"Let There Be Light!"
Comparing Interfaces for Homes of the Future

Barry Brumitt & JJ Cadiz

September 21st, 2000

Technical Report
MSR-TR-2000-92

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

"Let There Be Light!" Comparing Interfaces for Homes of the Future

Barry Brumitt & JJ Cadiz
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
+1 425 936 3263
{barry; jjcadiz}@microsoft.com

ABSTRACT
As smart devices proliferate in the home and become universally networked, users will experience an increase in both the number of the devices they can control and the available interaction modalities for controlling them. Researchers have explored the possibilities of speech- and gesture- based interfaces in desktop scenarios, but it is also necessary to explore these technologies in the context of an intelligent environment, one in which the user's focus moves off of the desktop and into a physically disparate set of networked smart devices. We report results from a lab study where six families tried various traditional and non-traditional interfaces for controlling lights in the home. Results indicate that speech is the preferred interface for controlling lights. More importantly, the study indicates location awareness and gaze tracking are vital when building a usable intelligent environment system for the home, even though users did not perceive added value from these technologies.

Keywords
Home automation, multimodal interfaces, speech interfaces, gaze, intelligent environments, mutual disambiguation

1 INTRODUCTION

There is an ongoing trend towards increasing amounts of technology in the home. The PC, in particular, has moved from something only owned by the enthusiast to a device found in over 50% of US homes. Beyond PCs, "smart" devices such as security cameras, remotely controllable lighting, and mobile phones are also found with increasing prevalence in the home. With the advent of technologies such as Bluetooth, as well as proprietary RF and powerline standards, these smart devices are becoming increasingly connected and, as this trend continues, will be able to behave in a coordinated fashion. We see the first signs of this trend in specialized devices, such as an MP3 player that can broadcast to the stereo in another room over phone lines.

While current devices tend to perform only standalone functions (like the MP3 player or security cameras mentioned above), the increasing connectivity of devices should allow more complex interactions. For example, the lighting in the home is likely to be controllable either from the PC, special-purpose remotes, or any generic input device capable of displaying a simple UI. A camera used for infant monitoring might be able to direct its output to whatever display is most conveniently located for the parent. This is very different from the current paradigm where each input device can talk to a small fixed set of devices in the home. An environment that has this kind of awareness of its users and an ability to maintain a consistent coherent interaction with the user across a number of heterogeneous smart devices is called an Intelligent Environment [2]. This class of computing has also been referred to as "Ubiquitous Computing." [20]

Since Intelligent Environments typically contain many diverse input devices, off-the-desktop multimodal interfaces are a natural result. In this domain, multimodal interfaces refer not only to using more than one modality simultaneously for a given task, but additionally to enabling the selection of any of multiple modalities for the task. Intelligent environment research systems [1,2,21] have used diverse interface modalities (including speech, gesture, GUIs, remote controls etc.), but research in this area has tended to be focused on what is technologically feasible instead of on which modalities users prefer. Intelligent Environment researchers haven't explored what makes a usable home interface.

Unfortunately, desktop PC usability is not necessarily the same as home intelligent environment usability. Interface designers cannot assume that all users will have the same interaction affordances, i.e. a keyboard, mouse, and screen. In one room, a touch screen might be available, but only a speech interface might be present in another. Furthermore, while the PC and its peripherals are implicitly co-located (on the same desk, for example), in a home environment the devices are located throughout the home. This implies a system that receives the command "Show me the weather" must determine on which device that information should be shown instead of merely displaying the information on the display attached to the PC performing the speech recognition.

One of the barriers to finding out what makes a usable home system is that conducting valid lab and field studies is

difficult. Most usability labs weren't built to create a home environment, and field studies with fragile technology in the home can be complex. Furthermore, creating technology robust enough to test can be difficult given the integrated systems built by intelligent environment researchers.

Thus, this research addresses the issue of usability in the home via a low-cost "Wizard of Oz" lab study of how users might control lights in a home of the future. Given the substantial effort required to build intelligent environment systems, our hope is that these user data guide researchers toward developing systems that have the most potential to be usable and desirable. In particular, indications of which interaction modalities are most important to usability can drive the selection and development of costly and complex perception systems, such as those that understand speech, interpret physical gestures or determine objects' locations.

In the following section, we discuss previous research on the home environment. In section 3, we introduce our focus on lights and describe different light control interfaces. Section 4 discusses the methodology of our study while section 5 presents our results. We discuss how our findings may direct home automation research in section 6.

2 RELATED WORK

Some previous research exists on interfaces in the home, but its focus has been on specific tasks for home PCs instead of integrated intelligent environments [15,16]. More applicable to this research are several ethnographic field studies of technology use in the home. These studies have focused on the effects of Internet usage at home [4,5,6], the use of television set-top-boxes in the home [12], the use of tablet PCs in the home environment [9], and the home environment in general [10]. Venkatesh [18] complements this work with a theoretical framework of household-technology interaction, based on research beginning in the 1980's of technology adoption in the home.

These studies outline several concepts to keep in mind when designing technology for the home. Most important is the idea that "the implicit design assumptions of the personal computer are inappropriate for the home" [10]. The environments and tasks in the home are different from what is found in the workplace, as are the relationships. For example, Venkatesh [18] outlines several environments in the home where a variety of tasks take place, ranging from food preparation to watching movies.

In addition, the social dynamics of the home are quite different than other environments. While PCs at work typically only have one user, home computer systems may have several users, not all of whom have the same level of access. For example, when studying the use of set-top-boxes by families in England, O'Brien [12] noted that parents wanted protect their kids from mature content. With home computing environments, we should not assume that all users would be treated equally.

The different environment of the home also brings up the issue of form factor. Studies of tablet PC use in the home found that the top reason users enjoyed the tablet PC was that it was portable [9]. Ethnography research reported in [10] also supports the idea of using smaller, portable devices in the home. They write, "The data imply that ubiquitous computing, in the form of small, integrated computational appliances supporting multiple collocated users throughout the home, is a more appropriate domestic technology than the monolithic PC." These studies suggest that people may find integrated intelligent environments to be quite useful for their homes.

However, little research exists to tell us exactly what kind of interfaces would be best suited for these integrated, intelligent, home environments. This research complements previous work by presenting a lab study of different potential interfaces for intelligent environments in the home.

3 FOCUSING ON LIGHTS

Lights are devices that are universally well understood and have been a typical target of today's nascent home automation systems. Their simplicity and ubiquity drove the selection of light control as a focus for this study. Though many interaction modalities have been proposed for intelligent environments, there has been no explicit comparison of these alternatives for a single task. This paper reports on a lab study of light control usability in an intelligent environment.

Currently, lights are typically turned on via a switch located on the wall or on the light itself. However, with the addition of speech and vision systems, other options become viable. A person might talk or make gestures to control lights. Small screens with access to the position of the user and all lights in the room can provide dynamic, context-sensitive interfaces. Following are some non-traditional mechanisms for controlling the lighting in a room, along with descriptions of each:

Plain text computer display: Using any display and point-and-click technology, a list of lights can be displayed. Using this list, users may control lights via slider bars or similar widgets. However, one major issue with any such display is creating consistent, clear labels to describe individual lights.

Graphical computer display: If enough knowledge of the room's physical layout is available, the problem with creating consistent labels for lights can be resolved by showing the user a map of the room with all the lights in it. Furthermore, knowledge of the location of the display and the position of the user can be used to orient the map appropriately.

Voice only: If the room can hear and understand speech, users can control lights by making voice commands. However, this method suffers from the same problem as the

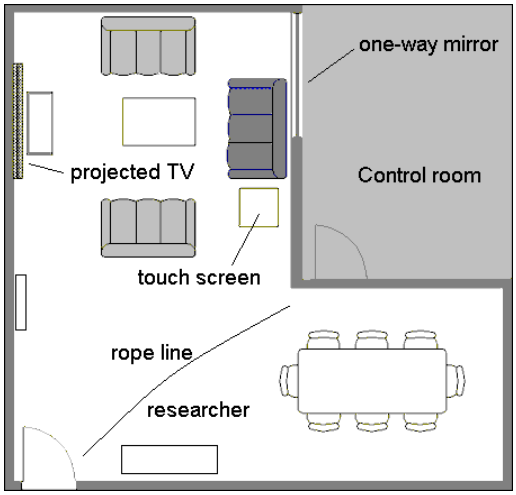

Figure 1: Layout of the lab where the study was conducted.

plain text computer display: in what way should users refer to individual lights?

Voice with location: One way people can refer to individual lights without using a specialized vocabulary is with commands such as, "Turn on the light to my left." For these commands to work, systems must be imbued with knowledge of where the user is located. A vision system [7] or an active badge system [19] could be used to gain this information.

Voice with location and gestures: Vision systems can also be used to recognize gestures [3]. With gesture recognition, users can successfully use multimodal commands to control the lights: for example, saying "Turn on that light", while pointing at a particular light.

Automatic behaviors: Perhaps the best interface is one where the user doesn't have to make any commands at all. Neural net systems may be able to infer the appropriate lighting based on the actions of the user [11]. For example, if a person enters a room or sits on a particular chair, then the lights should go on to provide appropriate illumination.

4 METHOD

To test these various methods of controlling lights in the home, we recruited six families, since they are typically the "users" of home environments. Families consisted of one parent and two children or two parents and one child, and no family members had any experience with a home automation system. A total of 10 parents and 8 kids participated in this study, and all family members received a free software product for their time. Parents ranged in age from 39 to 57 while children ranged in age from 10 to 18.

Although families were recruited for this study, participants did the tasks for the experiment individually. For each participant, the study had two phases: working with lights in the EasyLiving lab, and ranking the light interfaces using the card sort task.

4.1 Working with Lights in the EasyLiving Lab

For the first phase, we used the Microsoft Research EasyLiving lab. (The EasyLiving project at Microsoft Research [1] is concerned with architectures for intelligent environments and has built various methods for users to interact with devices in the space.) The lab is a mock-up of a small home living room (Figure 1). One portion of the lab (the lower right, in Figure 1) is set up as a conference room, which was roped off and not used for this study. The remaining part of the lab consists of an entertainment center, three couches, a center table, and some shelves. A television screen is projected on the front wall from behind the one-way mirror.

Participants manipulated the lights in the room using each of the six methods discussed in the previous section, with the exception of "automatic behavior", as it's difficult to "control" lights with automatic interfaces. Users controlled a total of 14 lights for this study. All lights were of the small spotlight type and were mounted in the ceiling in locations shown in Figure 1. The two computer display interfaces were made available (one at a time) via a touch screen that was placed on a table in the room.

The methods of light control that required speech and/or gesture recognition were conducted using the "Wizard of Oz" technique: although participants were told that the computer was responding to their commands, a researcher sitting behind the one-way mirror controlled the lights.

To do the tasks for this phase of the study, participants started by the door to the lab (the lower left portion of Figure 1). A researcher sat behind the rope line to give the participant instructions.

4.1.1 Study Scenario

Participants were told that they had agreed to help a friend set up her basement for a surprise birthday party. Participants were also told that the friend recently moved into a new intelligent home that had automated systems for controlling several things, including the lights.

To set up the basement, participants were told that they needed to do two tasks, and that they would do these two tasks several times, each time using a different method of controlling the lights. First, they needed to turn on all the lights in the basement. Second, they needed to adjust the lights to create a spotlight effect on the center coffee table, where the birthday cake and presents would be placed. The spotlight effect was to be achieved by dimming the lights around the edge of the table while leaving the light above the coffee table turned on to full brightness.

4.1.2 Conditions

Participants did the two tasks for the scenario several times. First, they were told nothing about the room except that it was "intelligent." This task was used to see what people expected of a futuristic intelligent home. When participants started this task, the touch screen was placed in screen saver

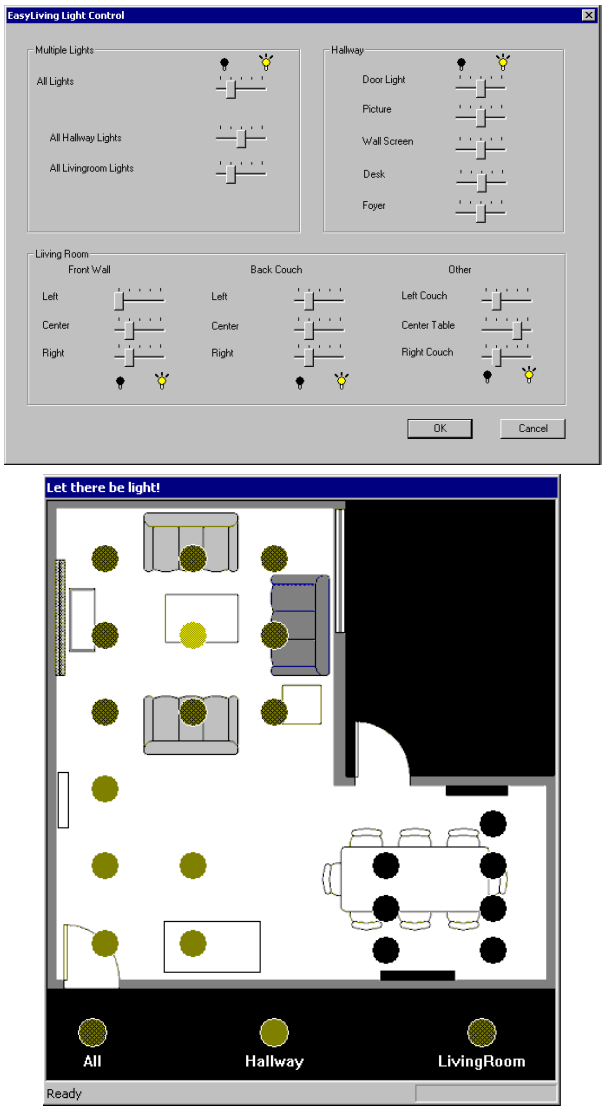

Figure 2: The two touch screen interfaces used for this study. On the top is the text interface, on the bottom is the graphical interface. In the bottom display, tapping lights would cycle them through 0%, 25%, 50%, 75%, and 100% brightness. Participants could not control the lights in the lower-right portion of the room (the conference room).

mode and displayed nothing so that it would not be an obvious avenue to complete the task.

Second, participants did the tasks for all of the methods listed in section 3 (except the method they used for the very first condition). Ordering of the conditions was balanced such that half of the participants used the touch screen interfaces first while the other half used the speech interfaces first. For the touch screen interfaces, participants were shown the screen with one of the interfaces and asked to do the tasks. The two touch screen interfaces are shown in Figure 2. For the speech interfaces, participants were told to do the tasks with the knowledge that the computer could hear them. If participants did not spontaneously gesture or refer to their location, participants were next shown the three Triclops cameras [17] they were "computer eyeballs." Participants were then asked to do the tasks with the knowledge that the computer could see and hear. If participants did not gesture spontaneously, they were told that the computer could understand "gestures" (researchers were careful not to say "pointing") and asked to try the tasks again. After using each method of controlling the lights, participants indicated on a 5-point Likert scale how much they liked the method and how easy they thought the method was to use.

For the last condition, participants were told they could use any of the methods to control the lights. This condition was used to probe participants' preferences after they knew everything the room could do.

After completing all of the above, participants were shown how automatic lighting features could work in homes. They were asked to sit down on a couch. When they did, the lights above them turned on. When they got up and left the area, the lights turned off. In addition, if participants did not use gestures or location information for the speech tasks, they were shown how one could control lights using that information.

4.2 Ranking the Light Control Interfaces

After completing the first phase of the study in the lab, participants were shown to an office outside the lab for phase two, where they were asked to do a card sorting task. In this task, eight methods of controlling the lights were presented on index cards:

Normal switches: lights are controlled using the switches that you typically find in homes today

Touch lights: lights are turned on and off by touching any part of them

The clapper: lights are turned on and off by clapping loudly

Computer wall display: lights are turned on or off by using a computer panel that hangs on the wall

[*] *Computer tablet:* lights are controlled using a computer tablet that can be moved around the room.

Speech: lights are controlled by talking to them

Speech + gesture: lights are controlled by talking to them and pointing at them

Automatic sensing: lights are turned on and off automatically by a computer that intelligently guesses how you would like the lights to be set.

[*] In the original design of the experiment, participants were to use a computer tablet to control the lights for one task. However, when running the study, we removed this condition without removing the card, thus we ignore the rankings for the "computer tablet" condition in our analyses.

	I liked this interface 1 = hated it 5 = loved it	This interface was easy to use 1 = very difficult, 5 = very easy
Touch screen – graphical interface	4.0	4.0
Touch screen – text interface	3.0	3.5
Speech	5.0	5.0
Speech with location	4.0	4.0
Speech with gesture	5.0	5.0

Table 1: Median scores (higher is better) of people's reactions to each method of controlling the lights.

Participants were first asked to sort the cards in order of their overall preference for the methods Participants were then asked to do the same for each of the following rooms in their home: living room, dining room, kitchen, bathroom, bedroom, and entertainment room.

5 RESULTS

For the first condition when participants were told nothing about the room except that it was "intelligent", we were convinced that a substantial number of people would immediately try to talk to the room to control the lights. However, only one person used speech. Of the remaining sixteen participants, twelve discovered and used the touch screen and four were stopped after trying unsuccessfully for three minutes. With the exception of the one participant who talked to the room, all participants wandered around the room looking for a device that looked like it might control the lights.

5.1 Ratings of Liking and Ease of Use

After using each method of controlling the lights, participants rated on a 5-point Likert scale how much they liked the interface and how easy they thought the interface was to use. The results of these data are shown in Table 1.

Mann-Whitney U tests were used to test for differences in these data. Overall, the only significant differences were found when comparing some of the speech interfaces to the touch panel interfaces. On the measure of liking, the speech interface was rated significantly higher than the touch screen graphical interface ($z = 1.97$, $p = .049$) and the text interface ($z = 2.7$, $p = .006$). The speech with gesture interface was also rated significantly higher than the touch screen text interface ($z = 2.8$, $p = .006$). On the measure of ease of use, the touch screen text interface was rated significantly lower than both the speech interface ($z = 2.1$, $p = .04$) and the speech with gesture interface ($z = 2.8$, $p = .005$). These data indicate that speech was the preferred interface.

5.2 Choosing a Method for the Final Condition

For the last condition when participants could use any method of controlling the lights they desired, fifteen people used speech and only two people used the touch screen, indicating that speech was the preferred interface. Of the fifteen people who used speech, eight used only their voice while seven used a combination of their voice and some type of gesture. Additionally, nine participants used speech vocabulary that assumed the system possessed knowledge about the location of things in the room, e.g. "Turn on the lights around the center table" or "Dim the ones behind the couch over here." Both people who used the touch screen used the graphical interface.

	Median rank (n = 18)	Difference using Sign test
Speech + Gesture	2	
Speech	2	p = 1.0
Automatic Sensing	3	p = .008
Computer wall display	4	p = .815
Touch lights	6	p = .031
Normal switches	7	p = .031
Clapper	8	p = .238

Table 2: Median rankings (lower is better) of the light control interfaces from the card sort task. The right column shows the statistical difference between rankings.

5.3 Rankings of Light Control Interfaces

We analyzed the card sort data where participants ranked the light control interfaces in two ways: first, we looked at how participants ranked the interfaces for the "overall" condition; second, we looked at whether participants ranked the interfaces differently for the six different rooms.

5.3.1 Overall Rankings of the Light Interfaces

Table 2 shows the median rankings of the light interfaces along with the statistical differences between rankings using the Sign test. Overall, these data indicate that speech interfaces were preferred. The speech and speech + gesture methods were ranked equivalently, while the speech method was ranked significantly higher than the automatic sensing method ($p = .008$). Automatic sensing was not ranked significantly higher than the computer wall display, but the computer wall display was ranked significantly higher than touch lights ($p = .031$).

5.3.2 Rankings of the Light Interfaces for Different Rooms

After providing an overall ranking of the light control interfaces, participants ranked the interfaces for different rooms in the home. These data should be approached cautiously since participants answered these questions hypothetically: they did not actually use each of the light control methods in the different rooms.

Friedman tests were used to determine if participants ranked each of the interfaces differently for the different rooms. Of the seven different light control methods, only two were ranked significantly different for the different rooms. Speech was ranked significantly different, ($X^2(5,$

	Mean rank for normal switches	Mean rank for speech
Living room	4.00	3.61
Dining room	3.56	3.42
Kitchen	3.33	2.86
Bathroom	2.53	4.50
Bedroom	3.50	3.17
Entertainment room	4.08	3.44

Table 3: The mean ranks (lower is better) of normal switches and speech across different rooms in the home. Ranks for the other methods of controlling lights did not differ significantly across rooms.

N=18) = 15.05; Kendall's W = .167; p = .01), and normal switches were ranked significantly different, (X^2(5, N=18) = 17.26; Kendall's W = .192; p = .004). Table 3 shows the mean ranks for these two methods. Normal switches were ranked highest for the bathroom, which is also where speech was ranked the lowest. Speech was ranked the highest for the kitchen.

Wilcoxon tests were used as follow-up tests to the Friedman tests to examine differences between pairs of rooms for each of the light control methods. For speech interfaces, the highest ranked room (kitchen) was only ranked significantly higher than the lowest ranked room, the bathroom (z = -2.69, p = .007), although it was close to being ranked significantly higher than the living room (z = -1.93, p = .054). In contrast, normal switches were ranked significantly higher for the bathroom relative to the second highest ranked room, the kitchen (z = -1.96, p = .050).

6 DISCUSSION

Before discussing the results of this study at length, we should note its shortcomings. Mainly, this research has the generalizability issues in common with other lab studies. We were measuring participants' reactions to different methods of home automation based on interacting with the room for only a few minutes. Opinions about the different methods of controlling lights might be more pronounced if we were to install each of the methods in the participants' homes for a month and then ask for their reactions. We also did not control for social effects; for example, people might not prefer speech if other human listeners were present. Furthermore, participants were giving us their initial reactions to the technology, thus a novelty effect may have influenced the data. However, first reaction data to new technology is valuable since participants began the study with very few, if any, preconceived notions about how home automation systems might work.

6.1 Which was the Preferred Interface?

The data indicate that participants prefer to use their voice to control lights. Speech interfaces had the highest initial reaction scores, they had the highest rankings in the card sort task, and when participants were given the choice of using any method of controlling the lights, they chose to use their voice.

However, although speech was the preferred method for controlling lights overall, there was some indication that speech interfaces would not be best suited for all rooms of the house. Speech was ranked highest for the kitchen, which could be explained by the amount of time people spend in the kitchen with their hands occupied or dirty. For the bathroom, speech was ranked lowest while normal switches were ranked highest, indicating that perhaps people don't perceive a need for advanced home automation systems in the bathroom. Alternatively, perhaps people don't want a computer listening or watching them in these more private rooms.

6.2 Does Adding Gesture or Location Recognition Help?

Given the research concentrating on systems that can see a person's location or recognize gestures, it's important to analyze whether, from a user's point of view, adding these systems helps. The data indicate that users don't perceive any added value in a system being able to see them. Ratings and rankings of speech interfaces vs. speech interfaces with location awareness vs. speech interfaces with gesture recognition were essentially the same.

However, we don't believe this means that homes of the future only need to be outfitted with microphones, or that vision systems research for home automation is misguided. People liked speech because it worked nearly perfectly, and it worked nearly perfectly because we had a "wizard" (another researcher) controlling the lights to make sure that they responded as perfectly as possible to user requests. But after observing how people tried to talk to lights, we believe a system would not be able to function well without the use of location-aware technology, and could function better still if gaze-tracking were available. This is primarily due to the difficulty in interpreting object references.

6.2.1 Vocabulary for Controlling Lights

When trying to control lights, people's speech commands have two elements: an action (on, off, etc.), and a reference to a particular light or group of lights. The vocabulary used for the action part of the command is quite small. When analyzing transcripts from the study, we found that just four words—"on", "off", "bright", and "dim"—were used in 81% of the 198 commands.

Unfortunately, the vocabulary used to refer to lights wasn't nearly as predictable. Of the 198 speech commands, 16% used a reference to an object to refer to a light ("the light above the couch"), 30% used an area reference to refer to a light ("all the lights in the living room"), 23% included relative terms such as "left", "right", "back", and "front", and 18% used an indirect reference ("this light", "all on this side", "over there in the corner"). All references to relative terms were given relative to the television, e.g. "the left

couch" referred to the couch to the left of a person facing the television screen.

Thus, while interpretation of the action part of the command is straightforward, interpretation of the object reference is complex.

6.2.2 Looking at Lights

There are numerous ways to address the issue of resolving which object a person is referencing. One can label all the lights, but this requires a person to learn a naming scheme for all the lights in the house. One can use vision so that people can point to lights, but pointing was not completely intuitive to people (only 2 of the 15 people whose data we could analyze pointed at lights before we told them that the computer could recognize gestures).

However, as we observed participants during the study, there was one action that everyone seemed to do when referring to lights: they looked at the light they wanted to control. In fact, in only 9% of the tasks did people never look at the light they were controlling for any of the commands they issued. In 25% of the tasks, people looked at the light during some of their commands, and in the remaining 66%, people always looked at the light they wanted to control.

Thus, for intelligent environments, computer vision systems may not be best used to add more features (such as gesture recognition). Rather, vision systems may be best used in conjunction with speech recognition systems to interpret people's commands as correctly as possible. From our observations, the best "gesture" to recognize is the place in a room where someone is looking when they utter a command. This approach has been proposed by [8] for desktop computers, and [13, 14] has studied the use of multimodal interfaces to reduce speech recognition error rates with pen-based computers. Oviatt [14] reported an improvement of as much as 41% when using multimodal architectures to recognize speech.

But using gaze in support of speech has two major problems. First, determining exactly where a person is looking is difficult unless they wear special devices, train the system to know their appearance, or position themselves carefully within the field of view of a camera. Wu & Toyama [22] have examined using vision to detect rough (rather than precise) gaze direction, thus reducing the training and field of view requirements. Second, people may stop looking at the light they want to control after they're confident that the system will do the right thing. When using a new set of light switches, people may look at the lights as they play with the switches, but after learning the switches, they most likely just flip the switch and assume the light will work as they expect. To a certain extent, we saw this behavior when people were using the touch panel displays. A follow-up study could explore this issue further.

Overall, researchers are beginning to focus more on the use of gaze with interfaces. Each of the previous two CHI conferences has devoted an entire paper session to the issue of gaze, but the focus has been on desktop computers or virtual world environments. Our data suggest that generalizing these interfaces to the intelligent environment domain may be fruitful.

6.2.3 The Importance of Location Information

In addition to gaze, there are also indications that location information is important for interpreting object references in the home.

First, there was a general trend indicating preference for the graphical UI over the text-based UI, though this difference was not statistically significant. Knowledge about the location of objects in the room is necessary to build such a UI.

Second, in the final tasks where the user could select any interaction method, eleven of fifteen subjects used location information in some way, either directly through gesture, implicitly through a reference to the location of a light, or indirectly via the graphical UI. The remaining users all used labels to refer to particular lights, e.g. "the back couch lights."

Third, as the number of controllable devices in the home grows, so will the complexity of determining labels for all of them. To refer to a given device, one can either use a unique name ("Center Living Room Light"), a unique property ("The tall halogen lamp"), a relative property ("The tallest lamp"), or an absolute or relative location reference ("The lamp to my left."). Labels and properties do not scale, as their use would require the user to specify all possible labels and properties whenever a new device is brought into the room or the location of objects in the room change significantly. All users who did not use gesture used labels to refer to particular lights, and these labels could be easily interpreted given a map of the room.

Fourth, while there are indications that gaze is a useful input modality, use of gaze necessarily requires location information to understand which devices are in the field of view of the user. Gesture recognition of any sort also requires this knowledge.

These facts support the idea that geometric knowledge of the environment is important to an intelligent environment system. Ideally, a person could bring home a new lamp, plug it in (to both power and data), specify its location (either automatically or manually), and then expect all different modalities to reflect this knowledge, providing interfaces to the new device.

7 CONCLUDING REMARKS

We have reported findings from an initial study of interfaces for the home. Data collected from this study support our intuition that speech interfaces are the preferred interface for the home for controlling lights. More importantly, we believe that the data indicates that location-

awareness and gaze tracking are important for intelligent environment systems, despite the perceived lack of value by the user. Gesture understanding, on the other hand, does not appear to be as necessary.

If a geometric model of the world is available (such as a continuously updated map) and the system is aware of the location of the user (as well as other objects in the world), then speech utterances like "to my left" and "above this couch" become feasible requests. The fact that most users, when not using gesture, referred to the lights they wanted to control in terms of their physical location implies the need for such a physical model so the expected speech requests can function. Without this model, the user is faced with the burden of specifying and maintaining the necessary labels to describe each device. It is important to explore the relative importance of obtaining location information about devices vs. people, as such information can drive the selection and design of perception systems.

In an environment with a large number of devices to control, there is a significant burden on the user to properly specify the correct device or devices needed for a given task. Gaze and speech coupled with geometric knowledge of the world provide one way of pruning down the list of all devices to those that are reasonable for a given task. Non-light-related requests for which this concept applies include, for example, "Display my email on the screen over here" or "Turn down the music in the next room."

We also believe the data indicate that computer vision technology is best used to determine people's gaze to create robust speech systems. Even coarse gaze direction could be used to increase the robustness of a verbal request. Future studies should explore this use of gaze with speech systems, especially in regard to the question of whether people stop looking at the object they want to control once their confidence in the system increases.

Future studies should also extend our preliminary data on appropriate interfaces for different rooms of the home. Substantially different tasks occur in different rooms, thus future research should determine whether this means that different interfaces should be used in different rooms, and if so, what interface best matches the activities that occur in each room. In addition, since we limited our study to controlling lights, future studies should examine whether our data generalize to other home automation systems.

ACKNOWLEDGMENTS

We would like to thank Kris Keeker from the Microsoft Usability Test Coordination Group for scheduling the families for this study. We would like to thank Steven Shafer and Mary Czerwinski for guidance on the conception and design of this study. Thanks also go to Brian Meyers, Michael Hale and Amanda Kern for their assistance in implementing the light control software. Finally, thanks to Scott Tiernan for valuable discussions regarding data analysis.

REFERENCES

1. Brumitt, B., Meyers, B., Krumm, J., Kern, A., and Shafer, S. (2000). EasyLiving: Technologies for Intelligent Environments Proceedings of the International Conference on Handheld and Ubiquitous Computing 2000 (to appear).
2. Cohen, M. (1998). Design principles for intelligent environments Proceedings of the AAAI Symposium on Intelligent Environments, 36-43.
3. Jojic, N., Brumitt, B., Meyers, B. Harris, S., Huang, T. (2000). Detection and Estimation of Pointing Gestures in Dense Disparity Maps. Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 468-475.
4. Kiesler, S., Kraut, R., Lundmark, V., Scherlis, W., and Mukhopadhyay, T. (1997). Usability, help desk calls, and residential Internet usage. Proceedings of the ACM Conference on Human Factors in Computing (CHI '97), 536-537.
5. Kraut, R., Mukhopadhyah, T., Szczypula, J., Kiesler, S., and Scherlis, W. (1998). Communication and Information: Alternative Uses of the Internet in Households. Proceedings of the ACM Conference on Human Factors in Computing (CHI '98), 368-375.
6. Kraut, R., Scherlis, W., Mukhopadhyay, T., Manning, J., and Kiesler, S. (1996). HomeNet: A Field Trial of Residential Internet Services. Proceedings of the ACM Conference on Human Factors in Computing (CHI '96), 284-291.
7. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., and Shafer, S. (2000). Multi-Camera Multi-Person Tracking for EasyLiving. IEEE Workshop on Visual Surveillance, 3-10.
8. Kuno, Y., Ishiyama, T., Nakanishi, S., and Shirai, Y. (1999). Combining Observations of Intentional and Unintentional Behaviors for Human-Computer Interaction. Proceedings of the ACM Conference on Human Factors in Computing (CHI '99), 238-245.
9. McClard, A., and Somers, P. (2000). Unleashed: Web Tablet Integration into the Home. Proceedings of the ACM Conference on Human Factors in Computing (CHI 2000), 1-8.
10. Mateas, M., Salvador, T., Scholtz, J., and Sorensen, D. (1996) Engineering Ethnography in the Home. Conference Companion for the ACM Conference on Human Factors in Computing (CHI '96 short paper), 283-284.
11. Mozer, M. The Neural Network House: An Environment that Adapts to its Inhabitants. Proceedings of the AAAI Symposium on Intelligent Environments, 110-114.

12. O'Brien, J., Rodden, T., Rouncefield, M., and Hughes, J. (1999). At Home with the Technology: An Ethnographic Study of a Set-Top-Box Trial. ACM Transactions on Computer-Human Interaction, 6(3), 282-308.
13. Oviatt, S. (1999). Mutual Disambiguation of Recognition Errors in a Multimodal Architecture. Proceedings of the ACM Conference on Human Factors in Computing (CHI '99), 576-583.
14. Oviatt, S. (2000). Taming Recognition Errors with a Multimodal Interface. Communications of the ACM. 43(9), 45-51.
15. Plaisant, C., and Shneiderman, B. (1991). Scheduling On-Off Home Control Devices. Conference Companion for the ACM Conference on Human Factors in Computing (CHI '91 short paper), 459-460.
16. Tetzlaff, L., Kim, M., and Schloss, R. (1995). Home Health Care Support. Conference Companion for the ACM Conference on Human Factors in Computing (CHI '95 demonstration), 11-12.
17. Triclops & Digiclops camera systems, by Point Grey Research http://www.ptgrey.com/
18. Venkatesh, A. (1996). Computers and Other Interactive Technologies for the Home. Communications of the ACM. 39(12), 47-54.
19. Ward, A., Jones, A., Hopper, A. (1997). A New Location Technique for the Active Office. IEEE Personal Communications, 4(5) ,42-47.
20. Weiser, M. (1993). Some Computer Science Issues in Ubiquitous Computing. Communications of the ACM. 36(7), 75-84.
21. Wren, C., Basu, S., Sparacino, F., and Pentland, A. (1999). Combining Audio and Video in Perceptive Spaces. Proceedings of the 1st International Workshop on Managing Interactions in Smart Environments, 44-55.
22. Wu, Y., and Toyama, K. (2000). Wide-Range Person- and Illumination-Insensitive Head Orientation Estimation. Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 183-188.