

A Field Evaluation of the User Experience of a Mobile Recommender of Leisure Activities

Ellen Isaacs, Victoria Bellotti, Alan Walendowski, Nicolas Duchenaut & Mark Newman

Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304
magitti@parc.com

ABSTRACT

There have been many context-aware, personalized mobile recommender systems for leisure activities, but the user experience of such devices has received scant attention compared to feature descriptions. This paper reviews findings from existing evaluations and provides detailed results from a field evaluation of Magitti, a similar system offering recommendations of leisure activities for locals and tourists. It focuses on issues that are likely to crop up for designers of such systems, such as user control and management of recommendations, sharing the experience with others, map design, the role of content, user interface design issues, and other aspects of the user experience.

Author Keywords

Recommender systems, leisure guides, tourist guides, mobile computing, user experience, user studies.

ACM Classification Keywords

H.5.2. [Information Interfaces and Presentation]: User Interfaces – *Evaluation/methodology*

INTRODUCTION

People have long relied on paper-based guides to help them find places to go and things to do during their leisure time. With the growth of GPS systems and mobile devices, there has been a great deal of interest within the research community in supporting leisure and tourist activities with mobile electronic guides (see [4] and [17] for surveys of such systems). The tourism and leisure guide domain is a natural fit for these technologies because they allow users to walk around with a small device and receive information about their environment based on their current location. Some of these systems also take into account the user's interests, either provided explicitly by the user or inferred from their activity or through collaborative filtering. Most provide information on tourist attractions, and some also offer information about shopping, eating out, news, and events [3, 4, 17].

Although many context-aware mobile guide systems have been built, only a few have been evaluated from the standpoint of the user's experience [6, 10, 9, 14, 18]. Most of these evaluations have measured user satisfaction, and

they all find that users generally have positive experiences with the systems and rate them as useful tools to varying degrees. Just a handful, however, report on the details of the user experience or discuss how the design and feature choices affected the usability and usefulness of the system.

Guide systems for which published reports on user evaluations are available include Guide [6], Gulliver's Genie [10], CRUMPET [14, 15, 16], COMPASS [18], and Dynamic Tour Guide [9]. All these systems offer information about tourist attractions and sometimes nearby eateries that can be personalized to the user's interest, and some offer audio and video clips. All include maps and navigational aids, some with recommended routes to the sites. All of them were tested under real world conditions around a city or college campus, with the exception of COMPASS, which was tested via a simulation.

Based on this small amount of data, a few patterns begin to emerge about the use of context-sensitive mobile guides. Perhaps the only common result, not surprising in research prototypes, is that users don't like waiting for slow systems. More interesting is the frequent mention of issues with maps. Users reported wanting more details on the maps, more frequent updates of their position on the map, the ability to match the map's orientation with their own, the ability to pan and zoom, and text directions. O'Grady et al. [10] report that users' sense of direction affects their satisfaction with the system navigation design, which may account for the wide range of issues.

There are some hints that users may be wary of systems being too smart in anticipating their interests and being too automated. Users of COMPASS didn't like the system deciding that they were less likely to visit a restaurant shortly after they'd visited it [18], and users of CRUMPET were wary when they didn't understand why certain items were being recommended [14]. They also rated the idea of a "pro-active" guidance mode low. Gulliver's Genie users wanted more control over the system's decisions about what to download. In addition, the content of the system has an important influence on user's trust in the system, as small omissions or errors in content reduced user's satisfaction with the system [6, 10]. Users also found systems were less useful if they were missing certain

categories of information, such as transportation, events, and restaurants [10].

Among some of the other findings are that people get overwhelmed if there appears to be too much information or too many features [6], and that users sometimes have trouble using a stylus to control the device. [10].

This paper provides an evaluation of the design and user experience of another context-aware mobile recommender system named Magitti [5]. Magitti is different from the tour guides described in that it focuses on suggesting leisure activities not just for tourists, but mainly for residents who want to find new places to go and things to do. Magitti is designed to help people choose among many potential options rather than guiding people through a small set of attractions. Like other systems, however, Magitti makes recommendations based on the user's context – location, time of day, weather – as well as their stated and inferred preferences and prior behavior. Like MobyRec [12], it also bases recommendations in part on users' similarity with others. Magitti also provides supporting information about the places recommended, but rather than providing cultural and historical information about attractions, it has reviews written by professional critics as well as by other users.

Although there are differences among these systems, we believe much can be learned by studying Magitti that can be applied not just to electronic travel or leisure guides, but to other mobile information or recommender systems where the content changes based on the users' location. In this paper, we offer a rich set of observational, self-report, and objective information about people's use of the system in the hopes of illuminating some of the design implications for building such systems.

The following section describes Magitti's design and features. We then explain how we conducted the study and report the findings. Finally, we offer guidelines to those interested in designing such systems.

MAGITTI

Magitti was designed based on extensive field studies of young people in Tokyo who are interested in finding new places to share time with friends [13]. Based on the field studies, we determined that Magitti needed to run on a small, portable device that could be operated with one hand. We also learned that Magitti should allow users to spontaneously discover ideas for activities while out and about without much effort, although it also needs to allow people to plan ahead. As people complained about being overloaded with information, it should help people cull through the abundance of leisure information to find the most relevant choices. The most common leisure activities among young Japanese were eating out and shopping so we focused mainly on those domains, although we also included some events and activities.

User Interface

Magitti's Main Screen is shown in Figure 1 (left). It shows a scrollable list of up to 20 recommended items that match the user's current context and profile [5]. As users walk around, the list automatically updates to show recommendations relevant to their new location. From the main screen, users can see the name and address of a place, its rating (number of stars), and the beginning of a description. To get more information, they go to the Detail Page (Figure 1, right), which shows more of the description, the beginning of a professional review and user comments, and options to view more things to do before or after. It also lets them give their own rating of the place. The user can view additional screens to see the complete description, review, or set of user comments.

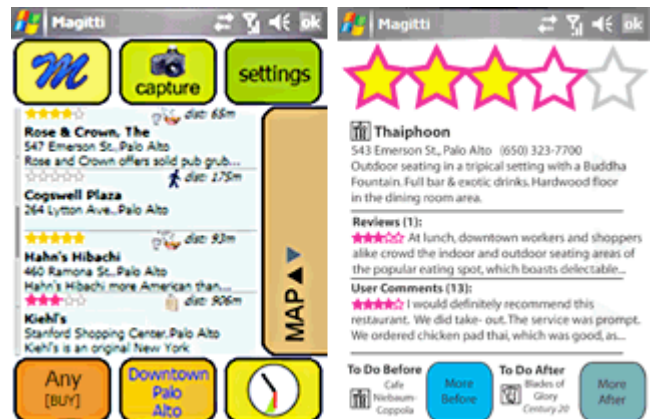


Figure 1. Magitti's Main Screen (left) and Detail Screen (right).

Having identified places of interest, users can open a map to see the location of the four recommendations visible in the list. They can view the map halfway while still seeing the list (hybrid view), as shown in Figure 2 (left), or they can slide it out all the way. The little figure on the screen represents the user's position, and the numbers on the map match up with those in the list. Users can pan the map by dragging it, and zoom using the + and – buttons on the map tab. As users move, the map updates to track their position.

Without entering any information about their interests, users automatically get recommendations. The system offers a mixture of recommendations for four main categories: restaurants (Eat), stores (Buy), events (See), and activities (Do), based on its learning about the user's likely interests at this time and place. In addition, the user can tune the recommendations in a few ways. They can filter the list to show venues for just one type of activity (e.g. Buy). They can specify their preferences for types of cuisines, shops, activities, and for attributes such as liveliness, price range, smoking, etc, as shown in Figure 2 (right). Also, whenever they rate a place their rating influences future recommendations, as Magitti compares their opinion to others' ratings and weights more heavily places that similar users like.



Figure 2. Main screen with map in hybrid view (left), and a preferences screen for rating types of cuisines (right).

Ordinarily, Magitti recommends activities for the current time and place, but the user can plan ahead by specifying a different time and/or location. They can also set a distance range for the recommendations. Finally, users can bookmark items and they can enter a keyword search, requesting the best matches or places that are nearby.

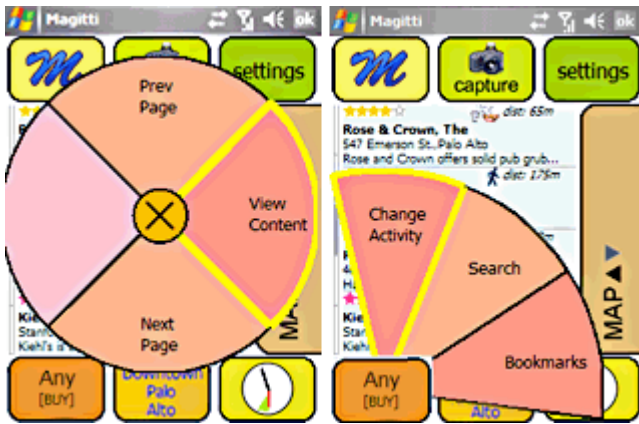


Figure 3. Marking menus for a recommended item in the list (left) and for the Activity button (right).

The large buttons on the interface reflect the fact that Magitti is designed to be held in one hand using the thumb to operate the controls via a touch screen. Most items on the screen have a marking menu [8] that appears when the user presses and holds with their thumb. Figure 3 shows a menu for an item in the list (left) and for the activity button (right). The user issues a command by sweeping their thumb to the desired menu option. Once the user learns the locations associated with each command, they don't need to wait for the menu to appear; they can simply sweep their thumb in that direction.

Magitti runs on Mio A701 Digiwalker smartphones running Windows Mobile 5.0. It has a built-in GPS and a resistive touch screen, which does not allow multi-touch gestures and is not as sensitive as the capacitive touch screens that have since become more common.

METHOD

Our goal was to get a rich understanding of the Magitti user experience, focusing on its usefulness, its ability to suggest appealing recommendations, and its usability.

Eleven people (seven men and four women) went out with Magitti in the Palo Alto, California area between one and four times each over several days. The subject pool was company employees and interns not involved in the project. Subjects ranged in age from late 20s to early 50s with a mean of 37. They ranged in their familiarity of the area, from new residents to old timers.

Before using Magitti, each person was interviewed about how they spend their leisure time and how they get recommendations. After receiving a short demonstration of Magitti, they were asked to go out with the device at least three times. Each person arranged their own outings, deciding where and when to go and whether to invite others. One outing per participant was shadowed by the first author, who took notes about their behavior but did not interfere except to troubleshoot. After each outing, participants filled out a Web-based questionnaire about their experience with Magitti. After completing all their outings, each user was interviewed again about their overall impressions. We also collected map traces of all Magitti outings and logged all user actions with the device.

We seeded Magitti with about 1050 places of interest, primarily in the Eat (497) and Buy (279) categories, with some items to See (158), and Do (122).

RESULTS

Participants went on a total of 32 outings (2.9 per person), visiting 60 places for an average of 1.9 places per visit. They visited 30 places to Eat, 27 to Buy, and 3 to Do. Sixteen (half) of the outings were accompanied by family members, a spouse, or a friend. Participants had an additional 23 sessions during which they requested recommendations for another time or place, set preferences, rated items, or explored its functions.

Overall Usefulness

Our main question was whether Magitti was useful in enabling people to discover new places to go during their leisure time with minimal effort. In the initial interviews, people said they tended to go to the same standbys because it was difficult to find good recommendations of new places. While using Magitti, however, they were successful at discovering new places, as 53% of the places they visited were new to them, and 69% of outings included at least one new place. The following characterizes the participants' familiarity with the places they visited:

- I had never heard of it: 38%
- I had heard of it but never been: 15%
- I have been there once or twice: 25%
- I have been there many times: 23%

Also, in 69% of the outings, they reported noticing at least one other new place that they wanted to visit in the future.

Participants gave Magitti good ratings on various measures of usefulness. When asked how helpful Magitti was during their outing, they rated it an average of 4.1 on a 1-5 scale (5=very helpful). When asked the degree to which Magitti enhanced or detracted from their outing, the average rating was 3.9 (1=detracted a lot, 5=enhanced it a lot).

In the final interviews, we asked people to indicate the degree to which they would use and trust Magitti compared with other sources of recommendations. They reported that they would both use and trust it more than professional reviews, user reviews, and advertisements, but less than their personal experience or recommendations by friends and family. This result is consistent with our field research, in which people reported trusting friends and family most.

We received reports of and observed many enthusiastic responses to Magitti. One appreciative user said in his survey, *"I love to have it with me. It adds fun to the outing and lets me explore more alternatives which I normally did not even think of, like a store or restaurant that I am not familiar with."* Another wrote, *"It gave me recommendations for places that I would never have gone to. It's especially nice to get a broad overview of how many places have what you're looking for, even if you've been to the shopping center before. I liked seeing them on the map relative to my position."* Another personified Magitti when he said, *"I saw the Mediterranean cafe and wasn't sure if I would like it or not and Magitti convinced me to try it."*

Our favorite story was of a woman who had been trying to find a park that she used to visit as a child but had never encountered as an adult in her nine years living in the area. She used Magitti to find a place to run and happened to choose that park. In her survey she wrote, *"It was a WONDERFUL trip down memory lane. THANK YOU!!! It really was magic."*

Not all the experiences were positive, however. The biggest frustration was caused by the same problem encountered by others, namely a slow and unreliable wireless network. We made several changes to improve performance where we could, but nonetheless slow responsiveness detracted from everyone's experience.

There were a variety of other factors that reduced Magitti's effectiveness. Perhaps the most interesting issue we uncovered was that people found it challenging to smoothly incorporate Magitti into an outing with another person. Other problems related to the effort required to get useful recommendations, in particular difficulties in culling the list to the most relevant and appealing options, and in operating the device with a thumb. We also saw some problems in looking down at a small handheld device rather than looking at their surroundings. As one user put it, *"Really, Magitti both enhanced and detracted from this outing. The outing was enhanced by the recommendations of places I didn't know about, but distracted by the time I spent fiddling with and looking at the device."*

In the following sections we characterize in detail these patterns of use and focus especially on the issues that affected the user experience.

Styles of Use

Watching people using Magitti, we observed three styles of use that can be characterized as the "What's Close?" "What's Best?" and "What's Good Enough?" strategies. Those who used the "What's Close?" style focused primarily on the map, often using the full-size map as their main interface rather than the list. These users tapped on the closest places to see the details, mainly looking at the descriptions and only occasionally checking a user comment or review. They generally stopped looking as soon as they saw one nearby that was appealing. For this group, it would have been helpful to provide a map-based view that listed *all* the closest places instead of showing just the four visible in the list. Indeed, many other tour guide systems are designed around a map rather than a list.

Those who used the "What's Best?" style wanted to find the best possible choice from all the options. They were quite thorough in their approach, looking through the entire list of recommendations, usually several times, and mentally narrowing it down. For shops, they then planned their route to visit as many as possible; for restaurants, they continued to narrow the list to just one. This group tended to look at many Detail Pages, reviews, and/or user comments, and they used the map toward the end to inform the decision or to plot the route. One person tried to remove items from the list by giving them low ratings. (This behavior was added later.) For this group, it would have been helpful to have a temporary hold feature that would let them mark the better choices and then see their "short list." They could then remove items from the short list until they were down to the one or few they wanted to visit. Another approach would be to let users remove items from the list temporarily so they could cull it down to just the contenders as they worked their way through it.

The "What's Good Enough?" style involved perusing part of the list, sampling Detail Pages of more interesting options, and then selecting one that was good enough, without necessarily reviewing all the choices. These users tended to view user comments and reviews along the way. Once they'd made a choice, they opened the map to find its location and plan a route. This style was served fairly well by the design as it was, although these users too might have made use of a temporary remove feature to help them keep track of the current candidates.

Control

One of the most common reactions was to desire more control over the type of recommendations listed. As one user explained, *"It's very nice that it recommends things without you having to do anything, but sometimes you want to ask for specific things."*

Most often, people filtered the list based on their activity. They changed activity type an average of 5.1 times per

outing, mostly to Eat (1.8 times per outing) and Buy (1.4). Although the system started up in “Any mode” (offering recommendations from across categories), they rarely switched back to this (0.7 times per outing). This supports the suggestion from other systems that people are skeptical of a system that is too automated. It also highlights the wisdom of making it easy to adjust incorrect inferences, one of the design goals of Magitti.

Sometimes people further wanted to specify a type of venue (e.g. shoe store, Thai food), particularly for shops. This was not possible to do at the time of testing, although it was added later. The only approximation available was to change one’s preferences to rate certain types of places higher and others lower. Consequently, people spent a lot of time adjusting their preferences each time they wanted to find a different type of store or restaurant. Doing so was cumbersome, as it was designed to be done once and adjusted only occasionally. As one user commented, “*I feel like saying what I want is so hard.*”

Other filters would also have been useful. We saw evidence of people mentally filtering the list by price (“*that sounds expensive*”) or by rating (“*only two stars*”). Most people at some point wanted to filter the list by distance, but that feature was not available. However, they could set the distance range for recommendations and six people did so, expecting to see only items within that range. As designed, the system prioritized closer recommendations but still tried to provide the usual 20 recommendations by supplementing the list with places a little further away. This turned out to be a poor design choice, as people were clearly trying to narrow their choices and wanted a shorter list. This finding supports prior observations that people get overwhelmed by too much information [6, 13].

People sometimes asked for the ability to manually remove items from the list, at least temporarily. If they had decided they weren’t in the mood for Mexican, for example, they didn’t want to keep seeing a certain Mexican restaurant in the list, even though they may have rated it highly earlier.

Finally, the location-based aspects of the system caused a problem in managing the list. As people moved around, the list updated to show new recommendations. Sometimes people had trouble returning to an item they had been interested in, since by then it had moved in the list. They didn’t always know in advance that they’d want to return to an item, so they wouldn’t have bookmarked or “short listed” it. A “recently viewed” list would help in cases when they viewed the item’s Detail Page, but not if they simply noted it in the list. In general, the unpredictability of the list and the locations of items in it caused some puzzlement and at times frustration.

Trust and Transparency

As Magitti’s goal was to offer useful recommendations, we asked people to rate after each outing how often the recommendations were “relevant and of interest” from 1=rarely to 5=almost always. They rated it an average of

3.8, a little less than “usually.” As one person put it, “*Most of the time, the list contained a mix of useful and not so useful recommendations.*”

Given that Magitti had only a few outings in which to learn from people’s behavior, that we had only a small set of ratings for collaborative filtering, and that not all of the intelligence was fully online at the time of the study, we took this as a good start. However, during the shadow sessions we noted factors that negatively affected people’s confidence in the recommendations. The most common and powerful effect occurred when someone walked near a place they liked and didn’t see it in the list, or at least in the top screen or two. When this happened, people either concluded that our content was incomplete (“*the list did not represent what downtown has to offer*”) or, more often, that the system didn’t know their tastes well enough. This echoes the findings of others who found that small omissions or inaccuracies reduce people’s trust in the system [6, 15]. Others have also noted this phenomenon about recommender systems in general [7].

As designed, Magitti used distance from the user as just one of several factors affecting recommendations – it sometimes showed a “better” recommendation down the block higher in the list than a “lesser” one next door. The thought was that people would want to choose from the best options near them, but instead they seemed to expect that the closest places would always be at the top of the list.

Similarly, people remarked upon recommendations that were too far away. Even if an item was close as the crow flies, if it was outside a comfortable walking distance they considered it a poor recommendation. This situation didn’t seem to have as strong an effect on people’s confidence overall, it was simply noted as strange.

Although people wanted distance to be a strong if not the primary factor in making recommendations, doing so could lead to other problems. The See and Do venues for activities were much further apart than those for Eat and Buy. So when users who had limited the distance setting asked for See or Do recommendations, they got few or no recommendations, which again led them to think the system content was incomplete. These findings highlight a need for a recommender to be sensitive to the density of venues, taking into account whether a user is browsing along the street or searching for the best place to go, and whether they are in a car or taxi. COMPASS zooms out its map if the user is traveling quickly, but its authors do not mention filtering content based on speed [18].

We observed that people put more weight on the first item recommended than we expected. If it was reasonable, they considered the recommendations good, even if some of the others were inappropriate. One person searched for “home furnishings,” and Magitti listed Williams-Sonoma first (a furnishings chain store) and McDonald’s second (likely because of a hit on “home” in a user comment). Still, the user discounted the second item because the first one was

on target. Conversely, if the first item didn't make sense, people were more skeptical of the rest of the list.

Surprisingly, people lost relatively less confidence when Magitti recommended a place that turned out to be closed, even though this should not have happened. (Our data about the hours of operation were occasionally incorrect.) They were disappointed but they didn't seem to judge Magitti as harshly, apparently because they didn't expect it to take hours of operation into account. Also, although more may have noticed, only one person complained that Magitti continued to recommend the place they had just visited. Guide's authors also noted the desirability of moving recently visited places lower on the list, although COMPASS users said they didn't want this behavior.

Of course, the worst case was when Magitti simply got it wrong. One person went to lunch with a friend, choosing a Vietnamese Pho restaurant. After lunch, they happened upon a nicer looking Pho place. It didn't appear in the list at all, but they found it by searching. Sure enough, it was rated four stars to the other one's three stars. His companion griped, *"It would have been good if Magitti had let us know that there was another similar but nicer restaurant right around the corner."*

It appears that many users were treating Magitti as an information guide rather than a recommender, expecting it to offer information about the closest places but not expecting it to know about hours of operation, to balance distance with ratings, or to personalize recommendations. A few, however, did indeed expect it to recommend the best place for them. Many users spent a fair amount of time speculating on how Magitti decided which activities and venues to list. Magitti used a complex set of algorithms that took into account many factors (location, time of day, preferences, similar users' opinions, prior behavior), and the lack of transparency of the algorithm – or even the user model – sometimes confused or even frustrated users. Others have also noted that trust issues arise when users don't understand the model for recommendations [7, 15].

Social Issues

Half the Magitti outings involved one or more companions, and four of those were shadowed by the experimenter. Social use of the small handheld device imposed some challenges. Users tended to walk with their head down staring at the screen. Of the eight who used it with a companion, five reported that their companion got annoyed at some point because they spent a lot of time looking down and not at their surroundings or companion. Some said that even when their partner wasn't bothered, they felt anxious about ignoring the other person.

One Magitti user said, *"it's rude to be looking at it a lot. You should look at it only when you need to make a transition."* He felt relieved when his wife was drawn into a store's product demonstration, saying *"This is good because she'll get really irritated if I fiddle with this too much."* Yet he said he enjoyed using it by himself. *"I*

definitely like having it when I'm alone because it gives you a sense of companionship." In fact, once while alone he used Magitti as an excuse to leave a restaurant after he'd sat down. He noticed another place in the list that looked better, so he pretended to be getting a call on the device, gesturing to it as the waiter saw him get up and leave.

Another person who used it with his wife said she felt *"like I was leading the whole thing and she was left out."* The issue of social control came up again when his wife adjusted his preference settings, hoping to see more recommendations that suited her tastes. Since he knew he would be returning the device shortly he didn't mind, but he said if it had been his own, he would have objected. This suggests that systems like this might need to be able to accommodate multiple profiles.

That same person also went for lunch with a friend who already had strong ideas about where to go. This participant used the more thorough "What's Best?" approach, so it took some social finesse for him to find a place with Magitti while still appeasing his friend. Although he tried to encourage his friend to look at the device, the latter showed little interest. In the end, they chose a restaurant the friend noticed when they walked by it, although he first looked it up in Magitti to check its reviews. He then used Magitti to find a place to go for dessert, and the friend seemed more willing to humor him as he walked around examining two options listed in Magitti before choosing one.

Some participants tried to share the device, but it was hard to arrange it physically so they could both see the small screen comfortably and even harder to walk with it. Large height differences made sharing even more challenging. When people shared the screen, they usually stopped outside a store, scanned the list to make a choice and then put it away, which was quite different from the ongoing use common among solo users.

In some of the more successful shared outings, the participant used Magitti ahead of time to narrow down the choices and then asked the companion to help select from a few options. As one user explained in her survey, *"Luckily, I had some time before my boyfriend was meeting me to read through listings/recommendations and consider places to eat. This would be harder to do with 2 people. By the time he met me, I had a few questions for him to help decide which to go to: price, long/quick dinner, food style. We could make our decision pretty quickly once he arrived - and we usually spend 30 mins walking up and down the street looking at places we've seen a million times before, and always going to ones we know."*

One pair was particularly interesting because they signed up to use Magitti together and participated in the initial interview together. They thought of the device as shared and passed it back and forth throughout their outings, although the one who was *"better with gadgets"* held it more often. The person holding the device generally read options aloud along with highlights of summaries, reviews,

or user comments as the other person commented or asked questions. Using it in this style, they seemed share the experience more than other participants with companions. Interestingly, they both said they preferred having the other person hold the device because it allowed them to look around while the other person read the information aloud.

Even though most travel or leisure guides are likely to be used with more than one person in a party, none of the other leisure system's evaluations reported on coordination issues among multiple people. We note that in our evaluation there was only one device per group. If instead each participant had a system running on his or her own mobile phone, the tension around sharing the screen would be eased, but the social dynamic would be altered in other ways. For example, everyone might become focused on their screens and be even less aware of their environment and each other. They also would have to coordinate their decision given the recommendations each was viewing separately. There is clearly more to be studied regarding shared experiences supported by mobile devices.

Device Issues

As mentioned, people often walked with their heads down concentrating on the device, which caused a few problems. First, it became tiring to use after a while, especially in bright sun. People often cupped their hands over the device or walked to the shade or inside doorways to reduce glare. Second, it meant that people were not aware of their environment. While walking around an obstacle on the sidewalk onto the street, one person said, *"The danger is that I'll just stare at it and not look at the road."*

Finally, looking down at the device kept people from noticing the very places they were looking for. One participant who went out with his friend repeatedly missed places that were easily visible, even after the friend pointed them out. The friend noted, *"This is like some sort of cartoon! He's so busy looking down that he doesn't see what's in front of him."* Another person complained, *"I had to stare at it all the time. I was missing the whole visual experience and I'm a visual person."* Again, no other system evaluation discussed problems related to looking down at a device while walking around.

Two people suggested offering an audio interface that would read aloud the names of the places and allow the user to request further information, thus allowing them to look around. Other systems use audio and video to provide background information about an attraction [10] or to give directions [9], but not to help users choose among options. Sotto Voce [2], a system meant for indoor use in a museum, provided audio clips of objects in the rooms. Its users were facile at engaging others in their audio or eavesdropping on others, and they seemed to have a high tolerance for multiple audio streams playing at once. It would be useful to know if the same would be true when a system is used in a noisy outdoor public place to choose an activity rather than to learn about items on a tour. To this point, it's

interesting that Gulliver's Genie users complained that the audio of that outdoor system was sometimes too quiet [10].

Content

The consensus was that Magitti was most useful for finding restaurants and events and less so for shopping. Users explained that when you're hungry, you're often open to many cuisine types and atmospheres, but when you're shopping, you focus on stores selling the things you want to buy. (Although people sometimes make impulse purchases, seeing the names of other types of stores in the list didn't intrigue them in the same way as spotting an interesting product in a shop window.) Restaurant reviews are helpful because you can't simply look in a restaurant window to tell if you'll like the food, whereas when you're shopping you can wander in and experience the goods without committing; browsing through stores is part of the leisure experience. Instead of store reviews, participants wanted information about the products sold (e.g., a directory, sales, photos, product reviews). Also, they said that frequently looking down at Magitti detracted from the experience of window shopping.

When asked which piece of information they relied on to select a place to go, participants said they relied most on the name (3.3, where 1=not at all and 5= very much so), followed by the summary description (2.7), then the rating (2.6), then user comments (2.5), and finally the formal reviews (2.1). The log data support this ordering in terms of number of views. We observed during the shadow sessions, however, that slow response times reduced people's willingness to look at review and user comments screens, so people may be more inclined to use them with a more responsive system.

These relatively low average ratings reflect the large variability in opinions of these information sources. Some people relied heavily on the user comments and one even considered them *"one of the best features of Magitti,"* whereas a few others didn't trust them at all. One person called them *"a lot of drivel,"* saying they too often talked about the personal lives of the people writing the reviews. Many thought the formal reviews were too long to read unless they were planning for a future outing, but a few used them almost exclusively. Most thought the star ratings were important, but a few ignored them. One said, *"often people will rate things too high just because a place is fancy,"* which he didn't care about. We conclude that, for a system like this, it's important to provide a range of supporting information about the recommendations so that people can find the type of information they trust.

Finally, we were reminded that no matter how well designed or feature-rich a device like this is, "content is king." We didn't have complete coverage of the area, and, as mentioned, people quickly noticed missing favorites, and then trusted other recommendations less as a result. In a few cases, people were misled by a store-provided description or a user comment and blamed Magitti if they

later found that the information was wrong. One woman dismissed one store as too expensive because it characterized itself as an “upscale shoe store.” She later happened upon it and found its prices were reasonable. She said, “*I wouldn't have come here based on the description.*”

Use of the Map

There was near universal enthusiasm for the map, although as others have found, people varied in their needs and style of use. As noted, the “What’s Close?” users used the map as their primary visualization of the recommendations, and the rest went to the map after making a selection. The data showed that people used the map an average of 9.7 times per outing, and had a slight preference for the hybrid position. One user explained, “*I really like the half-and-half mode. It's just enough to be usable,*” meaning that you could see just enough map and just enough information about the items in the list.

Eight of our eleven participants said they considered themselves good at reading maps, and good map skills were shown by others to improve users’ experience with a mobile map device [10], so our users’ success with the map may be unrepresentative. Still, we saw a few cases where people turned the device to orient the map to their physical location, which has been reported by others [10, 18]. An option to track orientation might be useful.

We saw only a few cases where people made use of being able to see multiple items on the map, and in those cases they didn’t want to see just the four visible on the screen but some subset of all 20 items in the list. It would probably be more useful to show several items on the map once people can sort or arrange the list to show only the options of interest in the ways discussed earlier.

We saw an interesting design challenge in fitting multiple items on the map as well as the user’s location. Initially, Magitti zoomed out the map to show all four locations visible on the screen plus the user, but sometimes this meant zooming out too far to be useful, so users immediately zoomed in. Later we modified the map to show only those recommendations that fit within a useful zoom setting, which meant that some items sometimes weren’t visible or got cut off. We saw one case where someone assumed his restaurant was on the map but then discovered he had been tracking the wrong number when he got closer, as his restaurant had been off the map when he’d started. It would help to have an indicator pointing to items that are off the screen.

Finally, people who were not that familiar with the area sometimes couldn’t tell which way to walk even after locating themselves and their destination on the map. So they walked enough to see which way the figure on the map went and then corrected from there. It would have been useful to have an arrow pointing in the direction of the recommended item. We also found that the user’s location on the screen sometimes did not update frequently enough, as have others [10, 14].

Users of Gulliver’s Guide also pointed to the map and navigational support as one of its strongest features, and still most of their suggestions for improvement centered around the map. Their findings and ours suggest the critical importance of maps to guide systems and the need to design them for users with different levels of map-reading skills.

Offline Use and Planning

It was common for people to use Magitti “offline” (i.e. from a fixed location) to plan a future outing, set preferences, rate items, or otherwise explore its content or functions. The logs showed an average of 2.1 such offline sessions per person in addition to the 2.9 outings. Users reported specifically planning their outings an average of 1.5 times, sometimes on their way to the area or even in the parking lot before setting out (which would have been treated as part of the outing in the logs). All but one participant used the device offline at least once, and the most did so six times for just two outings.

The planning feature was seen as important to users. “*The concept is great and allows planning ahead while I was sitting down having lunch.*” Another said, “*It was cool in a way because sometimes we say, ‘where do you want to eat’ and we’ll both say, ‘I don’t know.’ This way we can ask ‘What about this?’ ‘What about that?’*” Users of CRUMPET also found the planning function to be especially important [14].

We thought users might be more likely to read the formal reviews when offline, since they tend to be long, but this was not the case. They read an average of 1.6 reviews per offline session vs. 2.6 per outing, and they read user comments for an average of 2.5 places during offline sessions vs. 3.1 per outing. This result may be because these offline sessions were shorter than the outings and people used them to do some of the other activities, such as setting preferences and rating items, and not just to plan later outings.

Searching

People used the search mechanism relatively rarely, averaging only 1.2 searches per outing. Again, the variability among users was high. Only one user made extensive use of the search feature, conducting 18 searches in three outings. The rest did so between 0 and 6 times, with three people never using it and another three trying it once.

We know they avoided searching in part because it required pulling out the stylus and carefully tapping on a tiny keyboard to enter text. One man who went out shopping with his wife said to her, “*I’m going to enter ‘home furnishings’ into this for you. That’s how much I love you.*” However, the low number of searches also suggests that the recommendations were generally good enough that searching wasn’t necessary.

User Interface

Overall, the participants learned the layout of the screens quickly and became facile at controlling the system. The

relatively small and manageable number of features made it easy for people to learn what was available and where to find them. Although we have suggested adding a variety of functionality in this paper, we caution against overloading systems with so many features that users have a harder time learning what is available and making use of the features. Again, Guide's users commented that having too many choices made it harder to use [6].

Although the screen layout was easy to use, the thumb-based interface was difficult for some to master. Four of the 11 became comfortable using their thumb to issue commands, but the remaining seven wound up holding the device in one hand and using either their index finger or the stylus to operate it, at least part of the time. The devices used in the study contained a touch screen technology designed for stylus input that is not as sensitive to human touch as the capacitive touch screens that have since become more common. Interestingly, users of Gulliver's Guide complained about the difficulty of using the stylus to control that system and even suggested the idea of one-handed operation [10].

For most users, the marking menus worked well. All but two people learned to scroll up and down and move forward and back through the screens without waiting for the menus to appear, as these were the most common gestures. Over four days of use, only two people memorized the positions of the menu items on the Main Screen buttons, as these commands were issued less frequently.

CONCLUSION

We have provided a detailed account of the many user experience issues we uncovered when evaluating Magitti's use in the field in the hopes of informing the design of future such systems. However, in describing these usability issues in such detail, we do not mean to leave the impression that Magitti was not well received. On the whole, most people were delighted that it helped them discover new places with relatively little effort and they greatly enjoyed using it. As one person put it, "*More than anything, it was fun to use.*"

The people who had lived in the area a long time were most pleased with Magitti, as it opened up new experiences for them. One long-time resident expressed this well in saying, "*I think it makes life more interesting. It allows you to get out of your daily routine. Almost as if you're going to a different city.*" Another participant was reluctant to give Magitti up at the end of the study, saying, "*It's over! I'll miss this thing telling us where to go for dinner.*"

On the whole, this study showed that users appreciate the value of personalized location-based recommendations of places to go along with descriptions and reviews of those places. We close with a list of design guidelines revealed in this field evaluation.

- Although people need to feel confident that the system has complete and accurate information, at any given

moment they want to see a small set of the most highly relevant items so they don't have to wade through too many options.

- Although the temptation is to try to automate the system so much that the perfect short list of items appears automatically, people sometimes lose trust when the system is "too smart" that it reduces their control or their understanding of the options presented.
- It is equally important to support the user's own efforts to cull a list of recommendations on the fly, perhaps by making it simple to sort, filter, or "short list" the options to a set of strong candidates.
- People seem to vary the most with regard to the types and sources of information they trust (venue-provided vs. third-party descriptions, reviews by other users vs. professionals, ratings), and in their styles of use (searching for the best possible choice vs. finding something good enough). A good system needs to offer different types of information and support different representations of data to accommodate different users.
- Maps are a critical component of these systems. They need to allow users to easily and frequently track their position, and offer enough detail for users to interpret what they're seeing without being cluttered. When many items may appear on the map, designers need to make a tradeoff in showing enough places without overwhelming people. The density of nearby venues should be taken into account when determining a useful zoom level. It helps if users can pan and zoom the map and optionally track their orientation.
- When walking around, people seemed to expect to see information about items closest to them by default, even if they were not as highly recommended as others nearby. Still, information shouldn't always be restricted to just the closest items as people may often want to explore other options or view items by ratings or other attributes. The key is to help users readily interpret what types of recommendations they're seeing.
- Mobile guides for leisure activities are likely to be used with others, so the design needs to allow people to share the information and incorporate others into the experience.
- Although people like the convenience of small, lightweight mobile devices, they draw user's attention away from the setting and can get tiring to look at.
- People find restaurants and events a more natural fit for a recommender. For shopping guides, they expect to see information about products, such as photos, reviews, and discounts.
- Although it's useful to allow one-handed operation of the device, it wasn't easy for our users at least to get used to operating the device with their thumb. However, marking menus are a useful way of operating the touch screen with a minimum of clutter.

- It's important to allow people to use the system to plan ahead as well as to discover places spontaneously.

ACKNOWLEDGMENTS

We thank Dai Nippon Printing Co., Ltd. Media Technology Research Center and Corporate R&D Division for sponsoring this research. Diane Schiano helped seed some of the Magitti content. Bo Begole, Ed Chi, Mike Roberts, Kurt Partridge, Bob Price and other PARC researchers built the Magitti system prototype. We are grateful to the participants for using imperfect prototype software and giving candid feedback.

REFERENCES

1. Abowd, G., Atkeson, C., Hong, J., Long, S., Kooper, R., & Pinkerton, M. Cyberguide: a mobile context-aware tour guide. *Wirel. Netw.* 3, 5 (1997), 421-433.
2. Aoki, P. M., Grinter, R. E., Hurst, A., Szymanski, M. H., Thornton, J. D., and Woodruff, A. Sotto voce: exploring the interplay of conversation and mobile audio spaces. *Proc CHI 2002*. ACM Press, (2002) 431-438.
3. Asthana, A Cravatts, M. & Krzyzanowski, P. An indoor wireless system for personalized shopping assistance. In Cabrera, L. & Sattyanarayanan, M. (Eds). *Workshop on Mobile Computing Systems and Applications, IEEE Computer Society Press* (1994) 69-74.
4. Baus, J., Cheverst, K. and Kray, C.: A Survey of Mobile Guides. A survey of map-based mobile guides in: *Mapbased mobile services - Theories, Methods and Implementations*. Chapter 13. Springer-Verlag, Zipf, A., Meng, L. and Reichenbacher, T. (Eds). 2005.
5. Bellotti, V., Begole, J., Chi, E., Ducheneaut, N. Fang, J. Isaacs, E., King, T., Newman, M., Partridge, K., Price, B., Rasmussen, P., Roberts, M., Schiano, D. & Walendowski, A., Activity-based serendipitous recommendations with the Magitti mobile leisure guide, *Proc. CHI*, ACM Press (2008), 1157-1166.
6. Cheverst, K., Davies, N., Mitchell, K., Friday, A. & Efstratiou, C. Developing a context-aware electronic tourist guide: Some issues and experiences, *Proc. CHI 2000*, ACM Press (2000) 17-24.
7. Herlocker, J, Konstan, J., Terveen, L., and Riedl, J. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22 (2004), ACM Press, 5-53.
8. Kurtenbach, G., Sellen, A. & Buxton, W. An empirical evaluation of some articulatory and cognitive aspects of "marking menus". *Human Computer Interaction*, 8(1), (1993). 1-23.
9. Modsching, M., Dramer, R., ten Hagen, K. & Gretzel, U. Effectiveness of mobile recommender systems for tourist destinations: A user evaluation. *Interaction Challenges for Intelligent Assistants, AAAI Spring Symposium*, (2004).
10. O'Grady, M., O'Hare, G. & Sas, C. Mobile agents for mobile tourists: A user evaluation of Gulliver's Genie. *Interacting with Computers* 17 (2005) 343-366.
11. Parhi, P., Karlson, A., & Bederson, B. Target size study for one-handed thumb use on small touchscreen devices. *Proc. MobileHCI*, ACM Press, (2006) 203-210.
12. Ricci, F. & Nguyen, Q. N. Acquiring and revising preferences in a critique-based mobile recommender system. *IEEE Intelligent Systems* 22, 3 (2007), 22-29.
13. Schiano, D. J., Elliott, A., Bellotti, V. Tokyo youth at leisure: towards the design of new media to support leisure planning and practice. *CHI '06 Extended Abstracts*. ACM Press (2006) 309-314.
14. Schmidt-Belz, B., Laamanen, H., Poslad, S. & Zipf, A., Location-based mobile tourist services – First user experiences. *Proc. ENTER 2003*. International conference on tourism and communication technologies, Helsinki, Springer Computer Science, Heidelberg, (2003a).
15. Schmidt-Belz, B., Aspects of user trust in mobile guides. *Proc. Mobile HCI*. Udine, Italy (2003b).
16. Schmidt-Belz, B. and Poslad, S. User validation of a mobile tourism service. *Proc. Mobile HCI*. Udine, Italy (2003c).
17. Schwinger, W., Grün, Ch., Pröll, B., Retschitzegger, W., & Schauerhuber, A. Context-awareness in mobile tourism guides – A comprehensive survey. Technical Report, Johannes Kepler University Linz, IFS/TK (2002). [<ftp://ftp.ifs.uni-linz.ac.at/pub/publications/2005/0405.pdf>].
18. van Setten, M., Pokraev, S. & Koolwaaij, J. Context-aware recommendations in the mobile tourist application COMPASS, *Adaptive Hypermedia and Adaptive Web-Based Systems*, LNCS 3137, Springer, 2004, pp. 235-244..