

THE AMERICAN VOICE AND ROBOTICS "GUARDIAN" SYSTEM:
A Case Study in User Interface Usability Evaluation

Daniel F. Wallace and Kent L. Norman
Department of Psychology

and

Catherine Plaisant
Department of Computer Science

and

The Human/Computer Interaction Laboratory
University of Maryland
College Park, MD 20742

Supported in part by a grant from Maryland Industrial Partnerships
and American Voice and Robotics

Abstract:

American Voice and Robotics (AVR), in conjunction with the Maryland Industrial Partnerships (MIPS) program, contracted the Human-Computer Interaction Laboratory (HCIL) to evaluate the user interface of AVR's "Guardian" home automation system. Among their goals for the system were: Ease of use and learning of the system (intuitiveness), Aesthetic appeal, Unintimidating, Impressive, High Functionality, Fast and accurate input of user action, Clear feedback, Forgiving to errors, and Fun to use. Researchers found that users were initially impressed with the system, but that they were not fully satisfied with it. Specific initial recommendations, theoretical and methodological concerns, procedures, results, and direction of future research will be discussed.

Introduction:

American Voice and Robotics (AVR) is a relatively young corporation specializing in home automation, entertainment, and security management systems for homes in the upper brackets. These systems are designed to relieve the homeowner of some home management tasks. The AVR systems work together with existing technologies and home management /security systems to function as a central control system for the whole home. AVR has recently developed a user interface utilizing current touchscreen technology. The Maryland Industrial Partnerships (MIPS) program, which functions as a liaison between University resources and industry, has funded the major portion of the current research on the behalf of AVR. The Human-Computer Interaction Laboratory has provided the facilities, some hardware, and the technical expertise of their faculty, staff, and graduate students.

AVR, realizing a need for Human Factors in the interface, contacted the Human-Computer Interaction Laboratory (HCIL) for assistance in the evaluation and design of their new systems. AVR desired not only to modify their current system, but to establish a set of guidelines to direct the design of future systems as well. Human Factors is a genuine priority to AVR, and among their Human Factors goals for the system are that it be:

- 1) Easy to use for all users,
- 2) Easy to learn / Easy to train to use,
- 3) Unintimidating to the novice user,
- 4) Fast enough, however, for expert use,
- 5) Consistent in format throughout the system,
- 6) Highly functional,
- 7) Accurate for user inputs,
- 8) Clear in its feedback to the user,
- 9) Forgiving to errors (easy to correct errors),
- 10) Flexible, &
- 11) Easy to maintain.

Given the obvious marketing concerns, AVR also desired that the system have the following characteristics, that it would be:

- 1) Aesthetically pleasing,
- 2) Fun to use,
- 3) Impressive, &
- 4) Unobtrusive / Unannoying.

AVR and HCIL are currently in the process of implementing the findings and recommendations from the first phase of the study, and are making preparations for the second phase of research. The AVR "Guardian" system is focused more toward the security management aspects of home automation, and is the system to which this current study is directed. Future research and evaluation will be performed on the AVR "Ultimate" system. What follows in this report is a description of the methods employed to evaluate the Guardian system, and the findings of that evaluation, which will be referred to as Phase I. This

research was not intended to be theoretically oriented, rather, the focus of this study is to determine whether or not the system fulfilled the specified goals.

The AVR "Guardian" system is primarily intended to be a master controller and user interface for other sub-system controllers, therefore, the interface must primarily provide the user with the ability to program functions, and control devices from the system screen. The system must provide not only the means for controlling devices, but must also allow for the current status of all systems to be made available to the user in a clear, understandable form at all times. It is these sorts of issues that we will be addressing in this report.

PHASE I Methods :

The study of the AVR "Guardian" system utilized three separate methods of interface usability evaluation. These methods were: 1) An expert evaluation of the system, 2) Videotaping of users on the system, and 3) User responses on the *Questionnaire for User Interface Satisfaction* (QUIS Version 6.0 AVR) (Chin, Diehl & Norman, 1987).

Initial Expert Evaluation of the System:

Initial Expert evaluation: HCIL researchers familiarized themselves with the system, and spent a number of hours on the system without direct assistance from AVR personnel. They attempted to act as novice users of the system, so as to get a relatively naive opinion of the capabilities and limitations of the system. HCIL researchers also recorded any difficulties, inconsistencies, frustrations, ambiguities in feedback or system information, and any other violations of the specified HF goals of the system. Recommendations were then made based upon stated goals, and existing human factors guidelines and principles.

There were many initial recommendations specific to the AVR system, as well as many observations that are applicable to other systems as well. Some of the recommendations made by the HCIL to AVR during the first part of Phase I are cited below along with some brief discussion of the broader issues in the design of the interface. Many of these changes were incorporated before the user videotaping and QUIS testing phase due to the consensus among the project staff that very little new information could be learned from the current, unmodified system.

Brief Description of the Evaluated AVR System:

The hardware of the system that we evaluated consisted of an IBM-AT compatible micro computer, equipped with an extended 3 megabytes of RAM, and a color Mitsubishi monitor with an Elographics touchscreen. The touchscreen was the only medium of user input to the system, and system feedback to the user. The user's menu choices were registered in accordance to the "take-off" touch-screen selection protocol as described by Potter, Weldon, & Shneiderman (1987). That is to say, that options were highlighted as the user's finger touched each selection area, but actual selection of the item was not registered until the highlighted item was "released" by the removal of the finger from the screen surface. The relative merits of the input device were not investigated, as the touchscreen interface is fundamental to the AVR product, and is not an issue of interest at this time.

Only the *user interface* of the system was evaluated by the HCIL, so as a result, all of the home management functions that were referred to in the system interface were not actually being performed by the system. No feedback other than that provided by the screen (visual display) was available to the user to indicate whether or not a given action would indeed have been performed had the system been fully operational. The aspects of other forms of feedback, auditory or physical, though included in an actual system, will not be addressed here. The visual display and logical progression of menu frames are of primary interest in this study. A sample screen layout and the menu hierarchy are presented below.

Screen Layouts:

The typical AVR screen consists of a number of touch selectable areas identified by written text, surrounded by an elongated box. The color schemes used are strictly aesthetic, and colors are not generally used to communicate color specific information. The typical screen layout in the AVR system is shown in Figure 1 below. The screen represented is the opening menu encountered upon the first action performed by the user, touching the dormant (idle) screen.

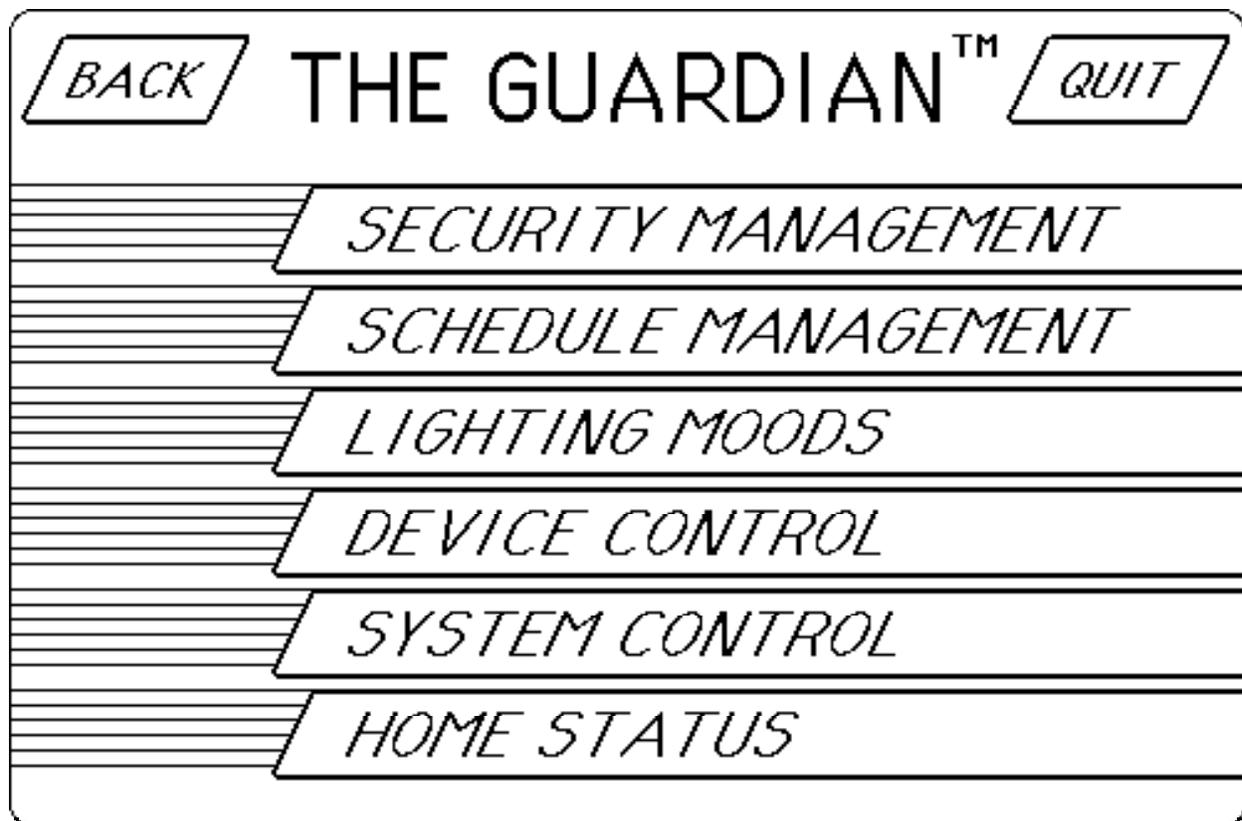


Figure 1: Typical touchscreen layout: The opening menu frame.

Menu Hierarchy:

The system's Main Menu (or opening menu frame) is called from the dormant screen by touching anywhere on the touchscreen and releasing. All menus in the AVR system are arranged in an hierarchical tree structure, and may be accessed by working systematically through the tree. The actual tree hierarchy studied is presented graphically below in Figure 2. Note that all frames also have the options "Back" and "Quit" which take the user to the Main Menu and Dormant Screen respectively. At the bottom of the tree, with the exception of the Schedule Management branch, the "Back" option must be selected to return to the Main Menu. The devices and functions that are abbreviated with two and three character symbols are various home management functions, and represent such functions as "Evening Party" (EP), and "All House Lights" (AHL). Any other significance need not be inferred by those tree elements.

General Impressions of the AVR System, Changes Recommended by the HCIL, and Relevant Issues:

The issues that are to be dealt with here are perhaps best approached in an individual, item by item manner.

Screen time-outs: The time-out period was altogether too short for many of the frames, especially for complex screens such as the house floor plans. We found that the floor plans (maps) and other complex screens disappeared before we could reasonably respond. If one should desire to move back to the dormant screen quickly, one may always use the "QUIT" option provided. The time-out didn't exist at all for some of the frames. Time-outs should have some consistent protocol for all frames. The primary reason that a time out is necessary, other than to preserve the monitor, is that in AVR systems having multiple touchscreens, input from any other touchscreen is ignored until such time as the first touchscreen quits or times out. A recommended protocol is to have a constant time-out of two minutes, at which time that station will automatically "Quit" if no input or touch is made. A warning to the user is suggested, however, of flashing the Quit button in reverse video for ten seconds prior to the time-out.

Cursor placement on touch screens: We would suggest that since the finger is most often below the visual field, given that the hand is below the field of view, the cursor which appears upon the placement of the finger on the screen should be slightly *above* the finger at all times (so as to be seen). An acceptable alternative may also be to provide a cursor that is centered under the finger, but extends "cross hairs" to provide visual feedback as to the location of the cursor.

Touch mechanism feedback: When one releases one's finger from the screen, thus selecting something, one should witness the action (selection). This could be done by briefly converting the item to reverse video, a flashing item, or the appearance of a "working..." flag. Users of this sort of system can use more system feedback (as opposed to the feedback provided by the affected changes around the home), since some system changes may not be readily apparent to the user from his or her location in the home. This is important in situations where the user may inadvertently slide his or her finger off of the item an instant before the finger is removed from the screen, thus selecting either nothing or an adjacent function. Without a "selected state" indicator, one would not be sure that an action had indeed been performed (In fairness to the AVR system, the researchers were informed subsequent to this evaluation that the system does have audio feedback capability, but that that function was disabled in the evaluated system).

Individual screens:

Dormant screen: There was a real need to indicate to the user that the main menu is coming when the dormant screen is touched. Initial touches were not immediately acknowledged by the main menu, due to the loading of main menu graphics, and repeated touches restarted the process of graphics loading. This created a situation where the impatient user could press the touchscreen repeatedly for some time before "giving up" and allowing the main menu to load.

Main menu: It was discovered that "Back" and "Quit" did the same thing. Back is typically "back to the main menu", so it should not be present in that screen. However, it might be good to keep the screen uniform across frames (it must still be consistent in taking the user to the main menu -- e.g. at the main menu the "back" key apparently does nothing to change the screen).

A greater concern to the researcher was that the word "Back" is not clear in its meaning. It may take awhile to understand that it takes the user back to the main menu. Therefore, it should be labeled "Main Menu", "Restart", "Opening Menu", or some other unambiguous label. Perhaps this issue is best resolved by the practitioner for the specific application, as to whether "Back" should mean "previous menu", "main menu", or "the most logical menu given the context".

Security management screens: These screens were somewhat difficult to use. To a large degree, the difficulties arose from the original design intent to mimic the appearance of a conventional security management control panel. This brings up a difficult issue in interface design: Should one try to emulate the "real- world" system (so as to provide a *familiar* interface) when the original interface is of a poor user interface design? If such a system is intended to replace the individual interactions with separate systems, it is advised that one revamp the entire interface design to provide a simple, clear, representation of the relevant objects and actions.

Specific Human Factors issues that were also noted while investigating, the security screens included:

- 1) Limiting the number of "apparent", or overt, menu choices made by the user results in greater satisfaction and a perception of smooth flow and transition between functions. The screen should not be removed to issue a system message when a message window would suffice.
- 2) When the floors of a home are presented in a menu, they should mimic the expected geography of a house (e.g. the basement should be at the bottom of the screen, the top floor at the top). A cross section of a house with selectable floors presented on the map screen may also be used and indeed may even be preferable.
- 3) When presenting a graphic map of regions that may be manipulated, "selectable" zones should be as large as practically possible, and these zones should be well labeled. Map symbols and icons, when used, should be distinct and clearly defined in a map legend.
- 4) When actions and functions are being performed, as much information as possible should remain visible (accessible) to the user. This is related to issue #1 above.
- 5) Where possible, allow users to perform repetitive tasks en masse (e.g. one may desire to "shunt" several security zones at once).
- 6) If "toggle" buttons are used on a screen, ensure that they are clearly identifiable as toggles and not just as the "on" function (e.g. "ARM" button may arm or disarm the system).
- 7) Ensure that the map legend is clearly marked as a legend, and not as selectable areas.

Device and system control screens: These 2 menu areas were similar in many respects, and both present a series of toggle screens; therefore, these will be addressed together.

As indicated above, toggle representation is very critical to the performance of any user interface that utilizes toggles. It is often hard to determine whether a device is on or off, let alone indicating a means for

changing the current status. A screen can easily become cluttered by many "on" and "off" words. One possible solution to the difficulties that toggles present may be to draw a realistic 3-D view of a toggle switch, showing which side is down and which side is up, indicating both the current status of the device, and the means for changing the status. For an example of such 3-D toggles, see figure 3 below.

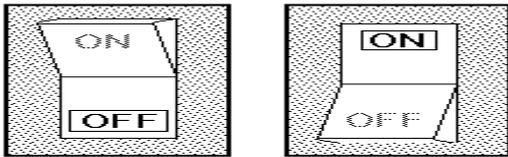


Figure 3: Sample 3-D switches

Independent schedule screens: Scheduling of devices presents a number of special problems to the researcher in interface design, as testified by anyone who has tried to program a VCR for both "on" and "off" times. Effective scheduling screens are currently under investigation by AVR and HCIL, focussing upon more graphic and directly manipulatable clocks and calenders, with error correction and multiple device and event capability.

Lighting moods screens: It is helpful to provide the user with some feedback to indicate to the user which "mood" has most recently been selected. Since a lighting mood may be easily changed by changing settings around the house directly, the "mood selected" feedback may be best indicated on that menu only until a "Quit" is selected.

Other issues: Perhaps icons can be introduced to this home management interface (associating each menu item with an icon). Although such may clutter the screen a bit, with care it will provide faster reading and selection times, and more accurate input for those persons using the system several times a day, every day.

It would be a great asset to implement a way to "undo" the most recent action performed. This function would eliminate any need to start over completely in the event that an error is made.

User Evaluation with Videotaping and the QUIS:

After the HCIL research staff provided AVR with the detailed evaluation of the existing system, HCIL and AVR personnel worked together toward implementing some of the recommended changes in the system immediately. The revised system was evaluated in two ways: 1) Subjects were videotaped during a testing session, and 2) Subjects filled out a user satisfaction questionnaire (QUIS 6.0). The primary purpose of the videotaping was to record any verbal comments that were made, as well as provision of a veridical record of any post hoc performance measures that may have proved to be of interest. The QUIS was employed as the primary measure of user satisfaction on a number of aspects of the system.

As expected, the videotaped users and the QUIS confirmed or revealed a number of potential areas of improvement with the system. These areas were identified from verbal and written comments from subjects, as well as an analysis of the frequency and types of QUIS responses. Frequency distributions for each QUIS question allowed not only visual inspection of the data, but allowed a good "feel" of the user's impressions of the system as well. Also, the statistical analysis of the QUIS allowed the researchers to make more definitive statements about the usability of the interface. A number of questions which did not relate directly to the current configuration of the system were dropped from the analysis so as to eliminate meaningless results. Several others were dropped that had an excessive number of "Not Applicable" responses. These dropped questions were specifically, numbers: 8.0-8.2, 10.0-10.2, 11.2, 15.0-16.1, 18.0-18.2, and 20.2.

Subjects: 52 students were drawn from the University of Maryland psychology subject pool. Ages ranged from 18 - 31, with a mean of 20.4 years old. Of the 52 subjects, 28 were female, and 24 were male.

Apparatus used: A revised AVR Guardian™ system (as per some of the recommended changes), was evaluated for the videotaping and QUIS portion of the study. Also used, was videotaping equipment (dual cameras, monitors, and VCR), a worksheet that presented a set of the functions and capabilities of the system in a structured format to be followed by the subjects, and the "QUIS" questionnaire version 6.0 AVR.

Procedure: Subjects were given a description of the procedures to be employed in the study, informed of the goals and purposes of the study, and were requested for informed consent to participate in the study. Two conditions were evaluated in the study. The first involved multiple subjects (2 or 3 users) working as a team, who were requested to explore the features and functions of the system. Subjects were given a worksheet to follow to assist in this exploration which provided the users with a number of tasks that they could attempt on the systems, and that would ensure that all subjects explored the same representative parts of the system (see Appendix A). The other condition involved the same sort of exploration of the system, but used only one subject at a time. Both of these conditions were evaluated for 15 minutes, and for 30 minutes of exploration time on the system.

While the subjects were using the system, their performance was also recorded on video tape with a composite image of the users and the screen image. The experimenter was available to answer any questions that the subjects may have had, however, the presence of the experimenter was as unobtrusive as possible. Upon completion of the

allotted time, the subjects were requested to complete the QUIS (Questionnaire of User Interface Satisfaction) (see Appendix B).

The QUIS is a measure of user satisfaction developed by the HCIL. This measure is well documented, and is a most reliable and valid measure of user satisfaction (Chin, Diehl & Norman, 1987).

Videotaping:

The primary reason for the videotaping of subjects was to provide a record of everything that transpired during the study, should there be a reason for going back to obtain a measure that was overlooked as an important source of information about either the system or the users behaviors and comments. There was no direct statistical evaluation of the videotapes, but they were useful to the researchers as a means of demonstrating procedures employed, and users performance on the system. Such data is also useful to *demonstrate* the shortcomings of the system to the consumers of human factors research, and to the designers that must correct the flaws in the system. Casual observation of these tapes has confirmed many of the original recommendations for the system, and have provided a medium to communicate this research to others.

QUIS findings:

The Questionnaire for User Interface Satisfaction (QUIS) provided insight into those areas of the system that need to be most scrutinized for improvement in the system. After acquiring some familiarity with the system, users were asked to complete a specially modified QUIS. Responses were compiled for each of the 52 participants in the study, along with any written comments that the subjects may have had. The reader will be better able to understand and appreciate the results by referring to the QUIS that is found in the appendix of this report.

Due to the nature of this study, less emphasis will be placed upon inferential statistics, as there is no system to which the current system is being compared. The greater emphasis will be upon the descriptive analysis of user agreement ratings on a number of statements regarding the system interface.

Realizing that some subjects will be more conservative in their responses, while others are more liberal, the researchers chose to utilize deviation scores from the mean of each of the respondents' ratings for many of the statistics. These will be referred to as mean deviation scores where used. The reason that these will be used, is to provide an indication of the *relative* rating of a given interface characteristic, and the relative importance of correcting that particular aspect of the interface. All statistics described as "significant" are considered to be significant at $p < .10$ unless noted to be otherwise. For those p-values set at .0015, that level was set by a Bonferroni adjustment to compensate for an inflated alpha error rate.

--Specific Findings:

Overall reactions:

There were six *overall reaction* ratings to which the users responded. These were specifically: 1) How impressed they were with the system, 2) Satisfaction with the system, 3) How stimulated they were with the system, 4) Ease of use, 5) Perceived "power" of the system, and 6) Flexibility of the system. Of these overall ratings, four of them were rated significantly lower than the mean response. These factors were: Satisfying, Stimulating, Powerful, and Flexible, indicating that these areas are subject to additional scrutiny. The other two overall ratings, Impressiveness and Ease of use, were not *significantly* less than the mean user response level (at $p < .0015$). From the four significant "Overall reactions" to the system, the researchers concluded that users found the system to be somewhat frustrating and rigid.

Further insight into the AVR system was obtained by more focussed comparisons between these "overall" items on the QUIS. An example of this sort of analysis is shown in Figure 4, indicating the finding, that the users tended to be more impressed with the system than they were satisfied with it. The bar graphs indicate that there is a general trend that the "Impressiveness" of the system is rated to be consistently higher than the "Satisfaction" of the user of the system.

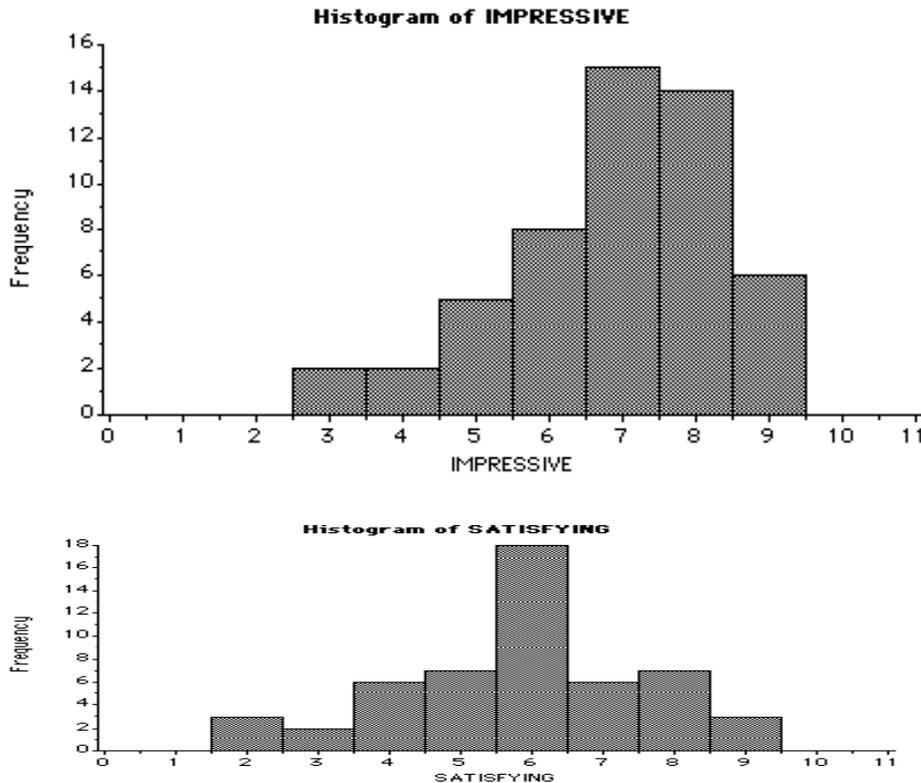


Figure 4: Graphic comparison between the histograms for "Satisfying", and "Impressive".

This perceived effect can then be subjected to more rigorous statistical methods. For this example, we found in a paired t-test that the result is indeed statistically significant, and has a mean difference of more than a full Lickert point, as seen in the following table.

Indeed, this result echoes the classic human factors concern: The user is impressed with the power and other aspects of the system, but is not satisfied with the usability of the system. Other related findings among the "Overall reactions" to the system that were statistically significant were that the perceived power of the system was greater than the satisfaction or the perceived flexibility of the system, and that the users perceived the system to be more "impressive" than "flexible." The actual mean differences and p-values for all significantly different overall ratings are shown in Figure 5, where the first system trait listed was rated significantly higher by users than the second trait listed:

Paired t-Tests				
	DF	Mean Difference	Paired t value	Prob. (2-tail)
EASY VS. SATISFYING	51	.808	3.329	.0016
POWERFUL VS. SATISFYING	49	.5	2.059	.0448
IMPRESSIVE VS. SATISFYING	51	1.038	4.283	1.0000E-4
EASY VS. STIMULATING	51	.615	2.25	.0288
IMPRESSIVE VS. STIMULATING	51	.846	3.673	6.0000E-4
EASY VS. FLEXIBLE	49	1.16	-4.202	1.0000E-4
POWERFUL VS. FLEXIBLE	47	.812	-2.866	.0062
IMPRESSIVE VS. FLEXIBLE	49	1.32	5.428	1.0000E-4

Figure 5: Paired t-tests between "Overall Reactions".

Demographic differences:

There was an overall significant effect of age upon the mean response level on the QUIS (In fact, there was more than a 0.10 Lickert point mean difference per year in age). Older subjects consistently rated all aspects of the system lower than other subjects. It was this finding that led to the use of the mean response deviation scores as the measure of choice for most of the results.

Additionally, there was a significant effect of the user's age in the rating of the impressiveness of the system above and beyond the previously noted base rate difference. Older subjects rated the system to be even less impressive than their younger counterparts.

Those subjects who were in groups of two or more rated the system to be significantly easier to use than did the individual subjects; however, there were no other significant effects of interest for any of the following demographics and conditions: Number of subjects in

group, Length of time on the system, Type of housing, Family income, Occupation or Major, or Previous experience with a variety of technical interfaces.

Individual System Characteristic Ratings:

The relative mean difference rating of a given system characteristic can serve as an indicator of the relative performance of the system on that particular characteristic. Therefore, the following results are intended to serve as indications of the system's performance in the particular areas cited.

The graph shown in Figure 6 on the following page represents in graphic form the mean response differences for each of the QUIS questionnaire items. The values on the ordinate represent the average deviations of each item from each of the respective means of the raters of the system.

Among all of the QUIS items related to the individual system characteristics, three proved to be rated significantly lower than all of the remaining items, as seen above. These items were 4.4, 9.0, and 9.2. Specifically, as stated, these items were:

4.4	Knowing where you are in the task (what you have done, and what you need yet to do)	confusing	1	2	3	4	5	6	7	8	9	clear
	NA											
9.	System keeps you informed about what it is doing	never	1	2	3	4	5	6	7	8	9	always
	NA											
9.2	User can control amount of feedback	never	1	2	3	4	5	6	7	8	9	always
	NA											

The prototype AVR system was rated to be confusing to the user with regard to the user's knowing where he or she was in the task (e.g. knowing what one had done, and what one yet needed to do) for question 4.4. This factor was rated more than a full point below the mean response level. Such a finding highlights the importance of keeping the user informed as to what aspects of the task have been accomplished and "where" the user is in the menu structure.

Similarly, in response to question 9.0, the users indicated that they felt that the system did not keep them sufficiently informed as to what the system was doing, indicating a need for more informative feedback. Question 9.2 indicated a related issue, that the users (rightly) did not feel that the system provided the user with the ability to control the amount of feedback provided. All of the issues stated above can be summarized by the issue of "control". Who is in control? The system, or the user? Of course, one of the keys to an

Figure 6: Mean response difference for each QUIS questionnaire item.

effective interface is the engendering of a feeling within the user that he or she is in control at all times. Many of these problems may be eliminated with the provision of a "MAC"-like "control panel" that will provide the user with the means to alter the amount of feedback and help from the system.

With regard to the issue of "help" from the system, it is important to note that the QUIS questions 15.0 - 15.3 (which dealt with the issues of help messages) were necessarily dropped from the analysis due to a substantial number (27 out of 52) of Not Applicable responses on those questions. This alone is sufficient to indicate that the users realize that the system is not providing any help to the users. In agreement with this indication, among the 25 raters that did respond to 15.0 - 15.3, they rated the "accessing of help" (QUIS 15.1) significantly lower (more difficult) than the mean response level, and the "amount of help" (QUIS 15.3) to be significantly inadequate.

As with the "Help" questions, the QUIS questions 10.0 - 10.2 (relating to the error messages on the screen) were not analyzed due to an insufficient number of user responses.

Of course, by no means were all aspects of the system poorly implemented. Items 1.0 - 2.3 indicate that the resolution of the screen is more than sufficient for the tasks that were evaluated in this phase of the study; however, it is not safe to say that the screen will remain sufficient should one attempt to invoke applications requiring high resolution graphics capability. "Going back to the main menu", item 4.3, was also rated to be significantly easier than average. This is an expected result in as much as there is a "back" key in every frame. Other favorable ratings were: 5.6, "Screen is left on long enough to read all of the items"; 6.1, "Consistency of screen headings"; 7.1, "Location of messages on the screen"; 13.0, "Remembering terms and use of commands; and 22.1, "Screen designs and layout are attractive."

Written Comments by Subjects:

In addition to the QUIS item findings, we also read and tabulated any written comments that the test users may have indicated on the worksheets and QUIS. This is done so as to accomplish two things, 1) to find indication of human factors concerns that we may have missed, 2) to determine the relative severity of the problems indicated, and 3) to serve as an indicator of the quality of the initial expert review of the system. Some of the comments are listed below, along with some brief discussion of each type of comment, and its frequency of occurrence out of the 52 subjects (indicated in parentheses after each representative indicative comment).

"The Security screens are difficult/complex/hard to understand" (n=9). Many of the difficulties with the Security screens have already been addressed, but it is indicative of the need to work on them that nearly a fifth of the respondents indicated that these screens were difficult to use. Other related comments related to Security screen difficulty were: *"It's hard to know what's shunted or armed"* (n=4), and *"The passcode system is unreliable and slow"* (n=3).

"It's hard to determine the on/off status or activation of devices" (n=19). A large percentage of the users indicated difficulty with both the determination of the on/off status of

References

Chin, J., Diehl, V. & Norman, K. (1987). Development of an instrument measuring user satisfaction of the human-computer interface. Technical report CAR-TR-328. Center for Automation Research, University of Maryland, College Park.

Potter, R., Weldon, L. & Shneiderman, B. (1987). Making touch screens effective: An experimental evaluation of three strategies. Technical report CAR-TR-327. Center for Automation Research, University of Maryland, College Park.

Appendix A: Worksheet used for user evaluation.

Thank you for participating in this evaluation!

The University is conducting this research in cooperation with American Voice & Robotics, Inc. AVR is a local company which is developing easy to use entertainment, security, and automation systems for the home.

The product you are about to use is a prototype. Your participation in this study will help us better understand the user friendliness and performance of this product. The results of this study will be incorporated into an improved version of the AVR system which will be available to the consumer market.

The device you will be working with is called a TOUCH SCREEN. In an actual application this device would be used to control a variety of system functions, such as security, lights, and other home appliances. In this evaluation, you will only be working with the TOUCH SCREEN, but keep in mind it is really part of a more comprehensive system.

When using the TOUCH SCREEN, the home owner simply touches the appropriate symbol or words on the screen to select his or her choice. The selection will then either activate a particular function or make appear a new "menu" of choices for the user to select. To select a particular "touch box", place your finger on it until it changes color. Once you are certain of your choice, release your finger from the screen to activate the selection. The system will not accept your choice until you release your finger.

Please take five minutes to familiarize yourself with the touch screen and explore the system features. To get going, simply touch anywhere on the screen. Once you are familiar with the use of the system please touch the "Back" box and try the following exercises. Writing down your thoughts and comments on the worksheet as you go will be a great help to the research team.

*Lets start with something simple:

Select Lighting Moods.

You now see a variety of "moods" that when selected will adjust the lights throughout the home to preselected intensities.

Try selecting "Quiet Evening" mood.

Remember in an actual home, the lights throughout the house would be adjusted to create a peaceful night time setting.

Now try to turn on all the lights.

Do you believe you made the right selection? _____

How did you know which mood you selected? _____

Touch "Back" to return to the main menu.

*Now will try to schedule an event:

Please touch "Schedule Management".

Try to schedule the lawn system to turn on at 4 pm, tommorrow, and run for one hour.

Do you think you've done it? _____

How do you know you scheduled the event correctly? _____

Any Comments? _____

*Next we'll try to turn on or off a couple of devices around the home.

Turn off the tennis lights. If they are already off, turn them on.

Put the System into "Vacation Mode"; if its already in "Vacation Mode" go to the next step.

Please indicate whether each of the following devices are on or off:

Vacation Mode

ON OFF

Weekend Switch ON

OFF

Paddle Tennis Heater 1

ON OFF

Comments about Device Control? _____

*Now we are going to try something a bit different.

Go back to the main menu and touch "Security Management".

We will first try to "shunt" a security zone. "Shunting" means you tell the system to ignore that particular area or zone, so you could, for instance, leave the door opened and still "arm" your security system. You must first select the zone by touching it and then shunt it.

Try to "shunt" the garage storage doors on the first floor of the home.

Is the proper zone shunted? _____

How do you know? _____

Unshunt the same zone.

is the proper zone unshunted? _____

How do you know? _____

Now "Arm" the security system; the pass code is 1-2-3.

Is the system armed? _____

How do you know? _____

What do you think about the security system screens?

Please select "Quit"

Is the system in vacation mode?

YES NO DON'T KNOW

You can continue to explore the system if you wish. Please let the experimenter know you are finished.

If you are interested in participating in an on-site system evaluation in a real home or in user focus groups, please leave your name and where we may contact you below.

Thank you VERY much!

University of Maryland & AVR

Do not copy without the author's written permission.

Part 4: User Evaluation of the AVR System

Please circle the numbers which most appropriately reflect your impressions about using this system.
 Not Applicable = NA. **Please add your written comments below the corresponding item.**

Overall reactions to the system:	unimpressive					impressive					
	1	2	3	4	5	6	7	8	9	NA	
	frustrating					satisfying					
	1	2	3	4	5	6	7	8	9	NA	
	dull					stimulating					
	1	2	3	4	5	6	7	8	9	NA	
	difficult					easy					
	1	2	3	4	5	6	7	8	9	NA	
	ineffective					powerful					
	1	2	3	4	5	6	7	8	9	NA	
	rigid					flexible					
	1	2	3	4	5	6	7	8	9	NA	

PART A: SCREEN

1. Characters on the touch screen	hard to read					easy to read					
	1	2	3	4	5	6	7	8	9	NA	
1.1 Image of characters	fuzzy					sharp					
	1	2	3	4	5	6	7	8	9	NA	
1.2 Character shapes (fonts)	barely legible					very legible					
	1	2	3	4	5	6	7	8	9	NA	
1.3 Contrast with the background	irritating					pleasing					
	1	2	3	4	5	6	7	8	9	NA	
2. Highlighting on the screen makes task easier	not at all					very much					
	1	2	3	4	5	6	7	8	9	NA	
2.1 Use of color for highlighting	unhelpful					helpful					
	1	2	3	4	5	6	7	8	9	NA	
2.2 Use of blinking	unhelpful					helpful					
	1	2	3	4	5	6	7	8	9	NA	
2.3 Use of color background change	unhelpful					helpful					
	1	2	3	4	5	6	7	8	9	NA	
3. Screen layouts make tasks easier	never					always					
	1	2	3	4	5	6	7	8	9	NA	
3.1 Amount of information displayed on screen	inadequate					adequate					
	1	2	3	4	5	6	7	8	9	NA	
3.2 Arrangement of information on screen	illogical					logical					
	1	2	3	4	5	6	7	8	9	NA	
3.3 Consistent arrangement of information	never					always					
	1	2	3	4	5	6	7	8	9	NA	

4. Sequence of screens	confusing							clear		
	1	2	3	4	5	6	7	8	9	NA
4.1 Next screen in a sequence	unpredictable							predictable		
	1	2	3	4	5	6	7	8	9	NA
4.2 Going back to the previous screen	difficult							easy		
	1	2	3	4	5	6	7	8	9	NA
4.3 Going back to the main screen	difficult							easy		
	1	2	3	4	5	6	7	8	9	NA
4.4 Knowing where you are in the task (what you have done, and what you need yet to do)	confusing							clear		
	1	2	3	4	5	6	7	8	9	NA
5. Screens items are easy to select	never							always		
	1	2	3	4	5	6	7	8	9	NA
5.1 Items are easy to find	never							always		
	1	2	3	4	5	6	7	8	9	NA
5.2 Item selection area size enough	too small								large	
	1	2	3	4	5	6	7	8	9	NA
5.3 Seeing cursor (finger light) location	hard							easy		
	1	2	3	4	5	6	7	8	9	NA
5.4 Knowing whether an item is selected	hard							easy		
	1	2	3	4	5	6	7	8	9	NA
5.5 System responds to selection when finger is pulled away from screen	unreliably							reliably		
	1	2	3	4	5	6	7	8	9	NA
5.6 Screen is left on long enough to read all items before it goes away	too short							long enough		
	1	2	3	4	5	6	7	8	9	NA

PART B: TERMINOLOGY AND SYSTEM INFORMATION

6. Use of terms throughout system	inconsistent	consistent								
	1	2	3	4	5	6	7	8	9	NA
6.1 Screen headings	inconsistent							consistent		
	1	2	3	4	5	6	7	8	9	NA
6.2 Terms on the screen (item labels)	ambiguous							precise		
	1	2	3	4	5	6	7	8	9	NA
7. Messages (feedback) which appear on screen	inconsistent	consistent								
	1	2	3	4	5	6	7	8	9	NA
7.1 Location of messages on the screen	inconsistent							consistent		
	1	2	3	4	5	6	7	8	9	NA
8. Instructions to the user	confusing							clear		
	1	2	3	4	5	6	7	8	9	NA
8.1 Instructions for commands or choices	confusing							clear		
	1	2	3	4	5	6	7	8	9	NA
8.2 Instructions for correcting errors	confusing							clear		
	1	2	3	4	5	6	7	8	9	NA
9. System keeps you informed about what it is doing	never								always	
	1	2	3	4	5	6	7	8	9	NA
9.1 Performing an operation leads to a predictable result	never								always	
	1	2	3	4	5	6	7	8	9	NA
9.2 User can control amount of feedback	never							always		
	1	2	3	4	5	6	7	8	9	NA

© 1988 Human Computer Interaction Lab, University of Maryland.
Do not copy without the author's written permission.

In-house Appendix C: Proprietary

General Impressions of the AVR System, and Changes Recommended by the HCIL:

Miscellaneous :

Does the use of the color red imply "stop" or "off" when used as text on the screen? Is it a potential problem in this system? One must maintain consistency, not only within a given system, but with the user's environment at large as well. Red is a common "warning" color, often indicating "STOP" (as in a traffic signal). For this reason, the use of red as a highlight color should be used carefully, if at all.

Is an italic font appropriate? It may be harder to read for some individuals. Limitations in screen resolution can result in italics actually becoming less readable to the users.

Perhaps icons should be introduced (associating each menu item with an icon). Although it may clutter the screen a bit, with care, it will provide faster reading and selection times, and more accurate input for those persons using the system several times a day, every day. Icons were not used in the original AVR system, however, research is currently under way to develop meaningful and memorable icons for the next generation systems.

There is currently no way to "undo" the most recent action. The only way to correct an error is to start over completely. In the deeper menus, this can be very irritating to the user.

Screen time-outs:

The time-out period is altogether too short for many of the frames, especially for complex screens such as the maps. We found that the maps (and other complex screens) disappeared before we could reasonably respond. If one should desire to move back to the dormant screen quickly, one may always use the "QUIT" option provided.

On the other hand, the time-out doesn't exist at all for some of the frames. For example, the home status screen did not time out even after four or five minutes (it should be relatively consistent between frames).

Cursor placement:

The cursor is actually below the finger (except at the bottom of the screen). Most often the cursor cannot be seen since it is obscured by the finger. We would suggest that since the finger is most often below the visual field, given that the hand is below the field of view, the cursor which appears upon the placement of the finger on the screen should be slightly *above* the finger (so as to be seen). The worst situation occurs when trying to select "Quit" and "Back". It is very difficult to get the cursor to select those items since the cursor is too low on the screen. Also, if the screen cursor is slightly above the finger, one must insure that

there are no "selectable" areas at the very bottom of the screen, as these will be very difficult to reach (due to the protrusion of the touchscreen frame).

Touch mechanism feedback:

There is one state missing in the selection of items. Currently the progression of events in the selection of an item is:

- a) Item is white = neutral state
 - b) Item is green = touched state
- Proposed step
- c) Item is ????? = selected state
 - d) Choice is executed

When one releases (thus selects something), either the action should take place immediately or some change should witness the action (selection). This could be done by converting the item to reverse video, a flashing item, or the appearance of a "working..." flag. Perhaps "touched" state indication could be some other color than green, while using "green" (for GO) as the flag to indicate that the item has indeed been selected. The users need more system feedback (as opposed to the feedback provided by the affected changes around the home), since some system changes may not be readily apparent to the user from his or her location in the home. This is important in situations where the user may inadvertently slide his or her finger off of the item an instant before the finger is removed from the screen, thus selecting either nothing or an adjacent function. Without a "selected state" indicator, one would not be sure that an action had indeed been performed.

Continuity between screens:

There is a need for some smooth continuity between screens:

Example: when selecting "lighting mood", the jump to the "lighting mood" screen could be smoother (remove the harsh black flash. Fade in/fade out between the two screens, or have the selected word slide to the top of the screen to become the title. If animation or fade is not possible, wipe the new screen on gently or replace the black screen with a screen of the same background color.

Some inconsistency exists in screen titles:

The selected item should consistently become title; but after selecting "Independent schedule" or "Device control" the next screen title is "Select a device". In this case, the type of title is also changing: "Select a device" is not a name but an action or instruction.

Inconsistencies in the "Back" key:

The "Back" key will take one back to the main menu from most screens, but in the main menu it is the same as "Quit". In the deep menus it would be preferable to have a "back" key to the previous menu as well. We feel strongly about this. A "previous menu" key would in effect perform the same function as an "undo" function (thus providing easy error recovery).

Individual screens:

--First screen (dormant screen with the AVR logo):

There is no real immediate feedback when one touches the screen:

If the finger touches the white area, the cursor is invisible; however, if one touches on a black area, a green cursor appears. As stated before, the cursor is nevertheless hard to see.

There is nothing to signal to the user that the main menu is coming when the dormant screen is touched. What happens now is that when one touches the screen there is a several second wait before the main menu appears; if there is another touch, the system restarts the search for the main menu. Should one touch the screen several times nervously, nothing will happen until the user stops. Either a) the next screen should appear more quickly, b) an immediate feedback should warn the user that his touch has been acknowledged, (by tone or visual feedback) or c) the system should ignore the repeated touches.

--Main menu:

Back and Quit do the same thing. Back is typically "back to the main menu", so it should not be present in that screen. However, it might be good to keep the screen uniform across frames (it must still be consistent in taking the user to the main menu -- eg. at the main menu the "back" key apparently does nothing to change the screen).

The word "Back" is not clear. It took a while to understand that it takes the user back to the main menu. Therefore, it should be labeled "Main Menu", "Start", "Restart", "Opening Menu", or some other unambiguous label.

--Security management screens:

These screens were especially difficult to use and understand. To a large degree, the difficulties arose from the original design intent to mimic the appearance of a conventional security management control panel. This brings up a difficult issue in interface design: Should one try to emulate the "real- world" system (so as to provide a *familiar* interface) when the original interface is of a poor design? Since the AVR system is intended to replace the individual interactions with separate systems anyway, why not revamp the entire security interface to provide a simple, clear, representation of the relevant objects and actions.

Flow between screens is choppy and confusing-

The screen displaying the message "Checking the security system" is particularly perturbing. This message wipes out the entire screen, thus preventing any further examination of the floor plan of the house (map screen) to determine the next action to perform. This message should appear as a small flag *on* the map screen to give a better anchor to the user and to allow him or her to appreciate the modifications performed on the map. The visual effect of going back and forth from the light blue screens of the AVR system screens to the black map screens and back again can leave the user with the impression of helplessness since there is a lot of apparent action going on without anything that one can do to regain control of this runaway system. Smooth flow and transition is the key.

There is currently no discernible way of getting out of the map screen:
Should one accidentally select the security mode, the only way to get out of the map mode is to let the system time-out.

Floor Choice Screens-

The floors presented in the menu are the reverse of the expected geography of a house (e.g. the basement should be at the bottom of the screen, the top floor at the top). A cross section of the house presented on the map screen may also be

used and may indeed be preferable. The current system presents the floors in inverted fashion as:

BASEMENT
FIRST FLOOR
SECOND FLOOR

Consider a default floor choice. Why not go directly to the first floor (or anyway the most often modified and checked floor for a given house). The map screen can then directly access the other levels using "up" and "down". In the current system, one must go back to the main menu to access a new floor.

The map screen-

As indicated earlier, the Quit and Back commands have been forgotten.

Security zones are difficult to select. So long as the zone is marked in some way when selected, enlarge the touch area to make it more easily selectable.

The map doesn't indicate which zones (doors and windows) are available that one can check and modify.

The map screen should not disappear when the message "Checking security system" appears! This message could just as easily appear in a corner or in an overlapping window. It is likely that the user could save otherwise wasted time by examining the screen for the next action while the systems are being checked. The user can also use this time to make sure that the action performed by the system is the action that he or she desired to perform. The system may provide more feedback at this point as to which action is currently being executed.

The symbols used are potentially confusing-

The cursor used is the same symbol (green square) as a shunted door.

When a zone is selected, it turns blue. But as soon as one touches the screen somewhere else (for example when you are approaching the "Arm" button), the mark disappears. Thus the user is unsure which if any zone is being acted upon, in spite of the message shown at the bottom of the screen.

There is no way of selecting several zones at the same time. It would be helpful and a lot faster to be able to arm and disarm, or shunt and unshunt several windows at once. This is potentially frustrating to the user, and we certainly found it to be so.

The Arm and Shunt buttons are confusing-

The "arm" button is not well labeled. It should be made clear that it is a toggle and not necessarily an arming of the system. Stating "Arm/Disarm" may be clearer. Granted the map screen is supposed to conform with the security panel, but the toggle representation should be the same here as in the rest of the system (for example show some sort of ON/OFF marks). The current status of the system and the means of changing the status must be consistent across functions in the system.

The "arm" button is not well represented in that the light which indicates its state is too far away. The red light symbol which indicates whether or not the system is armed is too close to the map legend and can thus be easily misinterpreted as being either part of the legend or the arm/disarm toggle.

The "shunt" button is similarly confusing.

The map legend is easily misinterpreted as being an area to perform actions.

--Device and system control screens:

These 2 menu areas are similar in many respects, and both present a series of toggle screens; therefore, these will be addressed together.

The current toggle representation is very confusing. It is hard to determine what device is on and what is off. The screen is very cluttered by the "on" and "off" words.

Current system:

When one touches the word, the ON/OFF mark toggles.

Screen Example:

vacation mode	ON (yellow on green)	OFF (red)
weekend switch	ON (red)	OFF (yellow on green)
platform heater	ON (red)	OFF (yellow on green)

Main problem:

The ON/OFF active state is not clear. There is a contradiction between the message "off" and its being printed in green (implying GO or ON), as there is in seeing that the word "on" is present for every item on the list (some printed in red).

Some Possible Alternatives:

Display only the current state.

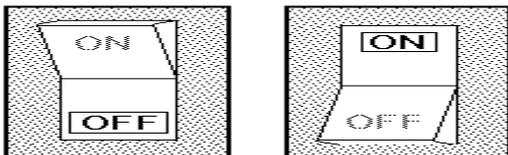
vacation mode	ON (green)	
weekend switch		OFF (red)
platform heater		OFF (red)

The ON and OFF area could be touchable separately.

vacation mode	ON (big, green)	off (small)
weekend switch	on (small)	OFF (big and red)
platform heater	on (small)	OFF (big and red)
-----	-----	-----

Zones:	=	=	=
Effect:	toggle	Turn ON	Turn OFF

Another possibility may be to draw a realistic 3-D view of a toggle switch, showing which side is down and which side is up, indicating both current status and the means for changing the status. For example:



Other ideas and remarks:

Granted that since a toggle mechanism is not easy to manage, it may be easier to touch the word "ON" to turn something ON, than to first check its status and then decide if this status needs to be toggled.

Wording of device and system options:

"Switch" and "mode" should be uniform in their phrasing. They currently are not. It is recommended that they be rendered in the following manner, since they are equivalent in meaning.

vacation mode		vacation switch
weekend mode	OR	weekend switch
etc...		etc...

--Independent schedule screens:

It is a very deep menu area. It takes 5 screens (5 selections) just to change a single device schedule, 10 selections to select both on and off times. These figures do not even reflect the number of user actions required to increment dates, hours, and minutes on the schedule screens.

Screens are not labeled to indicate which device is currently being scheduled. This is very important.

It lacks a "go to the previous menu" button. Should a mistake be made at or near the last frame, it would be costly in terms of the time required to rectify the error.

After selecting "Independent Schedule" the menu title should not be changed to "select device"; screen titles should be consistent.

It is very difficult to select dates and times, especially if you want to schedule a device to start on the half-hour and end on the hour. It requires the user to press the increment button on the screen at least 30 times to advance the minutes to the half-hour, and then another 30 times to set the ending time back to the hour! No user will stand for that sort of punishment for very long, rather, he or she will simply stop using that function.

The 2 menus :

- Turn device on
- Turn device off

and the

- "select date"
- "select time"

menus could be replaced by more efficient, graphically oriented screens. Select the date from a calendar and set the hour on a watch with hands or on a time line displaying the on and off states. The use of a digital-watch time setting mechanism is very painful and doesn't take advantage of touchscreen possibilities; one may do just as well or better by using keyboard entry.

Another potential difficulty is that there are insufficient warnings of user error, and insufficient instruction as to what comes next in the task. For instance, the user is not reminded that he or she must select an "off" time for the device after selecting the "on" time. Also, the user must select the date first, on one screen, and then select "Finished" in order to reach the menu to select the "on" or "off" time desired. It is a very uncomfortable feeling for the user to select "finished" when one knows that one is not finished with the task.

--Lighting moods screens:

There is no feedback to tell the user which mood is currently on, and when or whether the selected mood is actually on.

--Password screen:

It is not visually appealing.

The screen is so large that it is uncomfortable to swing the arm around the distances necessary to cover the whole screen. Also the wide swing makes it easier for an unwelcome individual to perceive the password from an oblique angle.

The numbers are not laid out in any standardized form. ATM machines and telephones have the most common layout, so it is sensible to mimic that structure. The keypad on the screen could potentially look something like this:

1	2	3
4	5	6
7	8	9
Enter	0	Clear

The touchscreen response time is sufficiently slow that the user may inadvertently enter the wrong password because the screen was not able to keep up with the user's finger. A faster touchscreen would be desirable if possible. If not, methods must be developed to insure that the entry of the password is performed accurately and easily.