Proceedings of Human Systems Integration Symposium 2003, Engineering for Usability, Vienna, VA, June 23–25, 2003.

Joseph L. Gabbard, M.S., Deborah Hix, Ph.D., J. Edward Swan II, Ph.D., Mark A. Livingston, Ph.D., Tobias H. Höllerer, M.S., Simon J. Julier, Ph.D., Dennis Brown, M.S., Yohan Baillot, M.S.

Usability Engineering for Complex Interactive Systems Development

ABSTRACT

Usability engineering is a cost-effective, user-centered process that ensures a high level of effectiveness, efficiency, and safety in complex interactive systems. This paper presents a brief description of usability engineering activities, and discusses our experiences with leading usability engineering activities for three very different types of interactive applications: a responsive workbench-based command and control application called Dragon, a wearable augmented reality application for urban warfare called Battlefield Augmented Reality System (BARS), and a head-mounted hardware device, called Nomad, for dismounted soldiers. For each application, we present our approach to usability engineering, how we tailored the usability engineering process and methods to address application-specific needs, and give results.

INTRODUCTION AND MOTIVATION

Usability engineering is a cost-effective, user-centered process that ensures a high level of effectiveness, efficiency, and safety in complex interactive systems (Hix and Hartson, 1993). Activities in this process include user analysis, user task analysis, conceptual and detailed user interface design, quantifiable usability metrics, rapid prototyping, and various kinds of user-centered evaluations of the user interface. These activities are further explained in Section "Activities in Usability Engineering."

Usability engineering produces highly usable user interfaces that are essential to reduced manning, reduced human error, and increased productivity. Unfortunately, managers and developers often have the misconception that usability engineering activities *add* costs to a product's development life cycle. In fact, usability engineering can *reduce* costs over the life of the product, by reducing the need to add missed functionality later in the development cycle, when such additions are more expensive. The process is an integral part of interactive application development, just as are systems engineering and software engineering. Usability engineering activities can be tailored to allow individualizing as needed for a specific project or product development effort.

The usability engineering process applies to any interactive system, ranging from training applications to multimedia CD-ROMs to augmented and virtual environments to simulation applications to graphical user interfaces (GUIs). The usability engineering process is flexible enough to be applied at any stage of the development life cycle, although early use of the process provides the best opportunity for cost-savings.

We have led usability engineering efforts on many different types of interactive military system development projects. This includes a responsive workbench-based command and control application called Dragon (Durbin et al., 1998), a wearable augmented reality application for urban warfare called Battlefield Augmented Reality System (BARS) (Gabbard et al., 2002), and a head-mounted hardware device, called Nomad (Microvision, 2003), for dismounted soldiers. In this paper, we present a brief description of key usability engineering activities (Section "Activities in Usability Engineering"). Within this context, we discuss our experiences with various usability engineering activities for each of the three interactive systems (Section "Usability Engineering Case Studies: Developing Complex Interactive Systems"). For each system, we present our approach to usability engineering, and how we tailored the process and methods as necessary to address application-specific needs, and give results. Our general conclusions focus on 'lessons learned' in improving both the usability engineering process and resulting complex interactive systems.

ACTIVITIES IN USABILITY ENGINEERING

As mentioned in the Introduction, usability engineering consists of numerous activities. Figure 1 shows a simple diagram of the major activities. Usability engineering includes both design and evaluations with users; it is not just applicable at the evaluation phase. Usability engineering is not typically hypothesis-testing-based experimentation, but instead is structured, iterative user-centered design and evaluation applied during all phases of the interactive system development life cycle. Most existing usability engineering methods were spawned by the development of traditional desktop graphical user interface (GUIs).

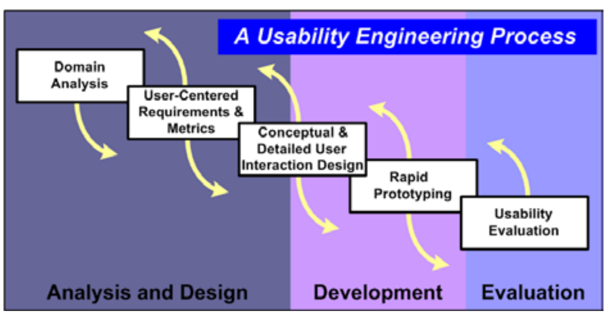

Figure 1. Typical user-centered activities associated with our usability engineering process. Although the usual flow is generally left-to-right from activity to activity, the arrows indicate the substantial iterations and revisions that occurs in practice.

In the following sections, we discuss several of the major usability engineering activities, including domain analysis, expert evaluation (also sometimes called heuristic evaluation or usability inspection), formative usability evaluation, and summative usability evaluation.

Domain Analysis

Domain analysis is the process by which answers to two critical questions about a specific application context are determined:

- Who are the users?
- What tasks will they perform?

Thus, a key activity in domain analysis is *user task analysis*, which produces a complete description of tasks, subtasks, and actions that an interactive system should provide to support its human users, as well as other resources necessary for users and the system to cooperatively perform tasks (Hix and Hartson, 1993; Hackos and Redish, 1998). While it is preferable that user task analyses be performed early in the development process, like all aspects of user interface development, task analyses also need to be flexible and potentially iterative, allowing for modifications to user performance and other user interface requirements during any stage of development.

In our experience, interviewing an existing and/or identified user base, along with subject matter experts and application "visionaries," provides very useful insight into what users need and expect from an application. Observation-based analysis requires a user interaction prototype, and as such, is used as a last resort. A combination of early analysis of application documentation (when available) and interviews with subject matter experts typically provides the most effective user task analysis.

Domain analysis generates critical information used throughout all stages of the usability engineering life cycle. A key result is a top-down, typically hierarchical decomposition of detailed *user task descriptions*. This decomposition serves as an enumeration and explanation of desired functionality for use by designers and evaluators, as well as required task sequences. Other key results are one or more detailed *scenarios,* describing potential uses of the application, and a list of *user-centered requirements*. Without a clear understanding of application domain user tasks and user requirements, both evaluators and developers are forced to "best guess" or interpret desired functionality, which inevitably leads to poor user interface design.

Expert Evaluation

Expert evaluation (also called *heuristic evaluation* or *usability inspection*) is the process of identifying potential usability problems by comparing a user interface design to established usability design guidelines. The identified problems are then used to derive recommendations for improving that design. This method is used by usability experts to identify critical usability problems early in the development cycle, so that these design issues can be addressed as part of the iterative design process (Nielsen, 1993). Often the usability experts rely explicitly and solely on established usability design guidelines to determine whether a user interface design effectively and efficiently supports user task performance (i.e., usability). But usability experts can also rely more implicitly on design guidelines and work through user task scenarios during their evaluation. Nielsen (1993) recommends three to five evaluators for an expert evaluation, and has shown empirically that fewer evaluators generally identify only a small subset of problems and that more evaluators produce diminishing results at higher costs. Each evaluator first inspects the design alone, independently of other evaluators' findings. Then the evaluators combine their data to analyze both common and conflicting usability findings. Results from an expert evaluation should not only identify problematic user interface components and interaction techniques, but should also indicate *why* a particular component or technique is problematic. This is arguably the most cost-effective type of usability evaluation, because it does not involve users.

Formative Usability Evaluation

Formative evaluation is the process of assessing, refining, and improving a user interface design by having representative users perform task-based scenarios, observing their performance, and collecting and analyzing data to empirically identify usability problems (Hix and Hartson, 1993). This observational evaluation method can ensure usability of interactive systems by including users early and continually throughout user interface development. This method relies heavily on usage context (e.g., user tasks, user motivation), as well as a solid understanding of human-computer interaction (Hix and Hartson, 1993).

A typical cycle of formative evaluation begins with the creation of scenarios based on the user task analysis. These scenarios are specifically designed to exploit and explore all identified tasks, information, and work flows. Representative users perform these tasks as evaluators collect both qualitative and quantitative data. Evaluators then analyze these data to identify user interface components or features that both support and detract from user task performance, and to suggest user interface design changes, as well as scenario (re)design.

Formative evaluation produces both qualitative and quantitative results collected from representative users during their performance of task scenarios (del Galdo et al., 1986; Hix and Hartson, 1993). *Qualitative data* include *critical incidents,* a user event that has a significant impact, either positive or negative, on users' task performance and/or satisfaction. *Quantitative data* include metrics such as how long it takes a user to perform a given task, the number of errors encountered during task performance, measures of user satisfaction, and so on. Collected quantitative data are then compared to appropriate baseline metrics, sometimes initially redefining or altering evaluators' perceptions of what should be considered baseline. Both qualitative and quantitative data are equally important since they each provide unique insight into a user interface design's strengths and weaknesses.

Summative Usability Evaluation

Summative evaluation, in contrast to formative evaluation, is a process that is typically performed after a product or some part of its design is more or less complete. Its purpose is to statistically compare several different systems or candidate designs, for example, to determine which one is "better," where better is defined in advance. In practice, summative evaluation can take many forms. The most common are the comparative, field trial, and more recently, the expert review (Stevens et al., 1997). While both the field trial and expert review methods are well-suited for design assessment, they typically involve assessment of single prototypes or field-delivered designs. Our experiences have found that the empirical comparative approach employing representative users is very effective for analyzing strengths and weaknesses of various well-formed, candidate designs set within appropriate user scenarios. However, it is the most costly type of evaluation because it may need large numbers of users to achieve statistical validity and reliability, and because data analysis can be complex and challenging.

A Cost-Effective Evaluation Progression

As depicted in Figure 2, our applied research over the past several years has shown that progressing from expert evaluation to formative evaluation to summative evaluation is an efficient and cost-effective strategy for assessing and improving the user interface (Gabbard, Hix, and Swan, 1999).

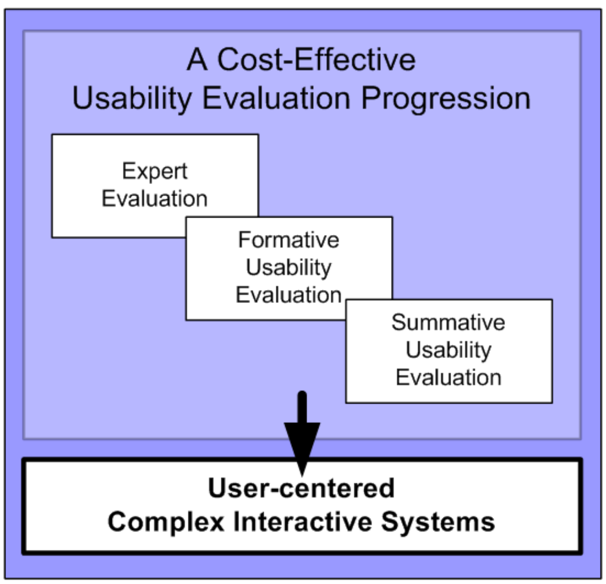

Figure 2. A cost-effective usability evaluation progression

For example, if summative studies are performed on user interface designs that have had little or no user task analysis or expert or formative evaluation, the expensive summative evaluation may be essentially comparing "good apples" to "bad oranges" (Hix et al., 1999). Specifically, a summative study of two different application interfaces may be comparing one design that is inherently better, in terms of usability, than the other one. When all designs in a summative study have been developed following this suggested progression of usability engineering activities, then the comparison should be more valid. Experimenters will then know that the interface designs are basically equivalent in terms of their usability, and any differences found among compared designs are, in fact, due to variations in the fundamental nature of the designs, and not their usability.

USABILITY ENGINEERING CASE STUDIES: DEVELOPING COMPLEX INTERACTIVE SYSTEMS

We next present three case studies in our experiences of applying usability engineering methods to three different complex interactive applications. The first, called *Dragon*, is a military command and control application developed on a responsive workbench. The next, called *BARS*, is an augmented reality system to be worn by mobile urban warfighters. The third, called *Nomad*, is a head-worn, see-through display that augments the real world with graphical and textual information. For each of these applications, we followed the usability engineering methods described above with great success, as discussed below.

Dragon Real-time Battlefield Visualization System

BACKGROUND / DESCRIPTION

For decades, battlefield visualization has been accomplished by placing paper maps of the battlespace under sheets of acetate and, prior to paper maps, was performed using a sandtable (a box filled with sand shaped to replicate the battlespace terrain). Personnel at the Naval Research Laboratory's (NRL) Virtual Reality Lab developed a virtual environment application, called Dragon, for next-generation battlefield visualization (Durbin et al., 1998).

In Dragon, a responsive workbench (Kruger et al., 1995) provides a three-dimensional display for observing and managing battlespace information shared among commanders and other battle planners. As described in (Hix et al., 1999), Dragon is a battlefield visualization system that displays a three-dimensional map of the battlespace, as well as military entities (e.g., tanks and ships) represented with semi-realistic models. Dragon allows users to navigate and view the map and symbols, as well as to query and manipulate entities, using a modified flightstick. Figure 3 shows a typical user view of Dragon.

Figure 3. User's view of the Dragon battlefield visualization system

USABILITY ENGINEERING APPROACHES AND METHODS

During early Dragon demonstrations and evaluations, we observed that the user task of "navigation" – how users manipulate their viewpoint to move from place to place in a virtual world – profoundly affects all other user tasks. This is because, when using a map-based system, users must always first navigate to a particular area of the map. Thus, all the usability engineering methods, including domain analysis, user task analysis, expert evaluation, formative evaluation, and summative evaluation, that we applied to Dragon focused on the key user task of navigation.

Domain Analysis

Early in its development, Dragon was demonstrated as a prototype system at two different military exercises, where feedback from both civilian and military users was informally elicited. This feedback was the impetus for a more formal domain and user task analysis that included subject matter experts from Naval personnel. Important Dragon-specific high-level tasks identified during our domain and user task analysis included planning and shaping a battlefield, comprehending situational awareness in a changing battlespace, performing engagement and execution exercises, and carrying out "what if" (contingency planning) exercises. In the user task analysis, we also examined how personnel perform their current battlefield visualization tasks. Navigation is critical to all these high-level tasks.

Expert Evaluation

During our expert evaluations, three user interface design experts assessed the evolving user interface design for Dragon. In early evaluations, the experts did not follow specific user task scenarios per se, but simply engaged in exploratory use of the user interface. Our subsequent expert evaluations were guided largely by our own knowledge of interaction design for virtual environments and, more formally, by the Dragon user task analysis, as well as a framework for usability characteristics for virtual environments (Gabbard, 1997).

Major usability design problems revealed by four major cycles of expert evaluations and subsequent redesign based on findings included poor mapping of navigation tasks to flightstick buttons, difficulty with damping of map movement in response to a user's flightstick movement, and inadequate graphical and textual feedback to the user about the current navigation task. We discuss these problems, and how we addressed them, in detail elsewhere (Hix et al., 1999). As our cycles of expert evaluations began to reveal fewer and fewer user interface design issues, we moved on to formative evaluations.

Formative Evaluation

Based on our domain and user task analyses, we created a set of user task scenarios consisting of benchmark user tasks, carefully considered for coverage of specific issues related to navigation. We thoroughly pre-tested and debugged all scenarios before presenting them to users.

During each of six formative evaluation sessions, each with an individual subject, we followed a formal protocol designed to elicit both quantitative (task time and error counts) and qualitative (critical incidents, especially related to errors, and constructive comments made about the design) user data. Time to perform the set of scenarios ranged from about 20 minutes to more than an hour.

During each session, we had at least two and sometimes three evaluators present. The evaluation leader ran the session and interacted with the user; the other one or two evaluators recorded timings, counted errors, and collected qualitative data. We found that the quality and amount of data collected by multiple evaluators greatly outweighed the cost of those evaluators. After each session, we analyzed both the quantitative and qualitative data, and based the next design iteration on our results.

Summative Evaluation

Our expert and formative evaluation work for Dragon revealed four variables most likely to influence virtual environment navigation tasks (Gabbard, Hix, and Swan, 1999). Subsequently, our summative evaluation manipulated and studied those four independent variables and their values, specifically:

- *Display platform* (CAVE™, wall, workbench, desktop): a standard immersive room, a single wall, a responsive workbench, and a standard desktop monitor, respectively
- *Stereopsis* (stereo, mono)
- *Movement control* (rate, position): how a subject's navigational hand gesture controls the resulting map movement
- *Frame of reference* (egocentric, exocentric): whether the user's actions with the flightstick appear to move the user through the world, or whether actions appear to move the virtual world around the user

Thirty-two subjects performed a series of 17 carefully designed and pre-tested tasks, each requiring the subject to navigate to a specific location, manipulate the map, and/or answer a specific question based on the map.

RESULTS AND DISCUSSION

Our summative evaluation yielded interesting results (Swan et al., 2003). A striking finding of our results was that the desktop had the best overall user performance time of all display platforms. Many user tasks required finding, identifying, and/or reading text or objects labeled with text. While all displays were set to 1024 x 768 pixels, the size of the projection surface varied enough to conjecture that pixel density is more critical than field of view or display size. Our observations and qualitative data support this claim. This research suggests we should further research user task performance using high-resolution displays. Interestingly, we also found no effect of platform at all in map tasks and geometric object tasks. This begs examination of the important question: "Why are we building large display virtual environments and incurring the resulting expense if the user benefit is not there?"

Battlefield Augmented Reality System (BARS)

BACKGROUND / DESCRIPTION

Urban terrain is one of the most important environments that current and future warfighters face. Because of increased urbanization, many future military operations will occur in cities. However, urban terrain is also one of the most demanding environments, with complicated three-dimensional infrastructure potentially harboring many types of risks (such as snipers or instability due to structural damage).

We are developing the Battlefield Augmented Reality System (BARS) (Gabbard et al., 2002) to mitigate these difficulties through the use of mobile augmented reality. Augmented reality is a display paradigm that mixes computer-generated graphics with a user's view of the real world (an example is shown Figure 4). The user wears a see-through head-mounted display that the system tracks in six-degree-of-freedom space (position and orientation). Computer graphics are created and aligned from the user's perspective with the objects to be augmented. By providing direct, heads-up access to information correlated with a user's view of the real world, mobile augmented reality has the potential to recast the way information is presented and accessed.

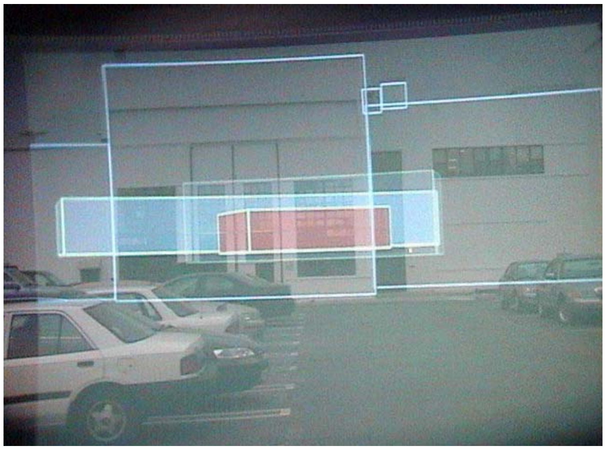

Figure 4. BARS user's view of the real world augmented with overlaid graphics.

Mobile augmented reality has many research challenges related to the design of the user interface, one of which is illustrated in Figure 4: the "*Superman X-ray vision problem*" (Stedman et al., 1999). This problem encapsulates the fundamental advantages and disadvantages of mobile augmented reality. With such a system, a user has "X-ray" vision and can see information about objects that are not visible. However, the user loses occlusion cues, which are extremely important for perceiving depth.

USABILITY ENGINEERING APPROACHES AND METHODS

Domain Analysis

Team members participating in domain analysis activities for BARS included personnel from the Naval Research Laboratory (software and system developers), Virginia Tech (usability engineers), and a USMCR Captain, who served in the critical role of subject matter expert. Our first domain analysis product was a specific *scenario* for BARS, to represent a realistic and significant warfighting task situation in an urban warfare setting (Gabbard et al., 2002). The scenario was developed over a couple of days with the subject matter expert. We performed early phases of user task analysis based on military documents (e.g., Beevor, 1998; Bowden, 1999) that describe protocol and tactics within an urban terrain. This allowed us to verify procedures (i.e., potential user tasks) as well as user information (i.e., data) needs, and military doctrine manuals (e.g., Thompson, 2001; US ARMY, 1993) that define specific terminology and symbology to ensure that the scenario was as accurate, thorough, representative, and concise as possible. Many times, information (such as terminology and symbology) captured during domain analysis is transitioned into the development effort and eventually manifests itself in the user interface. This is, in fact, the desired outcome since a well-conceived, user-centered domain analysis should lead directly to user interface design (and implementation) decisions.

We analyzed the scenario to produce a list of *user-centered requirements*. This list is typically the final outcome of domain analysis activities, and it is given to system engineers to aid in their development of an application. Interestingly, producing the user-centered requirements drove an important design decision. We realized that our user-centered requirements identified a list of features that could not be easily delivered by any current augmented reality system. Therefore our development team decided to take a step back and conduct some basic research and development underpinning these requirements. For example, one BARS user-centered requirement said that the system must be able to display the location of hidden and occluded objects (e.g., a tank located behind a visible building). This raised numerous user interface design questions such as how such occluded objects should be presented graphically to a user (the 'X-ray vision' problem). To address such issues, we began with expert evaluations.

Expert Evaluation

During six cycles of expert evaluation, we designed approximately 100 mockups depicting various potential designs for representing occlusion, using a variety of drawing parameters including:

- *Drawing style* (i.e., solid, dashed, dotted) lines or polygons
- *Outlined or filled* (shaded) polygons
- *Intensity* of lines or fill
- *Thickness* of lines

We were specifically examining several aspects of occlusion, including how best to visually represent occluded information and objects, the number of discriminable levels of occlusion, and variations on the above drawing parameters. In each cycle of expert evaluation, team members individually examined a set of occlusion representations (set size ranged from 5 to 30 mockups in a cycle), which were created using Adobe Photoshop and Microsoft Powerpoint employing video to capture real-world scenes as background images. Then as a team, we compiled our assessments, to get consensus on our conclusions and to determine how to design the next set of representations, informed by results of the current cycle. Our findings showed that line intensity appeared to be the most powerful (i.e., consistently recognizable) line-only drawing parameter, followed by line style. Further, both line-based and shaded representations were discriminable at only three or four levels of occlusion. Once we had iterated to an optimal set of representations, we used these representations to move on to formative evaluations using them.

Formative Evaluation

Continuing with our study of occlusion, we created a formal set of user tasks, based on our scenario. We then had five individual subjects perform the set of tasks while we collected both qualitative and quantitative data. Having anticipated the challenge of working in an outdoor, mobile, highly dynamic environment, team members had to consider novel approaches to usability evaluation. Our solution was to design and build a specially-constructed motion tracking ‘cage’ so that BARS could accurately track the user and accurately register graphics onto the real world. We also set up auxiliary evaluator’s monitors to provide evaluators an accurate depiction of a user’s view during task performance.

Our results showed that users performed approximately 85% of the tasks correctly and efficiently with less than 10 minutes of training using BARS. Other results supported findings from our expert evaluations, such as no more than three or four levels of occlusion are discriminable. We made new findings, such as the fact that the three-dimensionality of occluded objects was easier to perceive in shaded objects than in line-drawn objects. Users developed distinct strategies for using BARS, and all users had a very positive, enthusiastic reaction to BARS and its capabilities.

Summative Evaluation

Much as in our Dragon evaluations, our expert and formative evaluations of BARS led us logically to critical factors, in this case graphical techniques for displaying ordering and distance of occluded objects, that needed the statistical confirmation of summative evaluation. Specifically, we determined from our results that a critical, yet tenable set of factors and their values for summative study were:

- *Drawing style* – line, filled, line+fill
- *Opacity* – constant, increasing
- *Intensity* – constant, decreasing

Our reasoning behind choice of values for each factor is detailed in (Livingston et al., 2003). The study was run with eight subjects, who saw a small virtual world that consisted of representations of six blue buildings and a small red target object. The user’s task was to indicate the location of the target as it moved among buildings from trial to trial.

RESULTS AND DISCUSSION

At the time of this writing, we have not completed data analysis of our summative study, but we do have some preliminary results (Livingston et al., 2003). Subjects made 79% correct choices and 21% erroneous choices of the target location during trials. User errors fell into two categories: the target could be closer than the user’s answer, or farther than the user’s answer. Subjects were most accurate when the target was in the far position; only 17.3% of their erroneous choices were made when the target was in the far position, as compared to 38.6% in the close position, and 44.2% in the middle position. Other preliminary findings indicate that ‘line+fill’ drawing style yielded the best accuracy. Overall, our early results indicate that was have evolved an effective and efficient set of graphical representations for occlusion, by using our usability engineering methodology. Complete statistical and other findings will be reported in later publications.

Battlefield Information Display Technology

BACKGROUND / DESCRIPTION

The Battlefield Information Display Technology (BIDT) program was conceived by ONR to advance see-through, head-mounted, wireless display technologies for depicting tactical information for the mobile urban warfighter. These display advances are designed to integrate with emerging Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance systems. The current BIDT display is a monocular, head-worn Nomad augmented vision system manufactured by Microvision (Microvision, 2003). This display uses a low-powered laser beam to paint an image directly on a user's retina. This technology addresses a key drawback of phosphor-based see-through displays: no such display can come close to matching the range of luminance values encountered in outdoor settings, but lasers (and the human eye) can. Examples of the display device are shown in Figure 5.

Figure 5. User-based evaluations of the BIDT/Nomad display employed both military and civilian users.

Usability engineering for this project was somewhat different from many of our other projects (including the two others in this paper), because a usability engineering goal for BIDT is to assess a hardware device rather than a software user interface. Specifically, we are identifying critical design issues and performance parameters that directly impact user performance. For BIDT, we performed domain analysis activities such as identifying critical graphical elements to support urban warfighting scenarios and developing user-centered requirements. We also performed usability evaluation activities to elicit user feedback from both augmented reality experts and from military experts.

USABILITY ENGINEERING APPROACHES AND METHODS

Thus far in our on-going work with the BIDT program, we have applied several usability engineering approaches to the Nomad, including domain analysis, scenario development, user information requirements, user-centered requirements, and formative evaluations. We are continuing our usability engineering work with more formative and summative evaluations, which are not yet ready to report.

Domain Analysis

We performed extensive domain analysis of the urban warfighting domain. These efforts identified potential scenarios and associated user information requirements to be used for both the user interface design and to further usability engineering activities. We researched user information requirements at two levels: *what information* needs to be displayed and *what graphical elements* will likely be used to convey the needed information. In both cases, the set of user information requirements focused on supporting the urban warfighting domain.

To identify user-centered, task-based requirements, we leveraged scenarios from our BARS work (see Section "Battlefield Augmented Reality System (BARS)"), developed in conjunction with military experts. We captured information needs of different users, such as what objects a user needs to see, what data (e.g., about objects) a user needs access to, and what tasks a user needs to perform with BARS. We then translated information needs gathered from the scenarios into high-level user-centered requirements. We extracted user-centered requirements by systematically examining the scenarios and associated user tasks. Specifically, we examined each major task set within the scenario, as well as specific user interface and information needs at each step of those tasks. In essence, during this usability engineering activity, we enumerated what the user needs in terms of information and features at each step of a task sequence, where the union of all task sequences represents the most probable workflow for a particular scenario.

Formative Evaluation

Our BIDT user-based evaluations were fairly informal, and were designed to assess current Nomad display features and some prototype user interfaces, as well as important usability issues. They were also to provide a prioritized list of issues and recommendations for user-centered design changes to the Nomad.

We used two different approaches to structure our Nomad user-based evaluations:

Assessment by a Group: Demonstrating the Nomad display to a group of about a dozen students and faculty who participate in regular University-wide virtual reality research meetings at Virginia Tech. Each user donned the display and made comments to the group. Comments spurred group discussion aided by a projection of what that user was seeing onto the wall. Each user wore the Nomad 7 to 10 minutes.

Assessment by Individual Marine and Navy Users: Use of the display by Virginia Tech Naval ROTC instructors (active Marines) and Virginia Tech Naval ROTC students (Midshipmen in their senior year, but all with prior active military service). We guided these users through specific tasks and also performed structured interviews to elicit feedback on the display. Each user wore the Nomad about one hour.

We conducted all evaluation sessions indoors, but with users performing tasks that required them to look both outdoors and indoors (as shown in Figure 5.) Some of the questions we posed to users were designed to determine, for example, what a user could read on the Nomad display based on different focal lengths – that is, when focusing on real-world objects at near-range (approximately arm's length), mid-range (approximately 50 feet away), and far-range (out toward the horizon/infinity); how well a user could perform real-world tasks (e.g., operate a fax machine) with various graphics/text "in the way" (i.e., visible on the Nomad screen); and how much context switching (between Nomad screen image and real world) a user had to do, and how difficult this was.

We also gathered information on users' thoughts on the user interface prototypes and on suggestions of other applications for which the Nomad might be appropriate.

RESULTS AND DISCUSSION

Our early domain analysis efforts identified domain-specific information and data to be displayed in an urban warfighting scenario. This list contains 27 groups of information objects that we then organized (also based on our discussions with subject matter experts) into three categories of representative information: (1) geographic and environmental entities; (2) friendly forces, goals, and objectives; and (3) enemy assets. Our efforts also identified a list of graphical elements to present the 27 groups of information objects to a user.

Our work identified 31 user-centered requirements that address usability issues of the physical visual presentation device, or the display (in this case, the Nomad), grouped according to four categories:

- *Features and functions* – capabilities of the display itself that a warfighter may need to perform specific tasks within the urban warfighting domain, but that are also generalizable to other outdoor, mobile augmented reality settings

- *Visual characteristics of the display* – general properties of the visual display device, independent of the specific brand, model, etc.

- *Weight and power characteristics of the display* – issues affecting how heavy the display is and how much power it needs

- *Form factor of the display* – issues affecting physical design of the wearable display hardware

We also observed that intensity of the display image seemed more important to some users than complexity of the real-world background they were viewing through the Nomad. Different levels of image transparency obviously influenced how much of the real-world background a user could see. This in turn raised some compelling questions for further study. For example: What percent of the real world can be occluded by graphics/text, at what level of transparency, under what lighting conditions, for the user to perform particular types of tasks? How does a source of bright light behind the user affect the display (e.g., is it reflected on the display, perhaps obliterating some of the graphics/text)?

GENERAL CONCLUSIONS

From these three and numerous other projects, we have learned many lessons on how to improve the *process* of usability engineering. For example, the great benefits that a subject matter expert provides to usability engineering activities are constantly reinforced. These experts provide specific context-related information to help usability experts understand user task and information flow requirements. They also help direct and rank analysis foci so that evaluation resources are allocated to the most important usage issues.

Additionally, a key finding throughout our work is the successful progression from expert to formative to summative evaluations as a very cost-effective strategy for assessing and improving a user interface design. Expert evaluations identify obvious usability problems or missing functionality, thus allowing improvements to a user interface prior to performing user-based formative evaluations.

If expert evaluations are not performed prior to formative evaluations, the formative evaluations will typically take longer and require more users, and yet reveal many of the same usability problems that could have been discovered by less expensive expert evaluations. Once designs have been expertly and/or formatively evaluated, then experimenters can have confidence that those designs are essentially equivalent in terms of their usability, and thus facilitate a compelling comparative summative study.
Moreover, as indicated in both Dragon and BARS above, we found that results from formative evaluations inform the design of summative studies by helping determine critical usability characteristics to evaluate and compare.

Another important advantage of applying the complete progression of usability engineering methods is the timeliness of assessment efforts. This aligns each activity's strengths (such as level of detail or breadth of focus) with concurrent efforts in the software development process.

We expect to continue developing products such as those described in this paper, by continuing to apply and enhance as necessary the process of usability engineering.

REFERENCES

Beevor, A., *Stalingrad: The Fateful Siege 1942-1943*, Viking Press, 1998.

Bowden, M., *Black Hawk Down*, Atlantic Press, 1999.

del Galdo, E.M., Williges, R.C., Williges, B.H., and Wixon, D.R., "An Evaluation of Critical Incidents for Software Documentation Design", In *Proc. Thirtieth Annual Human Factors Society Conference*, Anaheim, CA. 1986.

Durbin, J, Swan II, JE, Colbert, B, Crowe, J, King, R, King, T, Scannell, C, Wartell, Z, and Welsh, T., "Battlefield Visualization on the Responsive Workbench", In *Proc. IEEE Visualization '98*, IEEE Computer Society Press, pp. 463-466. 1998.

Gabbard, J.L., Swan, J.E., Hix, D., Lanzagorta, M., Livingston, M., Brown, D., Julier, S., "Usability Engineering: Domain Analysis Activities for Augmented Reality Systems," In *Proc. SPIE* Vol. 4660, p. 445-457, *Stereoscopic Displays and Virtual Reality Systems IX*, Andrew J. Woods; John O. Merritt; Stephen A. Benton; Mark T. Bolas; Eds. Photonics West 2002, Electronic Imaging conference, San Jose, CA. 2002.
Gabbard, J. L., Hix, D, and Swan, J.E., "User-Centered Design and Evaluation of Virtual Environments", invited paper for *IEEE*

Computer Graphics and Applications. Nov/Dec 1999 (Vol. 19, No. 6), pp. 51-59. 1999.

Gabbard J. L., *A Taxonomy of Usability Characteristics for Virtual Environments*, Masters Thesis. Department of Computer Science, Virginia Tech. 1997.

Hackos, J. T. and Redish, J. C., *User and Task Analysis for Interface Design.* New York: John Wiley & Sons, Inc. 1998.

Hix, D., Swan, E. J., Gabbard, J. L., McGee, M., Durbin, J., and King, T., "User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment". In *Proc. IEEE VR'99 Conference*, pp. 96-103. Houston, Texas: IEEE Computer Society Press. 1999. (Winner of "Best Paper" award at this conference.)

Hix, D. and Hartson, H. R., *Developing User Interfaces: Ensuring Usability through Product & Process,* New York: John Wiley and Sons, Inc. 1993.

Kruger, W, Bohn, CA, Frohlich, B, Schuth, H, Strauss, W, and Wesche, G., "The Responsive Workbench: A Virtual Work Environment", *IEEE Computer*, 28(7), pp. 42–48. 1995.

Livingston, M.A., Swan, J.E., Gabbard, J.L., Hix, D., Höllerer, T.H., Julier, S.J., Baillot, Y., and Brown, D., "Resolving Multiple Occluded Layers in Augmented Reality," Submitted to *ISMAR Conference*. 2003.

Microvision, Company website, see http://www.microvision.com. 2003.

Nielson, J., *Usability Engineering*, Academic Press. 1993.

Stedmon, A., Kalawsky, R., Hill, K. and Cook, C., "Old Theories, New Technologies: Cumulative Clutter Effects Using Augmented Reality," *IEEE Intl. Conf. on Information Visualization '99*, London, UK. 1999.
Stevens, F., Frances, L. and Sharp, L., *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering, and Technology Education*. NSF 93-152. 1997.

Swan, J. E., Gabbard, J.L., Hix, D., Schulman, R.S. and Kim, K.P., "A Comparative Study of User Performance in a Map-Based Virtual Environment," In *Proc. IEEE VR'03 Conference*, pp. 259-266, Los Angeles, CA: IEEE Computer Society Press. 2003.

Thompson, L., *Hostage Rescue Manual*, Greenhill Books. 2001.

US Army, *An Infantryman's Guide to Combat in Build-up Areas*, FM-90-10 -1, US Army Field Manual. 1993.

ACKNOWLEDGEMENTS

All three applications in this paper have had support from the Office of Naval Research under various Program Managers, including Dr. Helen Gigley, Dr. Astrid Schmidt-Nielsen, and Dr. Larry Rosenblum. Others too numerous to mention individually contributed to development of these applications. Dr. Richard E. Nance, Ms. Trish Hubble, and Ms. Joyce Moser of Virginia Tech's Systems Research Center have also given much support to our efforts over the past few years.

Joseph L. Gabbard (*jgabbard@vt.edu) is a Senior Research Associate on the faculty at Virginia Tech in Blacksburg VA. With more than 10 years of experience in usability engineering, he conducts research to develop and apply usability engineering methods to a broad variety of application types. He is also currently pursuing his PhD in Computer Science at Virginia Tech.*

Deborah Hix *is a Research Scientist on the faculty at Virginia Tech in Blacksburg VA. With more than 25 years of experience in usability engineering, she has done extensive research, training, and consulting in this field. She is co-author of a popular book on the topic, Developing User Interfaces: Ensuring Usability through Product & Process.*

J. Edward Swan II *is a Scientist with the Virtual Reality Laboratory at the Naval Research Laboratory, where he conducts research in graphics and human-computer interaction. His current interests include battlefield visualization and augmented reality.*

Mark A. Livingston *is a Research Scientist in the Virtual Reality Laboratory at the Naval Research Laboratory, where he works on BARS. His current research focuses on vision-based tracking algorithms and on user perception in augmented reality systems.*

Tobias H. Höllerer *is an Assistant Professor at the University of California, Santa Barbara. His research interests span several fields of human-computer interaction and experimental systems: augmented reality, virtual reality; wearable and ubiquitous computing and multimedia information systems; and computer graphics and visualization.*

Simon J. Julier *is a Research Scientist for ITT Industries at the Naval Research Laboratory. He is a technical lead on BARS. His research interests include mobile augmented reality and large-scale distributed data fusion.*

Dennis Brown *is a Senior Software Engineer for ITT Industries at the Naval Research Laboratory. He works on BARS and multi-modal virtual reality projects. His research interests include ubiquitous computing and data distribution.*

Yohan Baillot *is a computer and electrical engineer for ITT Industries at the Naval Research Laboratory. His research interests are in computer graphics, 3D displays, tracking, vision, mobile augmented reality and wearable computers.*