

Affective Feedback Devices for Continuous Usability Assessment

by

Matthew Rowan Norwood

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

January 14, 2000.

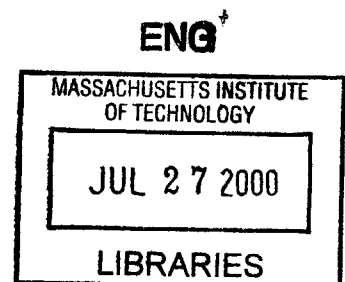
[February 2000]

© 2000 Massachusetts Institute of Technology. All rights reserved.

Author _____
Department of Electrical Engineering and Computer Science
January 14, 2000

Certified by _____
Rosalind W. Picard
Thesis Supervisor

Accepted by _____
Arthur C. Smith
Chairman, Department Committee on Graduate Theses



Affective Feedback Devices for Continuous Usability Assessment
by
Matthew Rowan Norwood

Submitted to the
Department of Electrical Engineering and Computer Science

January 14, 2000

In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

Automated user-feedback systems are set up in parallel with two different machine-user interfaces in actual work environments and evaluated in their effectiveness in providing the designers and service providers with user satisfaction data. One system is set up to provide feedback on a web-based software interface. The other is connected to a shared physical device, a coffee machine, using a tangible interface. The goal of both systems is to collect satisfaction data from users through natural, social, proactive, affect-based interfaces and correlate the data on the user's affective state with the machine's internal state to isolate critical states for service or redesign. By these means, the systems provide service providers and designers with design and service evaluations based on a human-value heuristic: the level of emotional satisfaction experienced by users. Patterns are examined from the collected data, revealing some data for redesign or re-prioritizing of services. Seven principles for automated feedback design are extracted from experience during the study.

Thesis Supervisor: Rosalind W. Picard
Title: Associate Professor, MIT Media Laboratory

1. INTRODUCTION

Providers of goods and services have long sought effective methods of communication with individual customers. Meaningful information from individual users would allow providers to formulate precise strategies for optimizing their products or services rather than settling for data from focus groups, marketing surveys, and trends in the market or the news. In order for the use of this kind of data to be within the means of most organizations, however, both the collection and analysis of the data would have to be efficient and scalable over a large population of customers.

Current modes of customer-to-provider communication tend largely to fail to meet one of the two criteria for efficiency: they are either too difficult to collect, or their content is too difficult to process on a large scale. While work is in progress to increase the efficiency of collecting and analyzing currently-used types of information, this paper proposes a new mode of customer-to-provider communication that is both efficient and scalable in its collection and analysis. The work aims to provide some insight into the nature of the customer-provider dialogue and open wider the door to a new paradigm in provider-customer relationships, one where the customer or user is always an integral part of the design or service management process.

The work explores a direct, computationally-based communication system that monitors a user's basic satisfaction or dissatisfaction with specific aspects of a product or service and reports them to the provider of the product or service. The systems are designed primarily to explore the quality of the interactions between users and machines by measuring the satisfaction of the user. The systems explored here use tools developed in the field of Affective Computing to evaluate a user's emotional reaction to an interaction with a product or service. Tying the user's emotional data with data about the specific context of the interaction, they are then used to enable the communication of this data in a meaningful way to the developers or providers of the product or service.

After detailing the design and implementation of the systems, results are provided from trial runs with actual users in the field. The data is analyzed for its

potential utility to designers and service providers. Data from users about interacting with the systems is also provided and analyzed. Finally, the viability of these types of systems, along with directions for future work, are discussed in light of the collected data and user reactions.

2. BACKGROUND AND RELATED WORK

2.1 Customer Satisfaction Rating

A number of factors contribute to customer satisfaction with a product or service. Generally, user satisfaction is tied to the usefulness and ease of use of the product relative to the investment of money and time in obtaining and using it. Most models of customer satisfaction are built using surveys of customers and examining market trends (Das 99), but designed products can also be tested for their user satisfaction ratings using tools and techniques from the field of usability engineering.

2.2 Usability and HCI

As the workplace becomes increasingly computerized and more jobs require daily interaction with computers, the quality of human-computer interaction (HCI) becomes more and more important to the quality of the lives of workers. Productivity on the job and personal happiness are intrinsically linked, and both are strongly affected (directly and indirectly) with the quality of HCI (Landauer 95).

Much of the evaluation of software interfaces in terms of HCI quality and usability is done by an editor, someone outside of the software development process looking at the interface from a user's perspective (Dameron 90). The editor is usually trained to look for potential interface flaws, but tries to maintain the perspective of a typical user of the interface. However, editors are usually HCI experts by profession instead of actual users of the type of software being developed. When users are

involved in HCI and usability evaluation, it is almost always in the form of paper questionnaires (Tullis 98) or small group meetings with designers (Starr 87, Dixon 97), putting a severe limit on the amount of user data that can be collected. As a result, user testing is a huge expense to product and service developers, and has been falling off in recent years due to time and monetary pressures (Landauer 95).

HCI quality can be measured using two different metrics: the ability of users to accomplish specific goals using a computer system (usability/usefulness), and the enjoyability of the HCI experience on an emotional level. The first standard is the subject of an entire field of industrial design, usability engineering, while the second is only addressed peripherally by usability engineers.

2.3 Affective Computing

Research is underway in the Affective Computing group at the MIT Media Lab to study the emotional dimension of HCI. Affective computing is computing that relates to, arises from, or deliberately influences emotions. Tools are being developed to improve emotional conditions in a computerized workplace. Agents have been developed which respond to user frustration, and an array of sensing methods have been tested to detect and respond to expression of a user's emotional state (Picard 97). However, few of these tools have been applied in an actual user environment – they have been largely confined to pilot studies and psychological experiments in the lab.

2.31 Frustration Detection in Human-Computer Interfaces

Drawing on psychophysiology literature, Raul Fernandez attempted to detect emotional state from patterns in biosensor data. Research in the psychophysiology community has proposed that affective states are mapped onto physiological states that we may be able to measure. In order to collect data carrying affective content, an experimental situation was designed so that a user, engaged in a computer task, would experience a frustrating incident while his or her physiological signals were being collected and recorded for further processing.

Using the physiological signals collected during the experiment, Fernandez proceeded to develop a user-dependent system that learns typical patterns of the physiological signals under the stimulus of frustration. Modeling affective signals is a hard problem. Some of the difficulties he encountered were the uncertainty associated with the effectiveness of the applied stimulus, and the non-stationary of the data. The analysis of the data set was carried out in a probabilistic framework by implementing machine learning algorithms for analysis of time series. Using Hidden Markov models, he obtained recognition rates which, for subjects with enough experimental data, were better than random for 21 out of 24 of the subjects (Fernandez 97.)

2.32 Computer Response to User Frustration

Jonathan Klein in the Affective Computing group carried out a study of the possibility of a computer system to effectively respond to humans' emotional states. He built an emotional support agent designed to respond to user frustration. Klein conducted an experiment to test the efficacy of the agent's intervention in improving mood and performance on later tasks, and the results showed some promise for such agents (Klein 98).

2.4 Automated Service Management

Studies have been made into the automation of service management, the provision of services to customers and how such provision is prioritized and otherwise managed. Much of this research has been within the field of computation itself, automating the management of computer-based services such as Internet service (Katchabaw 98). By allowing the service provider access to detailed quality-of-service data in a quantifiable form, algorithms for responding to such data are vastly simplified and able to be handled in large part by automatic computer systems.

2.5 Computer-Based User Feedback

A number of feedback techniques are already being used by various designers and service providers, and several other informal feedback techniques exist which have yet to be tapped by designers. Much of the computer-based customer discussion centers on technical issues, but non-technical providers have also begun to capitalize on computer-based feedback on their products and services (Brannback 98).

2.51 Email

Email has become the most common way for companies to receive feedback on their services and products using computers (Shiva 99). By providing customers with an email address on a product, web page, or business card, a provider can open up the possibility of addressing customer comments over the Internet. Using this information to address the quality of service provided has shown some promising results (Lowder 99), but one of the major stumbling blocks has been processing the unconstrained context of the emails from a large population of customers without employing a huge customer support staff. Some attempts have been made in the field of artificial intelligence at automating technical support systems (Giraud-Carrier 98) and automatically processing general emails and classifying their content (Shiva 99). Both approaches show promise, but fail to achieve the specificity required to obtain large amounts of useful data on specific parts of an interface or service.

2.52 Application-level Network Transmission

Some work has addressed the possibility of sending data directly from users to providers over a network without the user explicitly providing the content of the message. This allows the provider to specify the format and content of the data when the software is distributed, allowing highly specific information on the context of the computing environment to be conveyed. This has been used by some service and software providers to collect data on users' software use habits for marketing purposes (Robinson 99), but has also been used by companies like NoWonder software (www.nowonder.com) to report technical failures on the part of attached software, sending the designers detailed machine context information for later bug-

fixes. NoWonder's TalkBack software has been incorporated into recent versions of Netscape Communicator (www.netscape.com) and has negotiated support contracts for Microsoft Windows (www.msn.com). While providing detailed technical information for fatal system bugs, this approach unfortunately does not provide a channel for user experience-based complaints.

2.53 User-Collaborative Computer Support

Some user feedback and user support techniques have capitalized on the natural tendency of communities of users to cooperate among themselves to provide a user support network and discussion forum. Research has shown that organized communities of computer users can provide significant technical support to each other and empower each other in making use of computer technology (Clement 90), and some have developed techniques from transferring this kind of community from a shared office space to a shared virtual space. ThirdVoice software (www.thirdvoice.com) has developed a software agent and service that allows web browsers to post and read comments on online pages, which are visible to all ThirdVoice users when they look at the annotated pages. Posts can be filtered by the "groups" to which a user belongs, allowing for specific online communities to annotate web pages with comments visible to other interested parties. The graphical nature of the web-page markup tool enables much more specific annotation of interface features, lending itself to web page design critique. Indeed, ThirdVoice has already found itself used as a critic of online sites and services, empowering users to show other users – and providers – exactly what they think of the provider:

Third Voice has helped create a sort of "Consumer Voice" on corporate sites. Annotations on corporate Web sites act as a sort of online Consumer Reports and analysts agree that this would help keep a check on consumer product manufacturers as well as provide them with valuable consumer inputs to improve their products. (Purushotham, 99)

3. MOTIVATION

The purpose of this project is to explore possible designs and implementations for systems to detect and respond to user frustration and usability problems in an HCI or human-machine interaction task involving actual software or hardware being used in the workplace. By demonstrating the viability of affective computing technology as a way of gathering real-world user data and responding to user emotional needs in human-machine interaction in the workplace, the possibility is opened up to introduce a new tool to the fields of usability engineering, tech support, service management, and software development.

A workplace-based feedback system for a user interface or service would address several distinct problems:

- 1) Design processes, especially in the computer domain, tend to engage in very little user testing, even though user-centered design has been shown to have dramatic effects on the productivity of software users. One study have shown that in the absence of user testing, even experienced designers tend to discover an average of only 44% of interface design errors (Molich 90). In his book, The Trouble With Computers (Landauer 95), Thomas K. Landauer predicts a 100 percent increase in productivity of a piece of software designed with two cycles of redesign following feedback from user-testing methods. By using an automated feedback system in the workplace, software developers would have access to the kind of user data which must usually be gathered in the lab or by costly monitoring by trained consultants in the workplace (Brannback 98). In fact, designers usually have to settle for simulated user sessions for gathering data, which Landauer calls "compromise and approximation" in contrast to actual workplace data, the "gold standard" for usability design. Landauer states that compromise methods are used because "usually practicalities get in the way. Real workers and real designers are busy in different cities. Measurement and observation is too cumbersome, intrusive, and slow in the real workplace." Indeed, analyses of the inclusion of actual users in the design process have been pessimistic about the costs relative to the benefits using standard methods

(Mantei 88). By using an automated, computer-based feedback system over a network, monitoring could be accomplished at a distance with little to no intrusion into the workplace, and the data could be collected and analyzed in real-time by the designers.

- 2) As design and production cycle time decreases in the computer and high-tech machine industry, the products released to users are increasingly rough and unfinished – closer to prototypes than final products. Unfortunately, usability redesign – due partly to its time-consuming nature – is one of the first items to be glossed over in the rush to release. Computing products reaching the user have had less usability redesign and more bugs as time goes on. If this trend continues – and it certainly appears that it will – some kind of quick feedback system would allow designers to post quick patches for bugs and usability issues based on user feedback in the weeks following release. Using the web as an update forum, designers could be very honest about the unfinished nature of their products, posting patches as user feedback floods in through the automated feedback system.
- 3) User emotional support is a vital part of HCI which has been overlooked by most designers. Clifford Reeves and Byron Nass, in their book The Media Equation (Reeves 96), present experimental data to suggest that human relationships with computers are inherently social rather than utilitarian in nature. Based on these findings, other psychologists have studied the power computers have over our emotional state when we interact with them. We tend to have expectations of computers which can provoke extreme emotional reactions if unfulfilled – conversely, this power wielded by computers over our emotions can be used to assist us in managing our own emotional states. Jonathan Klein, Jocelyn Riseberg and Rosalind Picard, in the paper "Support for Human Emotional Needs in Human-Computer Interaction" (Klein 97), present a number of possible approaches to this technology. The networked feedback system, in the process of registering a user's reactions to other software, would be in an ideal position to respond with an emotional-support agent like the ones Klein describes. Indeed,

simply by letting users know that their reactions are being used in the redesign process would provide a certain feeling of empowerment which could greatly improve the emotional dimension of HCI.

- 4) Products or services that require occasional service from their providers usually receive such service only after a user, recognizing some problem with the product, takes the initiative to inform the service providers. Such a service request can include only as much information as the human user can convey about the product's problem to the service provider, and must be prioritized and prepared for using only this information (Brannback 98). If a system were implemented to not only convey up-to-date product information to the service provider, but also to provide the service provider with data on the levels of user satisfaction with the product over time, service calls could be addressed and prioritized in a fashion befitting their nature and their urgency, instead of relying simply on the low-resolution and potentially non-specific information relayed by a single user over email or telephone.

4. METHOD

Designing, implementing, and testing the feedback systems was executed over a series of months. The project began with an attempt to track user behavior in a networked software environment and monitor their reports for information on improving the user interface. At the conclusion of that study, a physical interface shared by all the members of a work environment was studied – the group's coffee machine. Each case is detailed chronologically, with explanation of design and implementation choices.

4.1 Case One: Feedback System for Network Software

The first system targeted for testing was a package of financial software running over a network. Negotiations were made with SAP Labs – a software firm open to

HCI innovation and experimentation – to implement the feedback system on a web-enabled version of their R/3 business software, SAPWeb, that had been put in place by the local institution to allow wider access. The locally-developed web interface had been the source of numerous interface complaints and, due to its limited user population, did not have an extensive support network in place. These facts, and the relative novelty of the product, meant that simple interface flaws might yet be revealed by users before later versions corrected them. The software system had recently been adopted by the organization being studied, and most users had only moderate experience using it. It was felt that this would leave them more open to looking at the software “from the outside” and evaluating it. Most users had moderate levels of technical experience: they were administrative assistants and financial officers who relied on computers for most of their daily tasks, leaving them experienced end-users. They were comfortable with their computing environments and with standard user interfaces, but were only somewhat familiar with the interface in question: ideal candidates for evaluating the interface.

Preliminary recruitment of subjects for the study yielded nine candidates. Users were recruited on a volunteer basis from the SAPWeb users in a single academic department to maintain some homogeneity in the types of tasks attempted with the software, as well as the training process the users had been through. Of the nine candidates, two were excluded because their roles as financial officers precluded regular use of a feedback system on SAPWeb. All remaining users were administrative assistants, although the length of their experience with the software varied. By the end of the study, two more users had taken other positions and were not available for follow-up data collection. The details of user information are contained in the Observations section.

4.11 Initial Questionnaire

Users were given a preliminary questionnaire to obtain basic user information and to inform the design of the feedback system. A copy of the questionnaire with results is provided in Appendix A. Some of the questions addressed the mode in which users usually express frustration with computer interfaces in the absence of a feedback system. Choices ranged from yelling at the computer to hitting things or

walking away. These choices capitalized on the tendency of users to interact with computers in a social mode, as detailed in The Media Equation by Byron Reeves and Clifford Nass (Reeves 97). Reeves and Nass's work suggests that users react to computers and other complex media by projecting human qualities onto them, creating the same kind of social and emotional dynamics that are seen in interactions between people. Current HCI paradigms usually fail to take this factor into account, and remedying this lack of "emotional intelligence" in HCI has been one of the major aims of affective computing (Picard 97, Klein 98).

In designing a feedback system for the network software, it was hoped that natural human means of social communication of emotions could be harnessed to provide a social interface between the user and the software. Suggestions were made on the questionnaire for possible ways to capture this kind of natural interaction with the computer: objects to be squeezed, microphones or cameras to capture user behavior or speech, or on-screen GUI features. Biosensors, a common tool in affective computing for sensing psychophysiological data, were left off of the survey under the assumption that they would be too alien and uncomfortable for everyday use in a desktop environment – later informal interviews with respondents confirmed this assumption.

The results of the survey are summarized in Figure 4.1. Given that the most popular choice was on-screen interaction, and given the ease of integrating this feature with a browser-based interface, a simple GUI was deemed the best approach to collecting feedback for this project. The other user choices, however, were kept in mind for later designs – some led to parallel development of feedback devices for context, like a squeezable/throwable bean bag and squeezable mouse, while the overall idea of applying pressure to an object was incorporated into Case Two of this study. It is interesting to note the lack of interest in passive sensing devices such as video cameras and microphones – most respondents, when interviewed, indicated privacy concerns as the major reason for avoiding these devices, even if the data were to be warehoused only in the form of specific feedback information.

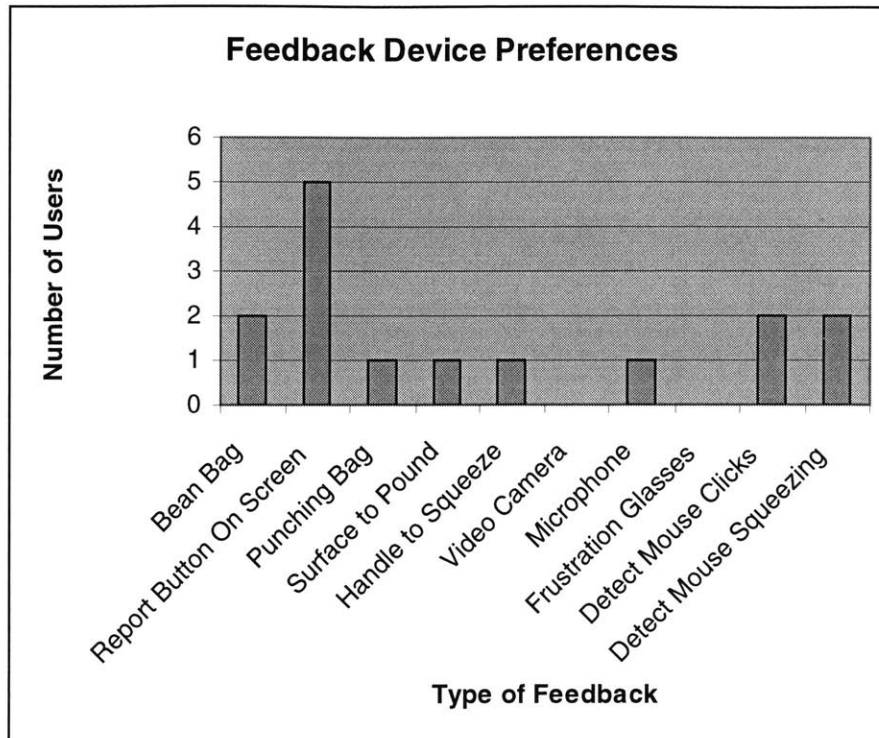


Figure 4.1 Responses to preliminary questionnaire for feedback device preferences.

4.12 Technical Requirements

The financial software runs over a web browser, using a secure connection via Kerberos encryption (Neuman 94). All users worked on Macintosh platforms, although they had the option of moving to Windows- or Unix-based stations if they so chose. Given this combination of details, the decision was made to design a feedback system that would run independent of platform and preferably in direct connection with the web browser to facilitate data monitoring and collection.

A major constraint on the project was the nature of the data involved in financial software transactions. The governing institution, understandably, had major concerns about the security of its financial transactions and the privacy of its financial data. The encryption scheme covering the software's network operations left very few options for monitoring user activity. Finally, a monitoring scheme was settled on which simply reported the frequency of user's complaints with timestamps, with the possibility of additional data supplied by the user. It was hoped that this approach would at least point out levels of user dissatisfaction with

the product, patterns of problems encountered, and the extent to which users were willing to use this kind of feedback system.

The initial designs built a client-server model which capitalized on existing technology and allowed for convenient user-end and designer-end adaptation. SAP has developed auditing software that keeps track of usage data on the server end of the interaction. This auditing software would have been used to keep track of user activity and react to it in ways specified by the designers, allowing the designers to refine server-end behaviors centrally or isolate specific users or classes of users for specific feedback schemes. On the client end, detection devices and software could be configured by the user, resulting in a personal environment suited to the style of a particular user and easily reconfigured if preferences changed. Unfortunately, the security scheme of the SAPWeb system precluded the use of the auditing software; the design model remains as a future project possibility.

4.13 Design Iterations

To build graphical user interfaces in a browser environment involving data transmission and storage, Macromedia Director (www.macromedia.com) was chosen as a development tool. In conjunction with limited CGI scripting for data transmission, the Director environment lent itself especially to the kinds of “game-like” feedback interfaces the users had requested in interviews as a relief from tedious and frustrating interface problems. Design iterations were guided by regular meetings and interviews with the users on what they wanted in an interface based on their typical moods when encountering HCI problems. Both individual and group meetings were held, and they served as a parallel feedback forum and brainstorming session as well as a collaborative user support environment as described in (Clement 90).

Since the financial transaction interface needed a significant amount of screen real estate, two options for situating the feedback system on the screen were considered: a separate browser window, opened using JavaScript upon opening a transaction, or a small frame in the same browser window as the transaction. Both

solutions were tested, although final designs all used the second solution with a very narrow frame at the top of the window to minimize crowding.

Early designs involved simple text prompts (“Click here to report frustration.”) which could be clicked with a mouse to activate a cartoon animation of a gloved fist punching a punching bag or a foot crushing a bug (the software “bug” in the system). Both were playful designs inspired by user recommendations, and designed on the principle that such cathartic scenes of violent release might be a way of venting frustration with a system, which has shown in some cases to be an effective stress management tool. This choice was made with reservations, however, since psychologists differ on the effectiveness of venting in emotion management (Goleman 95), some claiming that it actually makes subjects *more* angry. Nonetheless, the designs were developed in keeping with subjects’ requests under the assumption that users having some input into the design process would allow for the creation of a more agreeable system. Users were given both the option of choosing an animation upon starting a transaction or, in some versions of the system, an animation was chosen beforehand and placed in a separate widow. After activating the animation, the user was greeted with a text message acknowledging his or her feedback and thanking the user for helping with the design process. This acknowledgement was phrased in keeping with the principles of active listening and feedback to users addressed in (Klein 98).

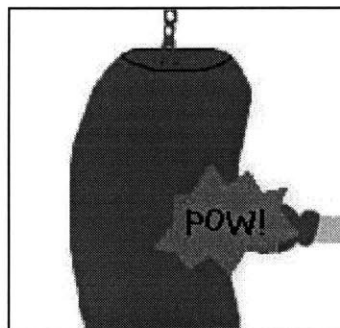


Figure 4.2 A frame in an animation from an early version of the feedback interface.

The final versions of the interface used a single strip along the top of the browser window to display a constant animation, another request from users. Animated

“bugs” crawled along the top of the browser window, running back and forth. When a user clicked on one of the bugs, a report would be registered and the bug would appear to be “stamped out”, its remains disappearing after a few seconds to be replaced by more bugs crawling around. The form of the “reports” varied by user: one user preferred to be able to immediately compose email in response to a perceived bug, while the others simply wanted the event registered for later collection. The timestamp of the report was written to a file on the local machine and collected at a later date. Unfortunately, the security constraints on the software prevented anything more informative than a timestamp to be recorded from a single click. Users were informed of this, which may have been a factor in their sporadic use of the feedback system as noted under Observations.

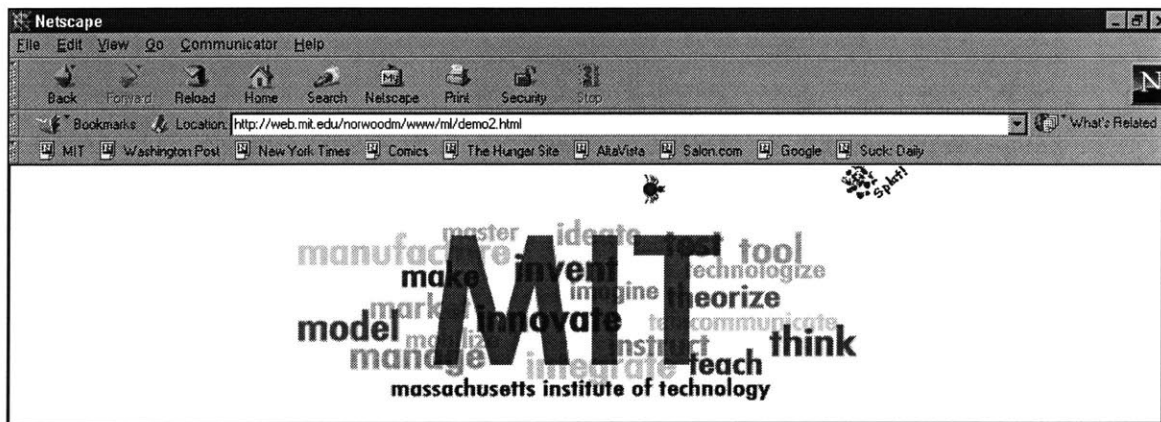


Figure 4.3) The final “bug-report” interface laid over the MIT home page. The bugs at the top crawl back and forth, “splatting” in response to being clicked.

4.14 Field Testing

The feedback agent was employed in the field for approximately six weeks in its final version. The one email feedback user of the seven didn’t use the feedback system at all, but data was collected from the other users. Unfortunately, the actual amount of data collected was insignificant – the causes behind the poor data yield and other problems in collection are discussed in the Observations section.

4.15 Follow-Up Questionnaire and Interviews

Following the six-week trial run, the users were given a follow-up questionnaire with questions similar to the first questionnaire. A copy of the follow-up

questionnaire with results is in Appendix B; this data is also discussed under Observations.

4.2 Case Two: Feedback System for a Dedicated Hardware-Based Service

After the conclusion of the first study, a second system was examined to shed light on a different context for automated feedback: an office coffee machine. The coffee machine provided an ideal test bed for an affective feedback system for a number of reasons. The communal nature of the machine meant that no one user was likely to take responsibility for managing its behavior independent of an automated system, and it would have a large population of users to provide feedback data. Users of the machine, a top-of-the-line multiple-function beverage dispenser, tended to get very attached to it and reliant upon its function; its frequent breakdowns left them anxious and frustrated. It was thus the focus for highly charged emotions, and an ideal candidate for affective feedback monitoring. Studying a coffee machine, a dynamic system whose function depends on regular maintenance and service, would allow an examination of the realm of service management and how it could capitalize on an automated feedback system. Finally, the physical nature of the interface would provide a contrast to the fully virtual nature of the first system studied.

4.2.1 Acorto and Mr. Java

The coffee machine in question, the Acorto Caffè 2000s, is a fully-automated espresso and cappuccino maker. It relies on a complex system of grinders, brew tanks, milk frothers, motors, sensors, and hydraulics, all controlled by a microprocessor running proprietary software, to deliver freshly, finely brewed coffee drinks and milk to users in a matter of seconds (Acorto 97). Its software system allows for service maintenance, allowing interface to an external processor controlling and monitoring the machine's behavior. In fact, such a project had already been implemented on the machine, a computerized drink monitor called Mr. Java (Kaye 00). Mr. Java kept track of drink orders on the Acorto and read data from specialized coffee mugs placed under the machine's spout, personalizing the

machine's behavior based on the information from the mug. Although the Mr. Java project had been dismantled at the time of the current project, the basic architecture – as well as the users' familiarity with the old system – was used as a basis for the new feedback system.

4.22 Objectives

The primary objective of the feedback system would be to correlate a rough measure of users' emotional reactions to the machine with the machine's behavior. Although monitoring usage data – as the original Mr. Java did – would give an idea of the service's popularity, direct affective feedback from users would allow much more precise problems to be identified. Specifically, different patterns of behavior and machine states might be linked with rates of user satisfaction, possibly pointing out specific machine states to be addressed in redesign or service.

4.23 Technical Requirements

The feedback system needed to incorporate two functions: interface with the user, and interface with the coffee machine. The first could be addressed with any of a number of affective computing tools, while the second could be accomplished using the Acorto's self-diagnostic capabilities and its serial port interface. The decision was made to build the entire feedback system around a separate computer installed in parallel with the Acorto, monitoring user data and Acorto data and relating them internally. This machine would also be responsible for displaying pertinent information to users of the service.

The Acorto runs a service port application which allows I/O through a serial port located on the front panel of the machine. A limited set of service commands are allowed, including prompts for internal state and monitoring drink orders. The full technical specifications for the 2000s software system are listed in (Acorto 97).

4.24 User-Feedback Design

After reviewing the data collected from computer users from Case One and taking into account the shared kitchen environment of the coffee machine, privacy was deemed an important factor in designing a feedback system. Various affective

metrics were considered, but were limited by a number of concerns. The device had to be reliable enough to work when unattended for days while being used by multiple inexperienced users. It was decided to allow users to remain anonymous, and also to give them the option of not providing feedback at all in order to preserve their privacy: the study was an intrusion into their familiar kitchen environment, and the device would have to be designed to avoid offending any potential users. In fact, this pointed out one of the key confounding factors in any study of user satisfaction: the administration of the measurement affects immensely the way the subjects feel about the service in question. Thus, every effort was made at each step of the design process to avoid annoying or frustrating users further.

A feedback mechanism was conceived using a simple “approve/disapprove” rating by having users press on one of two physical icons, or “phicons” (Ishii 97) shaped like a “thumbs-up” signal and a “thumbs-down” signal, on the machine’s surface. Data and prompts would be displayed to the user via a computer monitor mounted on the coffee machine at eye level. By allowing the user full control over what kind of feedback – if any – they presented, the device would be minimally intrusive to the current kitchen environment and users’ interactions with the Acorto. Also, using a physical interface would allow more natural feedback from the users and capitalize on the advantages of tangible interfaces (Gorbet 98). Finally, giving the users proper feedback in the form of acknowledgement of their report and verbal confirmation that their reports would be examined would encourage future feedback (Klein 98).

Simple feedback recorders were designed in the form of pressure-sensitive physical icons, which would register users pressing on them and gauge a response appropriate to the amount of pressure applied. For instance, a soft, brief press would presumably indicate a slight reaction, while a long, hard press on the phicon – or several presses – would indicate an extreme reaction. This feedback scheme also capitalized on the preference indicated during Case One to report feedback by pressure-sensitive devices.

To implement the pressure-sensitive phicons, a lightweight sensing architecture was employed in the form of pressure-sensitive resistors routed through an iRX board (Poor 97). The iRX board was equipped with a PIC chip programmed to

record presses on either of the phicons, distinguish between them, record the pressure, and pass the information on to the Java code running on the feedback machine processor. The pressure sensors were capable of distinguishing between fairly light presses on the phicons and maxed out near the limits of users' hand-pressing capabilities. The PIC code sampled the pressure-sensors at 100 Hz, summing recorded pressure over a 2.5-second interval from the beginning of a press. This sum was used as the overall strength of a user's reactions.

The feedback machine ran a Java application on a Windows 95 platform, reading input from the phicons and coordinating it with information from the Acorto and the information on the monitor. The user interface is displayed in Figure 4.3.

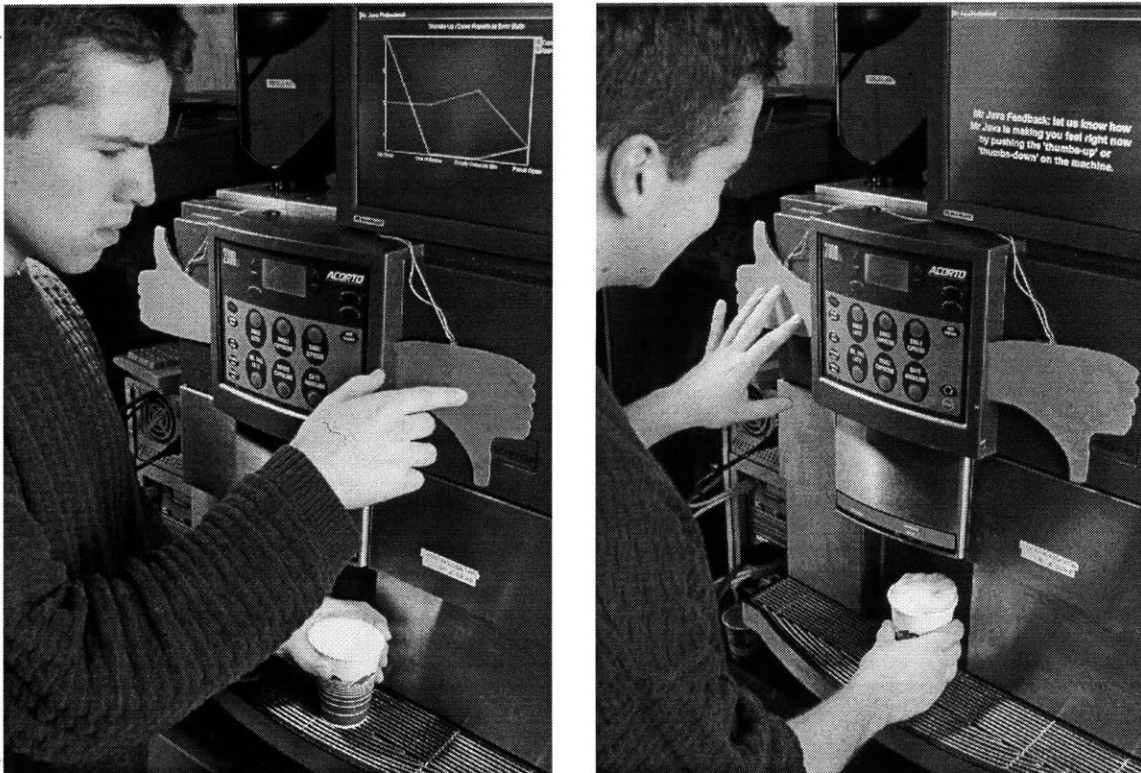


Figure 4.4 The Mr. Java Customer Support interface. The screen is situated at eye level, while the phicons are located next to the Acorto's control panel and above the spout.

4.25 Acorto Interface Design

The Acorto interface used Java code to communicate with the serial port on the service panel on the Acorto, synchronizing information from the coffee machine

with user input. The current state of the coffee machine was correlated with the current error state of the machine, and all drink orders were recorded.

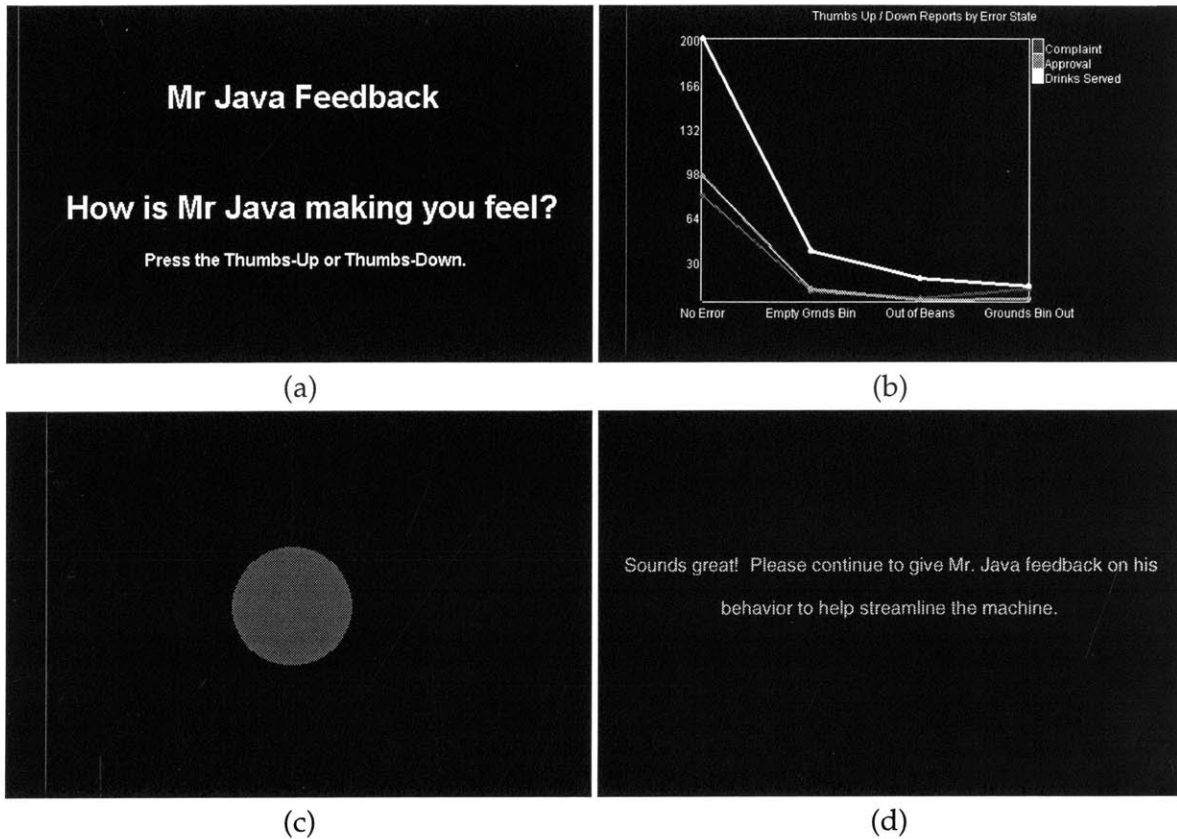


Figure 4.5 Screens displayed by Mr. Java during a user interaction. (a) and (b) alternated during idle activity, displaying a greeting and a graph of usage patterns. When users pressed a "thumb" phicon, they were greeted with visual feedback in the form of a circle of brightness and color corresponding to the strength and valence of the user's press, (c). Once the user removed his hand from the phicon, text appropriate to the intensity and valence of the report was displayed, (d).

4.26 User Display Design

The prompt for the user was designed to avoid the attribution to the machine qualities that it did not possess, such as “intelligence” or “awareness” of emotional states. User responses were greeted with an acknowledgement of the pressure and valence (approval or disapproval) of the report, along with acknowledgement that the feedback would be taken into account by the designers of the machine. A standard user display session is outlined in Figure 4.4. A list of Mr. Java’s display messages in response to user reports is provided in Appendix E.

4.27 Data Collection

The feedback system was put in place in the kitchen for approximately four weeks while data collection was in progress. During this time, the machine was functioning for approximately 20 days. The number of machine interactions – drink orders or attempted drink orders – along with the number of user reports, positive and negative, were recorded and categorized according to the error state of the coffee machine during the interaction. Unfortunately, data was not collected during periods of dysfunction, as the coffee machine needed to be running in order to interface with the feedback system. Periodic observations and maintenance of the machine and the feedback system were carried out during this period, and interesting deviations from normal pattern during the period were noted. The data is outlined in detail in the Observations section.

4.28 Final Questionnaire and Interviews

During the final days of the data collection, questionnaires were left for users to report their reactions to the feedback system during its tenure connected to the Acorto in the kitchen. Finally, individual perspectives were requested in the form of interviews with a few users on their reactions to the project. By these means, it was hoped to collect information on what worked and what remained to be redesigned in future systems soliciting user feedback. It was also hoped that the individual, detailed feedback might shed some light on interpreting the feedback data from the system.

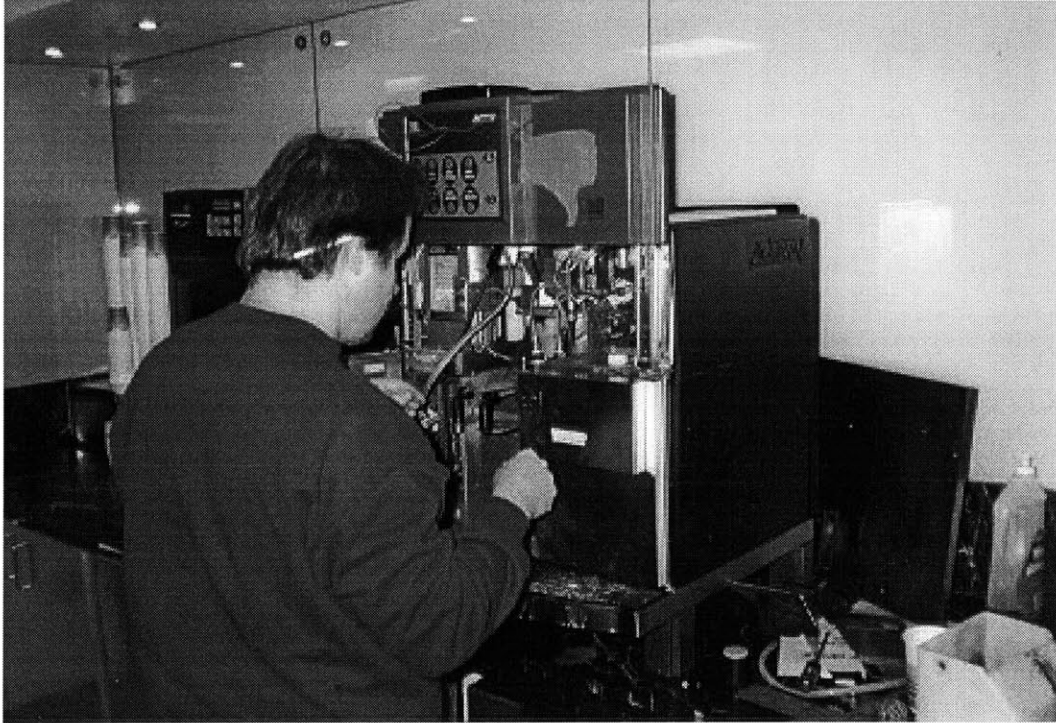


Figure 4.6 Dan, the region's Acorto technician, replacing a clogged milk valve on the machine.

5. OBSERVATIONS

5.1 Case One

5.11 Questionnaire

The data from the initial questionnaire in Case One is given in detail in Appendix A. However, a few pieces of information from the survey are interesting to note. The most commonly reported reactions to HCI frustration were yelling, leaning away from the computer, and clicking harder and more frequently on the mouse. The first two are events which might be monitored by microphones or video cameras, respectively, while the third could easily be monitored by a specially-designed mouse while preserving user privacy.

It is interesting to note the responses about the attribution of blame for frustration during HCI tasks. In *The Media Equation*, Reeves and Nass dismiss the idea that the emotional aspects of HCI are actually displacement of social attributes from the designers of an interface to the interface itself, but the issue may in fact be

more complex than this (Reeves 96). While most computer-elicited frustration is reportedly blamed on the computer itself or the software, in the case of the specific SAPWeb financial software the frustration was often attributed to the administrators who made the decision to adopt the new system. Interviews and meetings with users revealed quite a bit of dissatisfaction with the performance of the current financial software compared to the performance of the software it replaced. Users resented the replacement of what was, for them, an easy-to-use and functional application with a harder-to-learn, less-intuitive interface. They reported being more likely to become frustrated with the new interface when it failed a task that was straightforward in the old interface. Thus, administrative decisions and workplace context played a large role in eliciting HCI frustration from users, revealing a possible complication often overlooked by HCI designers: the limitations of controlled laboratory conditions in identifying potential sources of user frustration. It was hoped that the automated feedback system, running in an actual workplace environment, would be able to capture this data more accurately than psychological studies in a lab.

5.12 Feedback Data

Very little data was collected from the feedback system in Case One. Of the five users remaining at the end of the study, only two recorded any data beyond the initial test data during setup. Of this data, the only significant observation is a rough pattern of use where several reports were made in what appeared to be single sessions. The data gathered by the feedback system was not valuable at all for analysis of the software being studied.

More interesting than the recorded patterns of use were the explanations given in interviews and meetings for the lack of use. One user reported having lost track of how to open the SAPWeb filtered by the feedback agent, so had instead used the default web page. Another had the same problem and also declared that she found the animation too distracting to want to use it. A third user had made major overhauls to her computer system during the study, wiping out any stored feedback data, although she said that she had made at least some use of the bug-stomper.

In short, the data collection phase of this case revealed much more about the inadequacy of the feedback system and its implementation in the workplace than it did about the software being studied. During group meetings about the software and the feedback system, however, a forum was created where users could collaborate on overcoming interface obstacles with the product. This pointed out again the importance of integrating social forums for feedback and collaboration in parallel with any automated system.

5.13 Interviews / Group Meetings

Quite a bit of useful feedback on the SAPWeb interface was collected through user interviews and group meetings. Issues ranged from GUI comments like window operations, visual clutter, dialogue behavior and phrasing, to application-specific comments about the presentation and of data and behavior of the application in handling financial data. Also, typical work patterns affected the efficacy of the interface in many cases, such as a complaint that users who spend most of their time with a web browser open and little time with a specific application open are very unlikely to use anything but the functionality present on the Web.

Overall, the feedback reported directly by users in person was much more specific than the feedback collected by the feedback system. Had the system been less limited by security constraints and more widely used, however, it might have been able to provide insight into precise sequences of interaction impossible to illustrate in a conversational medium and quantifiable measures of in-context user frustration. The two feedback forums – automatic and personal – might be able to work together to provide designers with a comprehensive picture of directions to be taken in interface refinement. The feedback from SAPWeb user interviews and meetings is presented in detail in Appendix C.

A few users made comments during group meetings that pointed out an indirect benefit of the feedback system. Users mentioned having been annoyed by something specific about the SAPWeb software – losing data to a resized window, for example, or having a line-item fail to duplicate (see Appendix C) – and having clicked on the “bug-report” animation in response. Even though their direct reports to the software were incapable of registering the context of their complaints, they

seemed much more likely to remember the exact nature of the situation in group meetings after having used the feedback system. The feedback system seemed to act as a kind of ritual to fix the exact nature of the complaint in a user's memory, indirectly facilitating feedback.

Interestingly, follow-up interviews indicated that the subjects were less averse to being monitored on video or microphone than they were at the beginning of the study. One user remarked that using the feedback agent acculturated her to the idea of sending continuous feedback, and that a video or microphone would be an acceptable form of feedback as long as it was only active during SAPWeb sessions and it didn't keep a record of anything but specific feedback data. Another user professed openness to the idea of being monitored on a microphone under similar restrictions, although she admitted that many users would feel intruded upon in such a case. These testimonials suggest that unfamiliarity with the idea of limited workplace monitoring is a major factor in the opposition to such feedback measures, although it should be kept in mind that users were responding based on their completed roles in a pilot study, not a permanent workplace innovation. It should also be noted that neither video nor audio recordings were named as preferred methods of feedback: the data on the follow-up questionnaires confirms this. Still, such measures might be feasible as long as absolute anonymity was maintained at all times and recorded data was severely constrained in its content.

5.2 Case Two

5.21 Data Collection

Since the design of the feedback agent in Case Two was more straightforward and located in conjunction with a familiar interface, the prototype system was put in place directly to collect user feedback without a preliminary questionnaire. User feedback through alternative channels (e.g., email) was encouraged, however, and that data is examined in the next section.

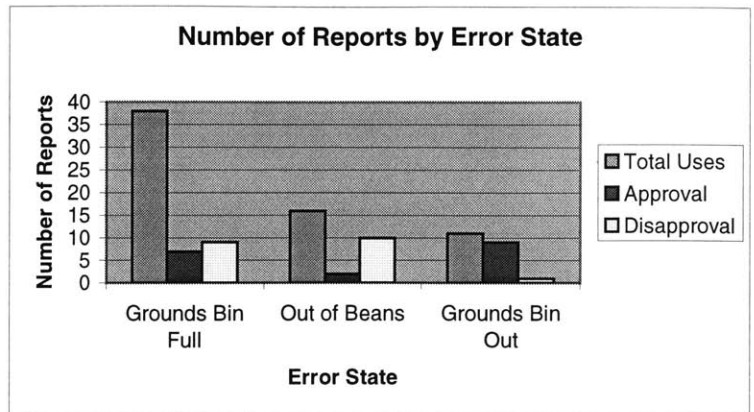
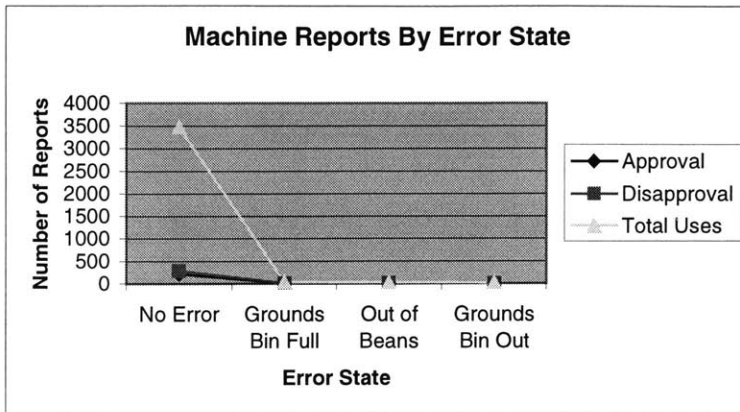


Figure 5.1 (a) A graph displaying the total number of user interactions and reports in each error state examined over the 20-day data collection period. Note that the reports and interactions in the “No Error” state overwhelm the others. In (b), the No Error state is eliminated from the analysis, showing the total counts of reports and interactions in the other states.

The data collected from the feedback system produced several interesting results. The most obvious observation, as is evident from Figure 5.1, is that most user interactions were enacted while the machine was in a “No Error” state. This suggests that the machine spent most of its time in this state. This could be confirmed by observing the machine casually over a period of weeks, but it is interesting to note that an automated feedback system could monitor the amount of time a given machine spent in an error-free state, automatically providing a service provider with valuable information about possible maintenance needs or possible misuse of the machine. This data could then be made instantly accessible over a network connection, allowing the service provider to prioritize service calls and reevaluate service contracts based on the amount of time a given machine spends in each error state. The error states examined in Figure 5.1(b) suggest that the “Grounds Bin Full” state was common, but a possible confounding factor is that that state does not interfere with drink orders, in contrast to the other two error states examined. Thus, users may have been more likely to interact with the machine in this state because they were still able to order drinks, which they wouldn’t have been able to do in the other two states.

Error State	Approval	Disapproval	Total Uses
No Error	216	290	3474
Grounds Bin Full	7	9	38
Out of Beans	2	10	16
Grounds Bin Out	9	1	11

Figure 5.2.) Numerical data on reports and interactions by error state.

Another interesting piece of data is the tendency of users to report on the machine's behavior more or less often depending on the state of the machine. Figure 5.3 summarizes this data by showing rates at which interactions resulted in user reports for different states. It also shows more clearly the data of primary interest: rates of approval and disapproval in various states.

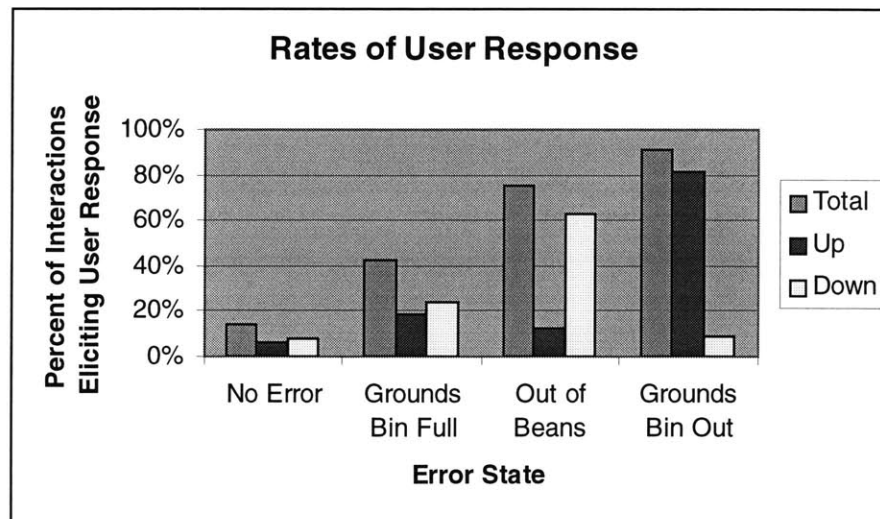


Figure 5.3) Rates of user response by error state.

A large number of "No Error" interactions went without user feedback, as seen by the low level of responses, and the negative responses only slightly outweighed the positive responses. These negative responses may have been due to a number of factors. One possibility is that the machine was behaving improperly and the error

was not detected by the diagnostic systems of the Acorto. Although the machine has extensive self-diagnostic capabilities, it is incapable of detecting, for instance, when the milk supply runs out. A second possibility is that users were displeased with the general behavior of the machine – the quality of the coffee, for instance. A third possibility, and one that was confirmed by later surveys, was that users were displeased with the new addition to the machine – the feedback display itself – and took it upon themselves to report their disapproval of the intrusion by reporting negatively on the system itself instead of (or in addition to) reporting their grievance through other channels.

The “Out of Beans” state yielded the highest ratio of negative-to-positive user reports. This may have been explained by the following peculiarity of the state: while a full or removed grounds bin is easily remedied by a few moment’s adjustment, an “Out of Beans” state requires several minutes of tedious work to remedy. In addition, there were occasions during the study when the local supply of beans was completely exhausted, leaving even experienced users with no possibility for a cup of coffee. Thus, although inexperienced users may have encountered some distress when running across the other error states, even experienced users were frustrated sometimes in the “Out of Beans” condition. This finding points out a possible re-prioritization for the facilities managers in charge of the Acorto: to make reliable bean supply a higher priority than basic machine maintenance. The rationale here is that while many users might be content to adjust a misbehaving machine, they are not in a position to supply beans for the machine if it runs out.

Finally, an oddity in the collected data is the high rate of approval in the “Grounds Bin Out” condition. Since the total number of interactions is so low here (11 interactions), this isn’t as odd as it looks. Only a few “positive” reports (nine) were made, and they might well have been the result of users playing with the system while the grounds bin was out. It is actually unlikely that the bin would have been out while users were attempting to use the machine, so this scenario seems all the more likely. If it were out, replacing it would have been a trivial task, so users might have replaced it and given the machine a “thumbs-up” report afterwards. All in all, this data category should not be taken to represent actual user

satisfaction with the machine in this state, except to point out that while the machine is in this state, no one seems to be particularly inconvenienced.

5.22 User Questionnaire And Interviews

At the conclusion of the study, questionnaires were supplied next to the machine for voluntary, anonymous comments on the system. A copy of the survey is presented in Appendix D. The ten user responses were rather divided on their opinion of the feedback system. The results are summarized in Table 5.4.

Can see Mr. Java reacting:	Clearly 20%	Sort of 40%	Not at all 40%
The graph on Mr. Java is:	Interesting 30%	Confusing 50%	Annoying 30%
Have privacy concerns:	Yes 0%	Sort of 10%	No 90%
Like the greeting on the monitor:	Very much 0%	Sort of 50%	Not at all 40%
Like the text feedback to the user:	Very much 10%	Sort of 20%	Not at all 50%

Table 5.4 Results from Mr. Java anonymous user questionnaire

The first question addressed a possible problem resulting from the placement of the monitor and the speed of the feedback to the user: either the user might not be able to read the monitor due to its angle, or the system might not respond quickly enough to display feedback to the user in time. The first was a real possibility because the angle of the monitor was often readjusted by users passing by, which might have made it impossible for users of another height to read. The second was a concern because the system was designed to make sure that a user had stopped pressing on a phicon altogether before displaying feedback text. This may have appeared to some users as a stalled or non-responsive system. Apparently, many users experienced one of these problems. It should be noted, however, that users with negative experiences were probably more likely to report them, leaving the 40% non-visibility rating a high estimate.

The other questions were more aesthetic and affective in nature. Privacy didn't appear to be a concern, but the aesthetics and sensibilities of the interface and the data displayed seemed to grate on some of the respondents. Again, negative reactions were more likely to be reported, but the degree of annoyance reported on

some of the forms seemed to suggest that a major overhaul of the system would be required if it were to be put in place permanently. In particular, users commented on the unsightliness of the computer CPU sitting next to the coffee machine, which suggested that a clean, elegant form for any add-on to an existing system would be needed to prevent aggravating current users. Several of the respondents' comments also indicated that they had used the feedback system while it was out of commission, as happened several times during its installation: an unfortunate consequence of installing directly onto a system currently in use. This data, however, pointed out the criticality of having the system perform as advertised as soon as installed and having it run smoothly for the duration of its installation: the users were very much annoyed at a problem-reporting system being out of commission because of problems of its own. Other users had very negative reports, but they seemed tied to frustration at the coffee machine itself not working. This suggested that users rarely discriminate between different parts of the offending system. This means that the task of isolating the specific subsystem causing the frustration becomes a much more complicated task, and a self-monitoring system would be required to account for the possibility of the feedback system itself being the cause of the frustration.

Two users volunteered to be interviewed in more depth about their experiences with the machine. Both of them had similar comments to make: that the binary feedback to the machine was too limited, and that they wanted more specific complaints to be logged. The alternative forum of email was deemed too much of a hassle for immediate grievances against the coffee machine, which was located on the other side of the building from their computer terminals. Their complaints pointed out the need for either a more complex set of user response choices – probably outweighed by the possibility of user confusion – or a more convenient forum for reporting detailed complaints.

6. CONCLUSIONS AND FUTURE WORK

6.1 Feedback Data On Two Cases

The Web-based software system and the coffee machine had some shortcomings in their ease of use and their function, as could have been expected. Examining the systems and soliciting user feedback shed some light on these problems, as the specific data revealed in the last section and the appendices illustrated. The automated feedback systems caught some useful data: for instance, the tendency for the “Out of Beans” state to particularly irritate users. The bulk of the feedback, however, was gathered through conventional means: surveys, questionnaires, memos, interviews, and meetings. The project suggests that while automated feedback agents may be valuable in collecting usability data, more conventional methods should hold sway until feedback designs are refined. The most valuable data provided by this project, then, is a collection of seven principles to keep in mind for future automated feedback design.

6.2 Lessons Learned About Building Automated Feedback Systems

6.21 Coordinate With Social Networks

Social, shared-space environments seem to be unparalleled for versatility in problem-solving. Social collaboration allows for experience and ideas to be shared effectively and quickly, and provide a wealth of user support: technical, emotional, and administrative. In addition, they empower end-users by demystifying technical products through group examination and shared experience. Social collaboration should be preserved and fostered in parallel with any automated feedback agent to provide additional support for users and to create a forum for monitoring the effectiveness and context of the feedback data.

6.22 Aesthetics: Keep It Clean

The look and feel of feedback interfaces and their physical support hardware was a major topic of user feedback on the systems. Users stated, more frequently than

any other preference, a desire for clean and clear feedback systems. Sloppiness on the part of the feedback interface generated immense dissatisfaction from users. Any feedback added on to an existing system is an intrusion into a familiar environment – real or virtual – and should be undertaken with an eye to blending in and complementing the existing elements familiar to users.

6.23 Reliability: Keep It Consistent

Inconsistent behavior was the second most common complaint from users. If a feedback system is put in place, it is vital that its design, implementation, and maintenance sustain consistent, meaningful behavior in the interface. A lack of faith in the feedback interface erodes users' willingness to interact with a new system, just as the interface flaws in the software from Case One eroded user confidence in the new software system. Users should be able to trust that the new feedback system is reporting their data reliably, or they may abstain from using it altogether and see it as nothing more than a nuisance in an already problematic environment.

6.24 Avoid Invasion of Privacy

User privacy is a major concern in an age where personal information is distributed among corporations and available through an ever-increasing number of venues. Many users have become understandably defensive about further encroachments into their private lives and their personal behavior, and few are willing to have their behavior monitored in any way not directly in line with their job descriptions. Thus, feedback systems should avoid the issue altogether whenever possible. After being debugged on small groups of user-subjects, they should draw anonymous data from large populations and preserve the option for users to abstain from giving feedback at all. Along with aesthetics and inconsistency, privacy is one of the touchiest issues in HCI. In adding a system on, the same rule can be applied to privacy as has been applied above to aesthetics and consistency: do not disrupt the original environment's aesthetics, do not disrupt the original environment's behavior, and do not disrupt the original system's level of privacy.

6.25 If It's Distracting, Make Sure It's Fun

A possible exception to the “do not intrude” mantra for the last three points is in the case of a feedback environment designed to be enjoyable in and of itself. In Case One, users requested a game-like feedback interface that would take their minds off of the frustration-eliciting behavior of the software and allow them to relax during a frustrating session. Similarly, many users in Case Two found the idea of hitting a “thumbs-up” or “thumbs-down” pad a diversion from the standard workaday routine of drinking coffee and sitting in front of a terminal. These reactions highlighted one of the psychological foundations of the project: the idea that, in order for users to evaluate an interface, they have to stand back from it and break out of the “user” mentality, adopting instead a broader view of the system. Games and playfulness are one way we maintain perspective on our environment. A friendly, informal relationship between users and the feedback device would allow users to report feedback in the same mentality they use when they joke about the system’s failings with their colleagues. By fostering this kind of “non-professional” relationship with users, a game-like, approachable interface might allow for more open, more insightful feedback. It would avoid the trap of having users “tell us what we want to hear”, letting them instead give free reign to their frustrations without cloaking them. Reeves and Nass discuss the tendency of users to be more “polite” in their interactions with a computer system when reporting its shortcomings than they are when reporting on the first computer to another computer (Reeves 96). By providing a “personality” for the feedback system which differs significantly from the style of the original system, this effect might be capitalized upon to encourage frankness.

6.26 Provide Convenient Alternative Feedback

In the case of feedback system inadequacy – whether due to design flaws, breakdown, or dislike by one or more users – an alternative feedback medium should be provided conveniently and clearly. This could mean incorporating an email system into the device or interface itself, or referring users in a desktop environment to a web page or phone number. The key here is to supply the communication medium *immediately* in order to capture the users’ initial reactions

the same way that the feedback device would. If the feedback system operates on a desktop computer, provide Internet- or phone-based options; if the system is connected to a public piece of hardware, provide a phone line for communications; if the system is installed near the designers' physical location, encourage users to report problems in person. The option should be preserved for users to remain anonymous, and for them to leave the complaint behind after they make it instead of being obliged to complete a follow-up interview, which might make the experience more trouble than it's worth and discourage future feedback. The alternative medium should be clearly denoted as a valid and desired way of communicating feedback so users will be encouraged to use it liberally: it is important to catch more than the extreme reactions to the system in question.

6.27 Collect Detailed Data

One of the most important lessons learned in retrospect on this project is the importance of collecting as much information as possible. Before observing usage patterns, it is difficult to assess what data might be important and what can be left out. For instance, in the case of the coffee machine, finding a way for the feedback system to collect data when the coffee machine was deactivated might have resulted in some very interesting data about user frustration when no coffee service is available at all. In addition, some data channels might not be effective, and it is important to provide as much data as possible to take up the slack for sources that turn up empty. That being said, the need to collect detailed data must be weighed against the concerns mentioned above about intrusiveness and privacy. An ideal system would minimize intrusiveness in all areas mentioned, while still monitoring as many different parameters as possible in the system and in the user's behavior.

6.3 Directions for Future Work

6.31 Technical Support

A highly promising tie-in for a feedback system would be context-specific technical support. By keeping track of the specific software context and machine state when a user reports an error, such a feedback system would be in an ideal

position to offer specific help to a confused user. Unclear aspects of the product or interface might be expanded upon when a user reports frustration with them.

Alternatively, a human technical support technician could receive a detailed report of the machine or software's behavior leading up to the grievance, allowing a much more efficient approach to trouble-shooting.

6.32 Emotional Management Support

Another possibility for customer support tied to customer feedback is the notion of emotion management support. As examined by Klein and Picard, this kind of support could not only result in happier and more productive users, but it could guide the feedback system in classifying the customer's complaint. Part of the active-listening approach to automatic emotion management support is a dialogue which tries to echo back to the user an evaluation of what the specific problem might be, with sensitivity to how the user is responding to it emotionally. This information, built in conjunction with a feedback system, would potentially provide a wealth of information to developers about specific patterns of interaction and their emotional consequences for users, while at the same time letting the user know that their specific concerns had been heard and were being addressed.

6.34 Networked Service Management

By adapting a feedback system to a regularly provided service – such as the coffee machine in Case Two – a service provider could make informed decisions about service management. If few users seemed positive about the service and many had complaints, the local manager might decide to discontinue its use or employ another service. A service technician could monitor the data via a network and prioritize service calls based on something more accurate than a phone call from a customer. If the final goal of a service provider is customer satisfaction, this value could be maximized much more accurately using a measure of users' emotional reactions to the service. For instance, a heavily-used and highly-valued coffee machine might receive immediate service for even minor disruptions in function, while an infrequently-used, peripheral coffee machine might be put on the back burner for service calls. The service provider could thereby prioritize their resources

to maximize actual customer satisfaction. In addition, the data from the feedback device might provide the technician with information about the technical nature of the problem by merging self-diagnostic data from the machine and outside reports from users to catch events missed by the diagnostics. The key here is integration of data sources and networking: instead of an isolated machine and isolated complaints from users, often days after the errors occur, a service provider would have access to real-time data from both sources, available remotely over a network.

6.35 Encourage HCI Measurement Standard For Competitive Business Practice

If automated feedback systems were to gain acceptance on a large scale, they might be able to bring about a revolution in usability standards for design. By adopting standard measures of user satisfaction into every software package and service on the market, a quantifiable measure of usability would be available to consumers in making decisions about product or service purchases. The kind of consumer reporting done through magazines and web sites like Deja.com would be disintermediated by allowing users access to direct statistics on various products' usability and customer satisfaction rates. Companies could advertise the high rates of satisfaction with their products or services, encouraging competitive development of usable interfaces and high-quality service. Of course, this kind of adjustment to the market would only be possible if an industry-wide standard for usability measurement were adopted – it would be horrendous for the industry and for users if negative feedback were discouraged in order to make the product look good.

6.36 Integration with other features

Detailed feedback might be enabled by building up more detailed user profiles and increased networking of information. The quality of the feedback is directly affected by the amount of information that can be provided about the context of the interaction; thus, personalization features such as Mr. Java's old identification function could be tied in to provide more meaningful feedback. Also, the proliferation of networks including appliances and non-desktop processors lends itself to developing this kind of accessible feedback in non-desktop environments and in previously non-

networked applications within desktop environments. Integration of a feedback system with this increased network proliferation would greatly increase its utility.

6.37 Adaptive Interfaces

Some promising work has been done with user interfaces that respond automatically to user preferences. While many applications now have personalized profiles for user interfaces, such as web browsers and operating system desktops, changes in these environments remain possible only by explicit user command. Preliminary work has begun in the Affective Computing group at the MIT Media Lab to develop interfaces that respond automatically to user reports of frustration. For instance, at the October SENS*BLES symposium in Cambridge, a modified version of the Microsoft Office paperclip "Help" agent was presented that displayed this kind of behavior. When a user showed signs of frustration with the agent by squeezing a pressure-sensitive mouse or yelling into a computer-mounted microphone, the paperclip volunteered to shut itself off. This was an example of an interface adapting to user preferences without being told explicitly to do so. The work is motivated by the idea that no interface will work perfectly for all users, and an interface should allow for multiple user styles in its design. This is a variant on the designer feedback paradigm: to design a software system so complex that it is capable of taking possible "flaws" – features not desired by one or more users – into account and "redesigning" (i.e., reconfiguring) the software on-the-fly. It would use many of the same feedback mechanisms as a feedback agent, and might therefore be integrated into the same system with fewer complications.

6.38 Networked Auditing

Some work environments need to keep track of user data for purposes other than usability feedback. Records of past interactions are often kept for network applications to simplify data warehousing and provide technically valuable data for error recovery or security. In an such an application with an existing auditing system in place, implementing a feedback system would be vastly simplified. In addition, the feedback data might help to integrate the data currently being collected, providing more detailed context for the audited data. For instance, a

security breach might be more easily detected if a system appeared to operate correctly internally, but was accompanied by user reports of strange behavior.

6.39 Synthetic Characters / Personalities

In keeping with the social, and often playful, nature of the feedback system design ideas, the integration of a synthetic character with its own “personality” might enhance the relationship between the user and the feedback system. Some work has been done in designing synthetic characters or conversational interfaces with personalities that match the user’s mood (Klein et al 97, “TurnStyles” and “Bruzard”); this kind of empathy would be vital for building the kind of non-professional, honest emotional relationship the feedback system would require. Again, the warning here is the same as that reiterated through this section: avoid unwanted intrusion into the user’s environment. A synthetic character might be very popular with some users, but others will invariably find it irritating. The control of users over their environments is key in avoiding the frustration the system would ultimately be trying to remedy.

CLOSING COMMENTS

Automated user feedback systems are a promising development of an increasingly fast-paced, networked, automated work environment and could eventually have spectacular effects on the way usability and quality of service is addressed by designers, businesses, and users. However, implementation of such systems has proven non-trivial, and a number of factors will have to be resolved before they become useful. They must be incorporated into a holistic design concept, as part of the system on which they report. Although this is true of all new features, natural incorporation of affective feedback systems is even more critical due to the highly charged circumstances in which they are likely to be used.

Although there is currently no substitute for human-managed feedback, automated feedback systems have shown some promise already as tools to be used in parallel with direct human communication. The findings of this study reveal the need for further investigation of such systems, which will reveal the extent of their eventual impact on design and market paradigms.

APPENDIX A: Network Software Feedback Questionnaire With Results

Results are from five users. Number of responses and percentages are given except in the case of "ranked" responses, which are scored in aggregate.

Questionnaire

1.) How often do you find yourself frustrated while using computer applications?

0%	0	Never
22%	2	Almost never
56%	5	Sometimes
11%	1	Often
11%	1	Very Often
0%	0	Always

2.) When you do become frustrated while using a computer application, what are the targets of your frustration? (rank the items from 1 to 6, with 1 next to the target toward which you are most likely to direct your frustration.)

- 2 The Computer
- 1 The Software
The programmers
- 3 The Company which made the software
Your employers/supervisors
Yourself
Other

3.) When you become frustrated with a computer application, do you tend to: (check all that apply)

78%	7	Yell / grumble / curse at the computer
0%	0	Hit the computer / bang the keyboard
56%	5	Click mouse harder / more frequently
11%	1	Lean forward toward the screen
67%	6	Push back away from the computer
11%	1	Other

4.) How long have you been using SAPWeb?

22%	2	<1 month
44%	4	<1 year
33%	3	>1 year

5.) How many times per week do you use SAPWeb?

0%	0	less than once / week
44%	4	1-5 times / week
22%	2	6-15 times / week

33% 3 16+ times / week

6.) How often do you find yourself frustrated while using SAPWeb?

0%	0	Never
0%	0	Almost never
44%	4	Sometimes
22%	2	Often
22%	2	Very Often
11%	1	Always

7.) When you do become frustrated while using SAP, what are the targets of your frustration? (rank the items from 1 to 6, with 1 next to the target toward which you are most likely to direct your frustration.)

The Computer

2 (tied)	The Software
	The programmers
1	The Company which made the software
2 (tied)	Your employers/supervisors
	Yourself
	Other

8.) When you become frustrated with SAP, do you tend to: (check all that apply)

66%	6	Talk / grumble / curse / yell at the computer
0%	0	Hit the computer / bang your desk
22%	2	Click mouse harder or more frequently / bang on mouse
0%	0	Type harder / bang the keyboard
22%	2	Lean forward toward the screen / focus on the screen
44%	4	Push back away from the computer
44%	4	Tense up your muscles
56%	5	Scowl at the computer
44%	4	Try to get away from your computer / desk
22%	2	Other

9.) If you could express your frustration with the application some other way, which of the following might you use:

22%	2	A beanbag to throw / squeeze
56%	5	A "frustration report" button on the screen
11%	1	A "punching bag" animation on the screen or something similar
11%	1	A surface to pound on your desk / monitor
11%	1	A handle to squeeze (free-floating or attached to your chair or desk)

0%	0	A video camera to monitor your visual frustration
11%	1	A microphone to capture your verbal frustration
0%	0	A pair of glasses which would record your muscle tension
22%	2	A mouse which would record the pressure and frequency of clicks
22%	2	A mouse which would record tension in your hand
44%	4	Other

10.) If your experiences with and reactions to the SAP web interface and the Sap GUI are different, please characterize the differences:

APPENDIX B: Network Software Follow-Up Questionnaire With Results

Results are from five users. Number of responses and percentages are given except in the case of "ranked" responses, which are scored in aggregate.

SAPWeb Follow-up Questionnaire

1.) How many times per week have you used SAPWeb since April?

- | | | |
|--------------------------|---|-----|
| a. Less than once / week | 2 | 40% |
| b. 1-5 times / week | 3 | 60% |
| c. 6-15 times / week | 0 | 0% |
| d. 16+ times / week | 0 | 0% |

2.) How often do you find yourself frustrated while using SAPWeb?

- | | | |
|-----------------|---|-----|
| a. Never | 0 | 0% |
| b. Almost never | 1 | 20% |
| c. Sometimes | 3 | 60% |
| d. Often | 0 | 0% |
| e. Very often | 1 | 20% |
| f. Always | 0 | 0% |

5.) When you do become frustrated while using SAPWeb, what are the targets of your frustration? (rank the items from 1 to 6, with 1 next to the target toward which you are most likely to direct your frustration.)

- | | | |
|----|----------|-------------------------------------|
| a. | <u>4</u> | The Computer |
| b. | <u>1</u> | The Software |
| c. | <u>3</u> | The programmers |
| d. | <u>2</u> | The Company which made the software |
| e. | <u>5</u> | Your employers/supervisors |
| f. | <u>6</u> | Yourself |
| g. | <u>5</u> | Other: <u>Accounting</u> |

6.) When you become frustrated with SAPWeb, do you tend to: (check all that apply)

- | | | | |
|----|----------|-----|---|
| a. | <u>3</u> | 60% | Talk / grumble / curse / yell at the computer |
| b. | — | 0% | Hit the computer / bang your desk |
| c. | <u>2</u> | 40% | Click mouse harder or more frequently / bang on mouse |
| d. | — | 0% | Type harder / bang the keyboard |
| e. | <u>1</u> | 20% | Lean forward toward the screen / focus on the screen |
| f. | <u>3</u> | 60% | Push back away from the computer |
| g. | <u>2</u> | 40% | Tense up your muscles |
| h. | <u>3</u> | 60% | Scowl at the computer |
| i. | <u>1</u> | 20% | Try to get away from your computer / desk |
| j. | — | 0% | Other _____ |

7.) Did you ever use the "bug-report" screen with SAPWeb?

- | | | |
|--------|---|------|
| a. Yes | 5 | 100% |
| b. No | 0 | 0% |

8.) How convenient was the bug-report interface to use?

- | | | |
|--------------------------|---|-----|
| a. Very convenient | 2 | 40% |
| b. Somewhat convenient | 1 | 20% |
| c. Somewhat inconvenient | 1 | 20% |

- d. Very inconvenient **0** **0%**
- e. No opinion **1** **20%**

9.) How effective did you find the bug report interface in making you feel better about your SAPWeb session?

- a. Very effective **1** **20%**
- b. Somewhat effective **2** **40%**
- c. Somewhat ineffective **2** **40%**
- d. Very ineffective **0** **0%**
- e. No opinion **0** **0%**

10.) Did you think of the bug report interface you used as more experimental or more practical?

- a. Very much experimental **2** **40%**
- b. Somewhat experimental **0** **0%**
- c. Some of each **2** **40%**
- d. Somewhat practical **1** **20%**
- e. Very experimental **0** **0%**
- f. No opinion **0** **0%**

11.) Do you feel that you would be more inclined to use the bug report interface if there were some kind of incentive: discounts online, software upgrades, etc.?

- a. Very much so **2** **40%**
- b. Somewhat **0** **0%**
- c. Perhaps **3** **60%**
- d. Probably not **0** **0%**
- e. Definitely not **0** **0%**

12.) Do you feel that you would be more inclined to use the bug report interface if you were sure that the company was acting directly on the data you sent?

- a. Very much so **3** **60%**
- b. Somewhat **1** **20%**
- c. Perhaps **0** **0%**
- d. Probably not **1** **20%**
- e. Definitely not **0** **0%**

13.) Do you feel that you would be more inclined to use the bug report interface if there were some kind of instant feedback from a person – a tech support staffer or financial officer?

- a. Very much so **2** **40%**
- b. Somewhat **1** **20%**
- c. Perhaps **1** **20%**
- d. Probably not **1** **20%**
- e. Definitely not **0** **0%**

14.) Having now interacted with one form of feedback system, what would you choose as the ideal way to send feedback to a company?

- a. Something similar to the bug report: an animation on the screen. **2** **40%**
- b. A beanbag to throw / squeeze
- c. A “frustration report” button on the screen **3** **60%**
- d. A surface to pound on your desk / monitor
- e. A handle to squeeze (free-floating or attached to your chair or desk)
- f. A video camera to monitor your visual frustration
- g. A microphone to capture your verbal frustration
- h. A pair of glasses which would record your muscle tension
- i. A mouse which would record the pressure and frequency of clicks **1** **20%**
- j. A mouse which would record tension in your hand **1** **20%**
- k. Other _____

[Note: Some subjects made multiple responses to question14.]

15.) For the choices you made above, please provide a brief description of why you feel the way you do about the various choices:

Responses to question 15:

The bug was a handy way to entertain yourself when you were frustrated. Having an email button to report feedback to someone at the company who would actually read your comments seems like it would be the most helpful.

I liked to be able to give feedback.

The mouse is my primary interface to using SAPWeb – so that would be the most effective way to register feedback (especially since I click my mouse hard and frequently when I'm frustrated)

*Something about a button to click that would send a signal somewhere seems more "productive" of a way to express frustration than the other ways... the mouse recording tension would be a great way to report **frustration** that doesn't require an **action** on my part.*

APPENDIX C: Data From Interviews: Case One

Complaints and observations are broken down by area of functionality addressed.

Window Operations

Dialog boxes and pages all blend together because of their uniform visual style.

-Suggestion: color-code windows and pages by function: Req forms, Inquiries, etc.

Certificate box pops up for every operation: it's big and slow.

-Suggestion: can there be a once-per-session security check? Or at least accompany the pop-up window with sound to indicate something new while you're typing?

Resizing window loses all data.

No line items in workflow box: have to switch pages to check.

Database Operations

Roles Database has locked out users in the past with no way to restore their privileges without creating a new user profile.

-Media Lab has particularly bad interface with the Roles DB

Logic behind Roles Database is unclear: what does the name mean?

Lots of authority taken away in Roles DB: you can do your own orders, but you can't grant access to other users.

Security procedure seems inconsistent.

Missing Functionality

No online source selection page: have to fax orders over \$2500

P.O.s faster to walk over to the financial office than to wait for in SAPWeb.

Can't complete open P.O.s in SAPWeb.

No way to add new vendors.

Line-item info must be retyped over and over.

No way to do items less than 1 cent each.

No credit card confirmation in SAPWeb.

Text Interface

Inconsistency: "Amount Available" versus "Amount outstanding": they mean the same thing, so decide on one.

No notification for orders over \$2500 – unclear error message.

No indication whether Search is working or whether it is stalled.

"Item Text" / "Short Text" : what is the distinction?

-to add more text later, you have to go through security again

"Material Group": what is it? Alerts appear demanding it.

No notification of P.O. issue.

No alert to let you know what E-Reqs are in your inbox.

General

Too many delays with no confirmation that it's working.

Not enough of SAP or E-Req functionality in SAPWeb.

APPENDIX D: Mr. Java Paper Questionnaire

More to say about Mr. Java?

Feel free to fill out one of these forms and leave it in the box on top of Mr. Java, or email Matt Norwood <rowan@media.mit.edu>. More information on Mr. Java: Customer Support can be found at <http://web.mit.edu/norwoodm/www/mrjava/>

Mr. Java: Customer Support

Written Feedback Form

I can see Mr. Java reacting to my feedback: **Clearly** **Sort of** **Not at all**

The graph on Mr. Java is: **Interesting** **Confusing** **Annoying**
Other (explain)

I have privacy concerns about Mr. Java: **Yes** **Sort of**
No

I like the greeting on the monitor: **Very much** **Sort of** **Not at all**

I like the feedback text on the monitor: **Very much** **Sort of** **Not at all**
all

Other comments:

APPENDIX E: Mr. Java Responses

Reponses are arranged from least to most extreme, as determined by the length and strength of user pressure on the feedback phicon.

Negative Reponses

Mild

Thank you for your feedback. The information will be used to help improve Mr. Java.

Thank you for reporting how Mr. Java is doing. Your feedback will be used to help improve Mr. Java.

Your report has been recorded, and the will be used to streamline customer interaction Mr. Java.

Moderate

Sorry Mr. Java is giving you problems. Thank you for your input, which will help in improving Mr. Java.

Sounds like Mr. Java isn't working quite right. Thanks for reporting it so that the problem can be addressed.

Sorry for any inconvenience Mr. Java may have caused you. Thank you for your input, which will be used to refine the machine.

Extreme

Sounds bad. Apologies for the inconvenience. Your feedback will be used to help address the problem.

Sounds bad. Apologies for any inconvenience Mr. Java may be causing. Your report will help to address the problem.

Sounds like Mr. Java's behavior is really displeasing. Thank you for your feedback, which is vital to helping to address the problem.

Positive Responses

Mild

Thanks!

Thanks for the input!

Glad to hear Mr. Java is doing well!

Moderate

Thanks for the feedback!

Your input will be used to analyze Mr. Java's performance and make it more consistent in its service.

Glad to hear Mr. Java is behaving well.

Your feedback is appreciated -- it will be used to ensure Mr. Java's service for other users.

Sounds good!

Hope you keep giving Mr. Java feedback on his behavior!

Extreme

Sounds great! Please continue to give Mr. Java feedback on his behavior to help streamline the machine.

Glad to hear it!

Your feedback will help to streamline the Mr. Java interface.

Great to hear!

Hope you keep letting Mr. Java know how he's doing so the developers can continue to refine the machine.

BIBLIOGRAPHY

1. Acorto Inc. (1997). "Caffe Acorto Model 2000s Software Requirements". Acorto Inc., Bellevue, WA.
2. Brannback, M. and Puhakainen, J. (1998). "Redesigning the service concept for the age of the Internet". TUCS Technical Report No 156.
3. Clement, A. (1990). "Cooperative Support for Computer Work: A Social Perspective on the Empowering of End-Users". CSCW 90 Proceedings.
4. Dameron, D. (1990). "The Editorial Role in Developing An Online User Interface". ACM.
5. Das, A.; Soh, C.; and Lee, P. (1999). "A Model of Customer Satisfaction with Information Technology Service Providers: An Empirical Study". SIGCPR '99, New Orleans, LA.
6. Fernandez, R. (1997). Stochastic Modeling of Physiological Signals with Hidden Markov Models: A Step Toward Frustration Detection in Human-Computer Interfaces. MIT Masters' Thesis, September 1997.
7. Giraud-Carrier, C. and Corley, S. (1998) "Inductive CBR for Customer Support". Department of Computer Science, Bristol University, England.
8. Goleman, D. (1995). Emotional Intelligence. Bantam Books, New York.
9. Gorbet, M., and Orth, M. (1997). "Triangles: Design of a Physical/Digital Construction Kit". Proceedings of Designing Interactive Systems (DIS '97), (Amsterdam, August 1997), ACM Press, pp. 125-128.
10. Ishii, H. and Ullmer, B. (1997). "Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms". Proceedings of CHI '97, March 22-27, 1997.
11. Katchabaw, M., Lutfiyya, H., and Bauer, M. (1998). "Driving Resource Management With Application-Level Quality of Service Specifications". ICE 98, Charleston, SC.
12. Kaye, Matsakis, Gray, Wheeler & Hawley (2000). "PC Dinners, Mr. Java and Counter Intelligence: Prototyping Smart Appliances for the Kitchen." To be published in a special Information Appliances issue of Computer Graphics & Applications, provisionally May 2000.
13. Klein, J. (1998). Computer Response to User Frustration. MAS Thesis, MIT Media Lab.
14. Klein, J., Riseberg, J., and Picard, R.W. (1997). "Support for human emotional needs in human-computer interaction". Juried position paper accepted for *Workshop on HCI Research and Practice Agenda Based on Human Needs and Social Responsibility*, CHI'97, Atlanta.
15. Klein, J. (1997). "TurnStyles: A speech interface agent to support affective communication". http://affect.www.media.mit.edu/projects/affect/AC_research/projects/turnstyles.html
16. Landauer, T.K. (1995). The Trouble With Computers. Cambridge, MA; MIT Press.
17. Lowder, J. (1999). "Web-based Student Feedback to Improve Learning". ACM.
18. Mantei, M. and Teorey, T. (1988). "Cost/Benefit Analysis for Incorporating Human Factors in the Software Cycle". Communications of the ACM, Vol 31, No.4.
19. Molich, R. and Nielsen, J. (1990). "Improving a Human-Computer Dialogue". ACM.
20. Neuman, B and Ts'o, T. (1994). "Kerberos: An Authentication Service for Computer Networks". IEEE Communications Magazine, Volume 32, Number 9, pages 33-38, September 1994.

21. Picard, R. (1997). Affective Computing. MIT Press, Cambridge, MA.
22. Poor, R. (1997). "The iRX 2.0". Media Lab, MIT, Cambridge, MA.
23. Purushotham, P. (1999). "'Third Voice' fields protests, praise". The Hindu Business Line, July 21, 1999.
24. Reeves, C. and Nass, B. (1996). The Media Equation. New York, NY; Cambridge University Press.
25. Robinson, S. (1999). "CD Software Said to Gather Data on Users". New York Times, Nov. 1, 1999, New York, NY.
26. Shiva, V.A. (1999). "EchoMail". General Interactive white papers, <http://www.interactive.com/>
27. Starr, P. and Keenan, T. (1987). "Structured Feedback: An Approach to Developing Knowledge of a System During Initial Model Development". Proceedings of the 1987 Winter Simulation Conference.