way in which the system works and of how much faith the user should place in it (cf. Jameson [17], section 15.7).

It is not a priori obvious what content should be presented in an explanation of a recommendation based on decision-theoretic planning or in what form the content should be presented. A particular challenge with this type of recommendation is that an individual recommendation may be understandable only as part of a larger recommendation policy.

In order not to rely entirely on our own intuitions, we followed the strategy of Herlocker et al. [16], generating a variety of different types of content and forms of representation and presenting them for feedback to a group of potential users (see Bohnenberger [12], section 7.4, for details). Although these subjects (33 students in a computer science lecture on *Intelligent Environments*) were not representative of the entire population of potential users and were not using the recommendations while performing a real shopping task, they did provide some interpretable feedback on the naturalness and perceived value of various elements of explanations.

4.1 Initial Implementation

In order to see how explanations are used and evaluated in a natural setting, we prepared a relatively simple initial implementation, making use of ideas that had emerged from the questionnaire study.

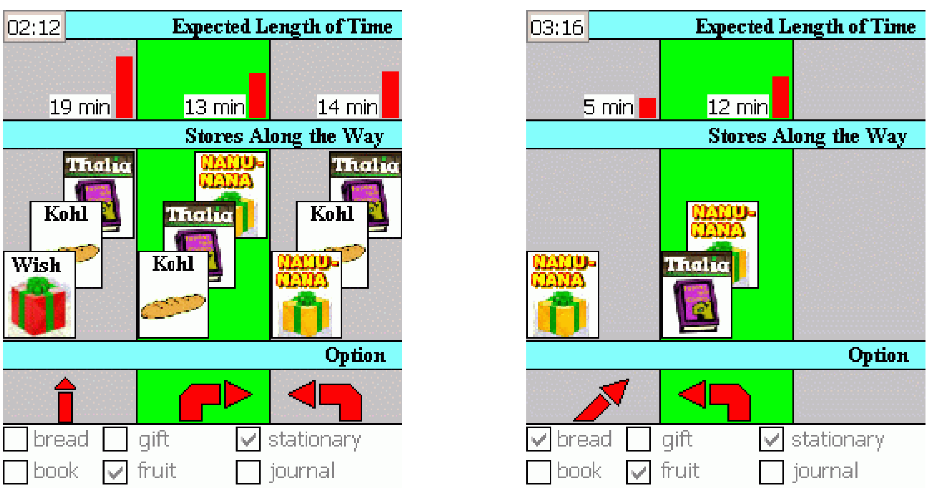

Fig. 4. Screen shots from the initial implementation of the explanation mechanism. (Each store is represented by a rectangle that shows the name of the store and the type of product that it sells. For example, Wish and Nanu-Nana are two alternative gift stores.)

What Is Presented to the User. Figure 4 shows two examples of explanations. The first one might be offered relatively early in the shopping trip used for the study in the mall, when the user has so far found just 2 of the 6 desired products. The (at most 3) walking options with the highest expected utility are always shown, the recommended one being placed in the middle of the screen and highlighted with a more salient background. The information provided about each option concerns two attributes that were judged especially important in the questionnaire study and that were fairly straightforward to implement: the expected length of time until completion of the shopping task if the option in question is followed and the next three (at most) stores that the user can expect to encounter—in both cases, on the (optimistic) assumption that the user will find each product in the next possible store and will proceed with the expected speed. In this example, the user can recognize that the option on the right is quite similar to the recommended option, apparently involving only a different order of visiting the first three stores and a slightly longer expected duration. The user might choose the right-hand option if for some reason she wanted to go to the gift store before the other two stores. The second explanation shown in Figure 4 would be offered a bit later on, after the user had found the bread. Note that these displays "explain" the system's recommendations only in a superficial way, by showing the properties of alternative options in such a way that the user may be able to see why one of them is probably preferable. More generally, we will see shortly (Section 5.1) that the term *explanation* may not be the most apt characterization of the supplementary information provided by our guide—or by other recommender systems.

Necessary Computations. The type of explanation just discussed is so simple that the explanations can be computed on-line on the PDA. The current implementation computes an explanation each time the user arrives at a beacon that corresponds to a location at which the user could walk in more than one direction. For each of the available options, the system simulates the application of the recommendation policy for a user who goes in the direction in question, assuming that every action is successful. For more sophisticated explanations, which took into account the inherent uncertainty involved in these actions, more sophisticated computations would be required (see Bohnenberger [12], section 7.4.3, for a discussion of some possibilities).

Deciding How to Make an Explanation Available. Aside from the question of how explanations should look, a system that can offer explanations must address the question of when the explanations should be presented to the user. A relatively simple policy was implemented for the new prototype:

– When the difference between the overall expected utility of the recommended option is only slightly greater than that of the second best option, the user must tap a button labeled "Explanation" to see the explanation.
– When the difference is of moderate size, the system displays on the normal screen a text explicitly urging the user to consider tapping on this button.
– When the difference is large, the explanation is presented immediately instead of the simpler display with a single animated arrow.

The rationale is that it is especially important for the user to see an explanation when the recommended alternative is much better than the alternatives, in case the user should be considering deviating from the recommendation. Although this rationale is not necessarily the best one, it seemed well suited to eliciting feedback in the study.

4.2 Follow-Up Study in the Shopping Mall

Method. Five subjects were recruited, two of whom had participated in the first study in the shopping mall. The method was the same as for the first study, except for the following additions:

1. When introducing the system, the experimenter demonstrated how an explanation could be requested (and how it might appear spontaneously) and he made sure that the subject understood the information presented in an explanation.
2. The questionnaire presented at the end included several new questions about the explanation mechanism, and the experimenter followed up on them during the debriefing.

Results. Despite the limited number of subjects, the answers given are sufficiently consistent and interpretable to yield a fairly clear picture.

Regarding the overall question of whether the subject found the explanations helpful, 3 subjects answered positively and 2 negatively. The main objection (mentioned even by one of the subjects with a positive attitude) was that under heavy time pressure, a user is not inclined to look at relatively detailed new information but would rather simply trust the system's recommendations.

One question concerned the circumstances under which explanations should be offered spontaneously: when the difference between the recommended and the second-best alternative was especially large or especially small. Of the 5 subjects, 4 chose the former alternative, in effect endorsing the rationale explained above. But one of the same subjects also noted the potential utility of an explanation even in the opposite case: "If the difference [in the predicted remaining duration] is only one minute, the user can decide himself whether he wants to take the longer route, maybe because he considers the fruit more important than the book."

One subject liked the explanations simply because they displayed information about the recommended option that went beyond what was usually displayed (i.e., the predicted remaining duration of the shopping trip and the next three upcoming stores).

Discussion. Despite the roughly even split between positive and negative overall judgments, the following summary statements (which take into account individual comments not mentioned above) seem to reflect the general consensus among the 5 subjects:

1. The appeal of explanations is relatively low in situations involving high time pressure, in which users tend to prefer simple displays that can be taken in at a glance (or better yet, with ambient vision, as is possible with the animated arrows). In addition to the obvious reason that reading explanations consumes scarce time, another reason for this preference may be the subjects' recognition (which is expressed in other comments) that the shopping guide's recommendations are relatively hard to improve on when

time pressure is involved: Given time pressure, the system is exploiting not only its knowledge of the products available in various locations but also its ability to adapt its recommendations continually to the approaching deadline—a task that humans are not especially good at.

2. The subjects do perceive several potential benefits of explanations, but these benefits are due only in part to the explanations' function of clarifying the reasons for the system's recommendations: An explanation can also help simply by conveying additional information about the recommended option and/or the available alternatives, thereby allowing the user to make a well-founded decision to deviate from the system's recommendation and/or to prepare mentally for what lies ahead.

3. Regarding the unrequested presentation of explanations, the strategy implemented in the current prototype is perceived as being basically reasonable, but some subjects would like to exert more control over the presentation strategy (for example, reducing or eliminating spontaneous explanations in the case of extreme time pressure).

5 Conclusions

5.1 Conclusions Concerning the Shopping Guide

The fact that even the follow-up study with the explanation component yielded significant ideas about further improvement of the shopping guide illustrates that the optimal design of a system of this type is not something that can be determined once and for all. Each test involving a new version and/or a new context may yield further ideas for improvement. But this fact is not surprising, since it also applies even to widely accepted software tools like calendar applications.

Our detailed discussion of results calling for improvements should not obscure the overall result that the majority of the subjects found the shopping guide to be an attractive tool for the support of shopping in some types of circumstances, characterized mainly by the unfamiliarity of the shopping environment and the existence of some form of time pressure. Both the basic functionality of the system and many of the specific interface features were found to be well adapted to this type of shopping task.

Aside from some improvements that can be realized straightforwardly, the necessary improvements that came to light concern mostly the need of users for certain types of additional information, ranging from general background information about the basic capabilities and limitations of the system to information that supports particular types of thinking during the performance of the shopping task (for example, deciding whether to deviate from a recommendation of the system). The results yield a good deal of guidance as to how this information should be presented; they also show that the presentation must be selective and subject to control by the user, since users are highly sensitive to the presentation of unnecessary information and since the appropriate amount depends on individual preferences and on situational factors such as time pressure.

With regard to information that conveys a basic understanding of how the system works, one approach currently being explored (see, e.g., Kröner et al. [18]) is to create a special retrospective mode in which the user can interact with the system outside of the natural environment of use. For example, at home at the end of the day, the system might walk the user through some of the day's events, presenting reminders of what happened and explanations of its actions. The idea is that when the user is free of time pressure and attention-consuming environmental events, she will be better able to build up a mental model of how the system works.

With regard to additional information about specific user options and system actions, the main challenge appears to be that of ensuring that the timing and mode of presentation are well adapted to the user and the situation. A good solution will probably involve some combination of (a) specification of long-term preferences by the user, (b) requests by the user for specific information during use, and (c) automatic situational adaptation by the system.

5.2 A More General Lesson

One general theme illustrated by this research concerns a fundamental tension that arises with systems that are employed while the user is interacting in a rich environment. On the one hand, users do not in general want to receive and process much information from a system of this type, preferring instead messages that can be perceived with minimal distraction from their interaction with the environment. On the other hand, users at least sometimes want to use their own knowledge and understanding to second-guess and override the information and advice provided by the system; and doing so will sometimes require getting more information from the system than the user needs in order to take the system's outputs at face value. Finding an acceptable resolution of this tension may be one of the trickiest challenges both for designers of pervasive computing systems and for their users.

References

1. Bohnenberger, T., Jameson, A.: When policies are better than plans: Decision-theoretic planning of recommendation sequences. In Lester, J., ed.: IUI 2001: International Conference on Intelligent User Interfaces. ACM, New York (2001) 21–24
2. Bohnenberger, T., Jameson, A., Krüger, A., Butz, A.: Location-aware shopping assistance: Evaluation of a decision-theoretic approach. In: Proceedings of the Fourth International Symposium on Human-Computer Interaction with Mobile Devices, Pisa (2002) 155–169
3. Asthana, A., Cracatts, M., Krzyzanowski, P.: An indoor wireless system for personalized shopping assistance. In: Proceedings of the First IEEE Workshop on Mobile Computing Systems and Applications. (1994)
4. Fano, A.E.: Shopper's Eye: Using location-based filtering for a shopping agent in the physical world. In: Proceedings of the Second International Conference on Autonomous Agents. (1998) 416–421
5. Shekar, S., Nair, P., Helal, A.: iGrocer: A ubiquitous and pervasive smart grocery shopping system. In: Proceedings of the 2003 ACM Symposium on Applied Computing. (2003) 645–652
6. Cumby, C., Fano, A., Ghani, R., Krema, M.: Building intelligent shopping assistants using individual consumer models. In Riedl, J., Jameson, A., Billsus, D., Lau, T., eds.: IUI 2005: International Conference on Intelligent User Interfaces. ACM, New York (2005) 323–325
7. Newcomb, E., Pashley, T., Stasko, J.: Mobile computing in the retail arena. In Terveen, L., Wixon, D., Comstock, E., Sasse, A., eds.: Human Factors in Computing Systems: CHI 2003 Conference Proceedings. ACM, New York (2003) 337–344

8. Kim, J., LaRose, R.: Interactive e-commerce: Promoting consumer efficiency or impulsivity? Journal of Computer-Mediated Communication **10** (2004)
9. Plutowski, M.: MDP solver for a class of location-based decisioning tasks. In: Proceedings of the First Bayesian Modeling Applications Workshop at the Nineteenth Conference on Uncertainty in Artificial Intelligence, Acapulco, Mexico (2003)
10. Brafman, R.I., Heckerman, D., Shani, G.: An MDP-based recommender system. Journal of Machine Learning Research **6** (2005)
11. Shani, G., Brafman, R.I., Heckerman, D.: An MDP-based recommender system. In Darwiche, A., Friedman, N., eds.: Uncertainty in Artificial Intelligence: Proceedings of the Eighteenth Conference. Morgan Kaufmann, San Francisco (2002) 453–460
12. Bohnenberger, T.: Decision-Theoretic Planning for User-Adaptive Systems: Dealing With Multiple Goals and Resource Limitations. AKA, Berlin (2005) Dissertation version available from http://w5.cs.uni-sb.de/∼bohne/.
13. Boutilier, C., Dean, T., Hanks, S.: Decision-theoretic planning: Structural assumptions and computational leverage. Journal of Artificial Intelligence Research **11** (1999) 1–94
14. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach. 2nd edn. Prentice-Hall, Englewood Cliffs, NJ (2003)
15. Landauer, T.K.: Behavioral research methods in human-computer interaction. In Helander, M., Landauer, T.K., Prabhu, P.V., eds.: Handbook of Human-Computer Interaction. North-Holland, Amsterdam (1997) 203–227
16. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 Conference on Computer-Supported Cooperative Work, Philadelphia, PA (2000) 241–250
17. Jameson, A.: Adaptive interfaces and agents. In Jacko, J.A., Sears, A., eds.: Human-Computer Interaction Handbook. Erlbaum, Mahwah, NJ (2003) 305–330
18. Kröner, A., Baldes, S., Jameson, A., Bauer, M.: Using an extended episodic memory within a mobile companion. In: Proceedings of the Workshop on "Memory and Sharing of Experience" at Pervasive 2004, Vienna (2004)