Performance Metrics for Intelligent Systems Workshop (PerMIS ’04), Gaithersburg, Maryland, USA, August 24–26, 2004

Evaluating System Capabilities and User Performance in the Battlefield Augmented Reality System*

Mark A. Livingston† J. Edward Swan II† Simon J. Julier‡ Yohan Baillot‡
Dennis Brown† Lawrence J. Rosenblum† Joseph L. Gabbard§ Tobias H. Höllerer¶
Deborah Hix§
Naval Research Laboratory
Washington D.C.

Abstract

We describe a first experiment in evaluating the system capabilities of the Battlefield Augmented Reality System, an interactive system designed to present military information to dismounted warfighters. We describe not just the current experiment, but a methodology of both system evaluation and user performance measurement in the system, and show how both types of tests will be useful in system development. We summarize results in a perceptual experiment being used to inform system design, and discuss ongoing and future experiments to which the work described herein leads.

1 Introduction

One of the most challenging aspects of the design of intelligent systems is the user interface – how the user will perceive and understand the system. Our application presents military information to a dismounted warfighter. In order to both refine the system’s capabilities and improve the warfighter’s performance of tasks while using the system, we measure human performance using our system, even while early in the design phase of the user interface. This paper describes an early experiment in the context of system evaluation and describes implications for both system and human performance metrics as they apply to such systems.

1.1 Application context

Military operations in urban terrain (MOUT) present many unique and challenging conditions for the warfighter. The environment is extremely complex and inherently three-dimensional. Above street level, buildings serve varying purposes (such as hospitals or communication stations). They can harbor many risks, such as snipers or mines, which can be located on different floors. Below street level, there can be an elaborate network of sewers and tunnels. The environment can be cluttered and dynamic. Narrow streets restrict line of sight and make it difficult to plan and coordinate group activities. Threats, such as snipers, can continuously move and the structure of the environment itself can change. For example, a damaged building can fill a street with rubble, making a once-safe route impassable. Such difficulties are compounded by the need to minimize the number of civilian casualties and the amount of damage to civilian targets.

In principle, many of these difficulties can be overcome through better *situation awareness*. The Concepts Division of the Marine Corps Combat Development Command (MCCDC) concludes [2]:

> Units moving in or between zones must be able to navigate effectively, and to coordinate their activities with units in other zones, as well as with units moving outside the city. This navigation and coordination capability must be resident at the very-small-unit level, perhaps even with the individual Marine.

A number of research programs have explored the means by which navigation and coordinated information can be delivered to the dismounted warfighters. We believe a mobile augmented reality system best meets the needs of the dismounted warfighter.

1.2 Mobile Augmented Reality

Augmented reality (AR) refers to the mixing of virtual cues from the real three-dimensional environment into the user’s perception. In this work, AR denotes the 3D merging

---

[*] Portions of this paper were originally published in [10] and [11].

[†] Virtual Reality Laboratory, Naval Research Laboratory. Corresponding email: livingston@ait.nrl.navy.mil

[‡] ITT Advanced Engineering and Sciences

[§] Systems Research Center, Virginia Polytechnic Inst. and State Univ.

[¶] Dept. of Computer Science, University of California, Santa Barbara

Figure 1. A sample view of our system, showing one physically visible building with representations of three buildings which it occludes.

of synthetic imagery into the user’s natural view of the surrounding world, using an optical, see-through, head-worn display.

A *mobile augmented reality* system consists of a computer, a tracking system, and a see-through HMD. The system tracks the position and orientation of the user’s head and superimposes graphics and annotations that are aligned with real objects in the user’s field of view. With this approach, complicated spatial information can be directly aligned with the environment. This contrasts with the use of hand-held displays and other electronic 2D maps. With AR, for example, the name of a building could appear as a ”virtual sign post” attached directly to the side of the building. To explore the feasibility of such a system, we are developing the Battlefield Augmented Reality System (BARS). Figure 1 is an example from BARS. This system will network multiple dismounted warfighters together with a command center.

Through the ability to present direct information overlays, integrated into the user’s environment, AR has the potential to provide significant benefits in many application areas. Many of these benefits arise from the fact that the virtual cues presented by an AR system can go beyond what is physically visible. Visuals include textual annotations, directions, instructions, or “X-ray vision,” which shows objects that are physically present, but occluded from view. Potential application domains include manufacturing [1], architecture [20], mechanical design and repair [7], medical applications [4, 17], military applications [11], tourism [6], and interactive entertainment [19].

1.3 Performance Measurement in BARS

BARS supports information gathering and human navigation for situation awareness in an urban setting [11]. A critical aspect of our research methodology is that it equally addresses both technical and human factors issues in fielding mobile AR. AR system designers have long recognized the need for standards for the performance of AR technology. As the technology begins to mature, we and some other research groups are also considering how to test user cognition when aided by AR systems.

We determined the task in which to measure performance first through consultation with domain experts [9]. They identified a strong need to visualize the spatial locations of personnel, structures, and vehicles occluded by buildings and other urban structures during military operations in urban terrain. While we can provide an overhead map view to view these relationships, using the map requires a context switch. We are designing visualization methods that enable the user to understand these relationships when directly viewing, in a heads-up manner, the augmented world in front of them.

The perceptual community has studied depth and layout perception for many years. Cutting [3] divides the visual field into three areas based on distance from the observer: near-field (within arms reach), medium-field (within approximately 30 meters), and far-field (beyond 30 meters). He then points out which depth cues are more or less effective in each field. Occlusion is the primary cue in all three spaces, but with the AR metaphor and the optical see-through, this cue is diminished. Perspective cues are also important for far-field objects, but this assumes that they are physically visible. The question for an AR system is which cues work when the user is being shown virtual representations of objects integrated into a real scene.

Our immediate goal is thus to determine methods that are appropriate for conveying depth relationships to BARS users. This requires measurement of the system’s performance in presenting information that feed the users’ perceptions of the surrounding environment. Then, we need to establish a standard for warfighter performance in the task of locating military personnel and equipment during an operation in urban terrain. For example, one goal of our work is to determine how many depth layers a user can understand.

2 Related Work

2.1 Perceptual Measures in AR Systems

A number of representations have been used to convey depth relationships between real and virtual objects. Partial transparency, dashed lines, overlays, and virtual cut-away views all give the user the impression of a difference in the depth [7, 16, 20, 12].

Furmanski et al. [8] utilized a similar approach in their pilot experiment. Using video AR, they showed users a stimulus which was either behind or at the same distance as an obstructing surface. They then asked users to identify whether the stimulus was behind, at the same distance as, or closer than the obstruction. The performance metric here is thus an ordinal depth measure. Only a single occluded object was present in the test. The parameters in the pilot test were the presence of a cutaway in the obstruction and motion parallax. The presence of the cutaway significantly improved users' perceptions of the correct location when the stimulus was behind the obstruction. The authors offered three possible locations to the users, even though only two locations were used. Users consistently believed that the stimulus was in front of the obstruction, despite the fact that it was never there.

Ellis and Menges [5] found that the presence of a visible (real) surface near a virtual object significantly influences the user's perception of the depth of the virtual object. For most users, the virtual object appeared to be nearer than it really was. This varied widely with the user's age and ability to use accommodation, even to the point of some users being influenced to think that the virtual object was further away than it really was. Adding virtual backgrounds with texture reduced the errors, as did the introduction of virtual holes, similar to those described above. Rolland et al. [13] found that occlusion of the real object by the virtual object gave the incorrect impression that the virtual object was in front, despite the object being located behind the real object and other perceptual cues denoting this relationship. Further studies showed that users performed better when allowed to adjust the depth of virtual objects than when making forced-choice decisions about the objects' locations [14].

2.2 Cognitive Measures in AR Systems

There have been few user studies conducted with AR systems; most such studies (including ours) have been at the perceptual level, such as those described above. The recent emergence of hardware capable of delivering sufficient performance to achieve stable presentation of graphics does enable such studies, however. One example of a cognitive-level study is the application of AR to medical interventions with ultrasound guidance [15]. In this trial, a doctor performed ultrasound-guided needle biopsies with and without the assistance of an AR system that had been designed for the task. A second physician evaluated the needle placement of the first. The analysis showed that needle localization was improved when using the AR system. The performance metrics in this trial were the standard for evaluating doctors' performance used by medical schools: needle placement at various locations within the target lesion. The physician uses the ultrasound to determine the ideal and actual needle locations. Thus the measure is tightly connected to the task, and in fact exists prior to the development of the AR system.

3 Experiment

As noted above, we have begun our performance measurements with the subsystem that depicts occluded surfaces. The first test we performed was a perceptual experiment to determine whether the system provides sufficient information for the user to understand three layers of depth among large objects that are occluded from view.

3.1 Design Methodology

From our initial design work and review by colleagues, we selected three graphical parameters to vary in our representations: drawing style, opacity, and intensity. These comprised a critical yet tenable set of parameters for our study. We used an urban environment that fit our laboratory facilities. By sitting in the atrium of our building, a user could wear an indoor-based version of our system (which is more powerful than the current mobile prototypes). The environment included one physically visible building and two occluded buildings. Among the two occluded buildings we placed one target to locate in one of three different positions: closer than the two occluded buildings, between the two, or behind both. This introduced the question of whether the ground plane (i.e. perspective) would provide the only cue that users would actually use. Because our application may require users to visualize objects that are not on the ground or are at a great distance across hilly terrain, we added the use of a consistent, flat ground plane for all objects as a parameter.

3.2 Hardware

The hardware for our AR platform consisted of three components. For the image generator, we used a Pentium IV 1.7 GHz computer with an ATI FireGL2 graphics card (outputting frame-sequential stereo). For the display device, we used a Sony Glasstron LDI–100B stereo optical see-through display (SVGA resolution, 20° horizontal field of view in each eye). The user was seated indoors for the experiment and was allowed to move and turn the head and upper body freely while viewing the scene, which was visible through an open doorway to the outdoors. We used an InterSense IS-900 6-DOF ultrasonic/inertial hybrid tracking system to track the user's head motion to provide a consistent 3D location for the objects as the user viewed the world. The IS-900 provides position accuracy to 3.0 mm and orientation accuracy to 1.0°.

The user entered a choice for each trial on a standard extended keyboard, which was placed on a stand in front of the seat at a comfortable distance. The display device, whose transparency can be adjusted in hardware, was set for maximum opacity of the LCD, to counteract the bright sunlight that was present for most trials. Some trials did experience a mix of sunshine and cloudiness, but the opacity setting was not altered. The display brightness was set to the maximum.

The display unfortunately does not permit adjustment of the inter-pupillary distance (IPD) for each user. If IPD is too small, then the user will be seeing slightly cross-eyed and tend to believe objects are closer than they are. The display also does not permit adjusting the focal distance of the graphics. The focal distance of the virtual objects is therefore closer than the real object that we used as the closest obstruction. This would tend to lead users to believe the virtual objects were closer than they really were.

Stereo is considered a powerful depth cue at near-field distances (approximately 1.0 meters, or about at arm's length). At far-field distances, such as the task we gave our users, stereo is not considered to be a strong depth cue; however, we wanted to be able to provide some statistical evidence for this claim. Many practitioners of AR systems have noted that improper settings of parameters related to stereo imagery (such as IPD and vergence) can lead to user discomfort in the form of headaches or dizziness. None of users reported any such problems; they wore the device for an average of 30 minutes. These issues will need to be addressed in future versions of the hardware for AR systems, but are beyond the scope of our work.

3.3 Experimental Design

3.3.1 Independent Variables

From our heuristic evaluation and from previous work, we identified the following independent variables for our experiment. These were all *within-subject* variables; every user saw every level of each variable.

Drawing Style ("wire", "fill", "wire+fill"): Although the same geometry was visible in each stimulus (except for which target was shown), the representation of that geometry was changed to determine what effect it had on depth perception. We used three drawing styles (Figure 2). In the first, all objects are drawn as wireframe outlines. In the second, the first (physically visible) object is drawn as a wireframe outline, and all other objects are drawn with solid fill (with no wireframe outline). In the third style, the first object is in wireframe, and all other layers are drawn with solid fill with a white wireframe outline. Backface culling was on for all drawing styles, so that the user saw only two faces of any occluded building.

Opacity (constant, decreasing): We designed two sets of values for the α channel based on the number of occluding objects. In the "constant" style, the first layer (visible with registered wireframe outline) is completely opaque, and all other layers have the same opacity ($\alpha = 0.5$). In the "decreasing" style, opacity changes for each layer. The first (physically visible, wireframe) layer is completely opaque. The successive layers are not opaque; the α values were 0.6, 0.5, and 0.4 for the successively more distant layers.

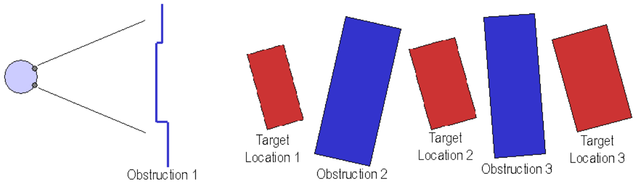

Figure 3. The experimental design (not to scale) shows the user position at the left. Obstruction 1 denotes the visible surfaces of the physically visible building. The distance from the user to obstruction 1 is approximately 60 meters. The distance from the user to target location 3 is approximately 500 meters, with the obstructions and target locations roughly equally spaced.

Intensity (constant, decreasing): We used two sets of intensity modulation values. The modulation value was applied to the object color (in each color channel, but not in the opacity or α channel) for the object in the layer for which it was specified. In the "constant" style, the first layer (visible with registered wireframe outline) has full intensity (modulator=1.0) and all other layers have intensity modulator=0.5. In the "decreasing" style, the first layer has its full native intensity, but successive layers are modulated as a function of occluding layers: 0.75 for the first, 0.50 for the second, and 0.25 for the third (final) layer.

Target Position (close, middle, far): As shown in the overhead map view (Figure 3), there were three possible locations for the target.

Ground Plane (on, off): From the literature and everyday experience, we know that the perspective effects of the ground plane rising to meet the horizon and apparent object size are a strong depth cues. In order to test the representations as an aide to depth ordering, we removed the ground plane constraint in half of the trials. The building sizes were chosen to have the same apparent size from the users' location for all trials. When the ground plane constraint was not present in the stimulus, the silhouette of each target was fixed for a given pose of the user. In other words, targets two and three were not only scaled (to yield the same apparent size) but also *positioned vertically* such that all three targets would occupy the same pixels on the 2D screen for the same viewing position and orientation. No variation in position with respect to the two horizontal dimensions was necessary when changing from using the ground plane to not using it. The obstructions were always presented with the same ground plane. We informed the users for which

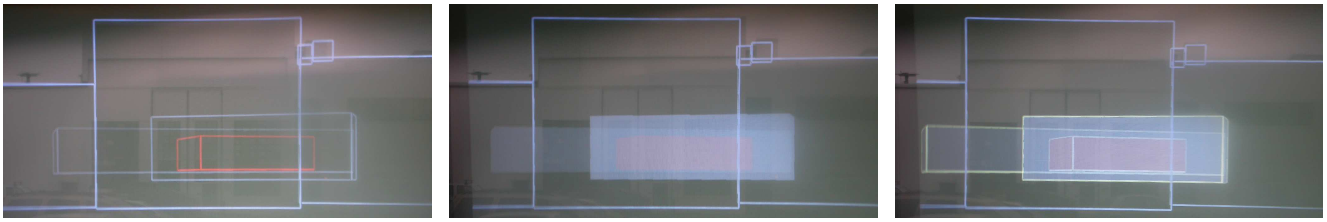

Figure 2. User's view of the stimuli. ***Left*****: "wire" drawing style.** ***Center*****: "fill" drawing style.** ***Right*****: "wire+fill" drawing style. The target (smallest, most central box) is between (position "middle") obstructions 2 and 3 in all three pictures. These pictures were acquired by placing a camera to the eyepiece of the HMD, which accounts for the poor image quality. The vignetting and distortion are due to the camera lens and the fact that it does not quite fit in the exit pupil of the HMD's optics.**

half of the session the ground plane would be consistent between targets and obstructions.

We did this because we wanted to remove the effects of perspective from the study. Our application requires that we be able to visualize objects that may not be on the ground, may be at a distance and size that realistic apparent size would be too small to discern, and may be viewed over hilly terrain. Since our users may not be able to rely on these effects, we attempted to remove them from the study.

Stereo (on, off): The Sony Glasstron display receives as input left- and right-eye images. The IPD and vergence angle are not adjustable, so we can not provide a true stereo image for all users. However, we can present images with disparity (which we call "stereo" for the experiment) or present two identical images ("biocular").

Repetition (1, 2, 3): Each user saw three repetitions of each combination of the other independent variables. It is well-known that users will often improve their performance with repetition of the same stimulus within an experiment. By repeating the stimuli, we can gain some insight into whether the user needs to learn how the system presents cues or whether the system presents intuitive cues. If there is no learning effect with repetition of stimuli, then we can infer that the users had whatever collective performance they achieved intuitively.

3.3.2 Dependent Variables

For each trial, we recorded the user's (three-alternative forced) choice for the target location and the time the user took to enter the response after the software presented the stimulus. We opted to ask the user only to identify the ordinal depth, not an absolute distance between the graphical layers. This implied the forced-choice design.

All combinations of these parameters were encountered by each user; however, the order in which these were presented was also randomly permuted. Thus each user viewed 432 trials. The users ranged in time from twenty to forty minutes for the complete set of trials. The users were told to make their best guess upon viewing the trial and not to linger; however, no time limit per trial was enforced. The users were instructed to aim for a balance of accuracy and speed, rather than favoring one over the other.

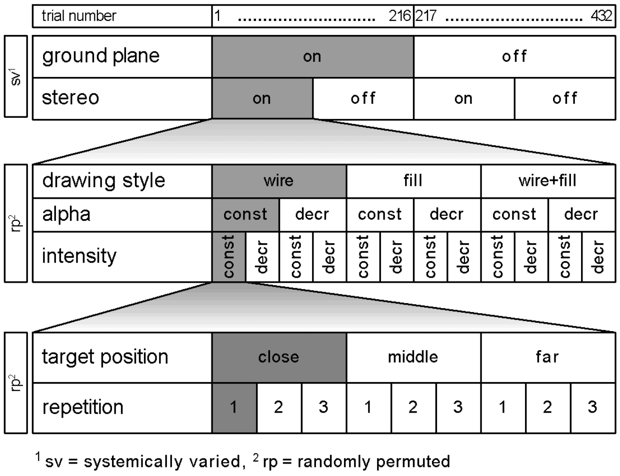

Figure 4. Experimental design and counterbalancing for one user. Systematically varied parameters were counterbalanced between subjects.

3.3.3 Counterbalancing

In order to reduce time-based confounding factors, we counterbalanced the stimuli. This helps control learning and fatigue effects within each user's trials and factors such as the amount of sunshine that change between subjects beyond our control. Figure 4 describes how we counterbalanced the stimuli. We observed (in conjunction with many previous authors) that the most noticeable variable was the presence of the ground plane [3, 18]. In order to minimize potentially confusing large-scale visual changes, we gave ground plane and stereo the slowest variation. Following this logic, we next varied the parameters which controlled the scene’s visual appearance (drawing style, alpha, and intensity), and within the resulting blocks, we created nine trials by varying target position and repetition.

3.4 Experimental Task

We designed a small virtual world that consisted of four buildings (Figure 3), with three potential target locations. The first building was an obstruction that corresponded (to the limit of our modeling accuracy) to a building that was physically visible during the experiment. The obstructions were always drawn in blue; the target always appeared in red. The target was scaled such that its apparent 2D size was equal, regardless of its location. Obstructions 2 and 3 roughly corresponded to real buildings. The three possible target locations did not correspond to real buildings.

The task for each trial was to determine the location of the target that was drawn. The user was shown the overhead view before beginning the experiment. This helped them visualize their choices and would be an aide available in a working application of our system. The experimenter explained that only one target would appear at a time. Thus in all of the stimulus pictures, four objects were visible: three obstructions and the target. For the trials, users were instructed to use the number pad of a standard extended keyboard and press a key in the bottom row of numbers (1–3) if the target were closer than obstructions 2 and 3, a key in the middle row (4–6) if the target were between obstructions 2 and 3, or a key in the top row (7–9) if the target were further than obstructions 2 and 3. A one-second delay was introduced between trials within sets, and a rest period was allowed between sets for as long as the user wished. We showed the user 48 sets of nine trials each. The users reported no difficulties with the primitive interface after their respective practice sessions. The users did not try to use head motion to provide parallax, which is not surprising for a far-field visualization task.

3.5 Subjects

Eight users completed the experiment (432 trials each). All subjects were male and ranged in age from 20 to 48. All volunteered and received no compensation. Our subjects reported being heavy computer users. Two were familiar with computer graphics, but none had seen our representations. Subjects did not have difficulty learning or completing the experiment.

Before the experiment, we asked users to complete a stereo acuity test, in case stereo had produced an effect. The test pattern consisted of nine shapes containing four circles each. For each set of four circles, the user was asked to identify which circle was closer than the other three. Seven users answered all nine test questions correctly, while the other user answered eight correctly.

4 Hypotheses

We made the following hypotheses about our independent variables.

1. The ground plane would have a strong positive effect on the user’s perception of the relative depth.
2. The wireframe representation (our system’s only option before this study) would have a strong negative effect on the user’s perception.
3. Stereo imagery would not yield different results than biocular imagery, since all objects are in the far-field [3].
4. Decreasing intensity would have a strong positive effect on the user’s perception for all representations.
5. Decreasing opacity would have a strong positive effect on the user’s perception of the “fill” and “wire+fill” representations. In the case of wireframe representation the effect would be similar to decreasing intensity. Apart from the few pixels where lines actually cross, decreasing opacity would let more and more of the background scene shine through, thereby indirectly leading to decreased intensity.

5 Results

There are a number of error metrics we apply to the experimental data. Figure 5 categorizes the user responses. Subjects made 79% correct choices and 21% erroneous choices. We found that subjects favored the far position, choosing it 39% of the time, followed by the middle position (34%), and then by the close position (27%). We also found that subjects were the most accurate in the far position: 89% of their choices were correct when the target was in the far position, as compared to 76% correct in the close position, and 72% correct in the middle position.

As discussed above, we measured two dependent variables: ***user response time***, and ***user error***. For user response time, the system measured the time in milliseconds (ms) between when it drew the scene and when the user responded. Response time is an interesting metric because it indicates how intuitive the representations are to the user. We want the system to convey information as naturally as the user’s vision does in analogous real-world situations.

For user error, we calculated the metric $e = |a - u|$, were a is the actual target position (between 1 and 3), and u is the target position chosen by the user (also between 1 and 3). Thus, if $e = 0$ the user has chosen the correct target; if $e = 1$ the user is off by one position, and if $e = 2$ the user is off by two positions.

We conducted significance testing for both response time and user error with a standard analysis of variance

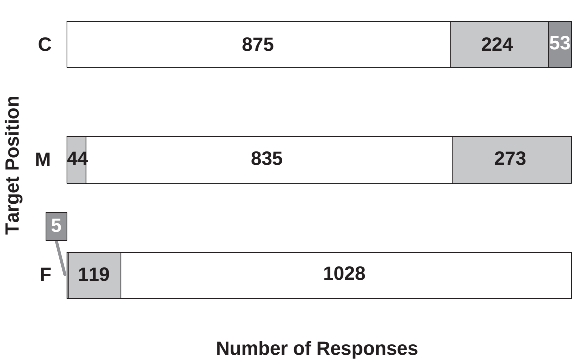

Figure 5. User responses by target position. For each target position, the bars show the number of times subjects chose the (C)lose, (M)iddle, and (F)ar positions. Subjects were either correct when their choice matched the target position (white), off by one position (light gray), or off by two positions (dark gray).

(ANOVA) procedure. In the summary below, we report user errors in positions (pos).

We briefly discuss the factors that affected user performance. As we expected, subjects were more accurate when a ground plane was present (.1435 pos) then when it was absent (.3056 pos). Interestingly, there was no effect of ground plane on response time $(F < 1)$. This indicates that subjects did not learn to just look at the ground plane and immediately respond from that cue alone, but were in fact also attending to the graphics.

Figure 6 shows that subjects were slower using the "wire" style than the "fill" and "wire+fill" styles. Subjects had the fewest errors with the "wire+fill" style. These results verified our hypotheses that the "wire" style would not be very effective, and the "wire+fill" style would be the most effective, since it combines the occlusion properties of the "fill" style with the wireframe outlines, which help convey the targets' shapes.

Subjects were more accurate with decreasing opacity (.1962 pos) than with constant opacity (.2529 pos). This makes sense because the decreasing opacity setting made the difference between the layers more salient. Subjects were both faster (2340 versus 2592 ms) and more accurate (.1811 versus .2679 pos) with decreasing intensity. This result was expected, as decreasing intensity did a better job of differentiating the different layers. However, Figure 7 shows that the effect on response time is due to the difference between constant and decreasing intensity when the target is drawn in the "wire" style.

As expected from training effects, subjects became faster with repetition. However, repetition had no effect on absolute error $(F < 1)$, so although subjects became faster, they did not become more accurate. This can be taken as a sign that the presented visuals were understandable for the subjects right from the outset. No learning effect took place regarding accuracy. Subjects became faster, though, which is a sign that their level of confidence increased.

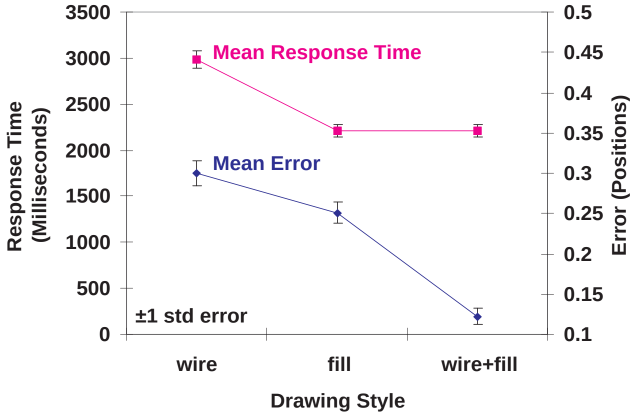

Figure 6. Main effect of drawing style on response time (□) and error (◇).

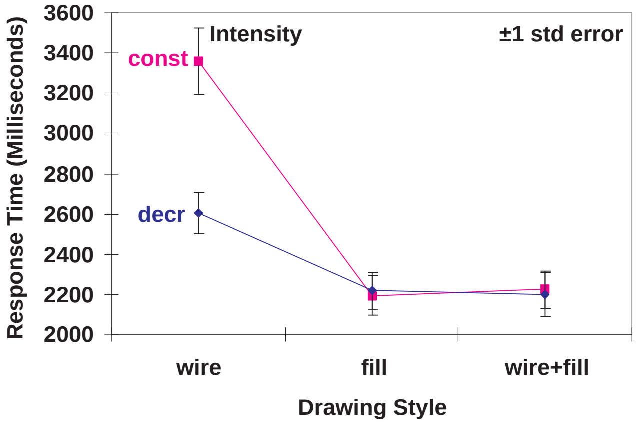

Figure 7. Drawing style by intensity (constant (□), decreasing (◇)) interaction on response time.

6 Discussion

In a broad context, we believe that our methodology will enable us to evaluate both system capabilities and user performance with the system. Human perception is an innate ability, and variations in performance will reflect the system's appropriateness for use by dismounted warfighters. Thus, we are really evaluating the system's performance by measuring the user's performance on perceptual-level tasks. The evaluation of cognitive-level tasks will enable us to determine how users are performing. Such high-level metrics can only be measured after the results of the perceptual-

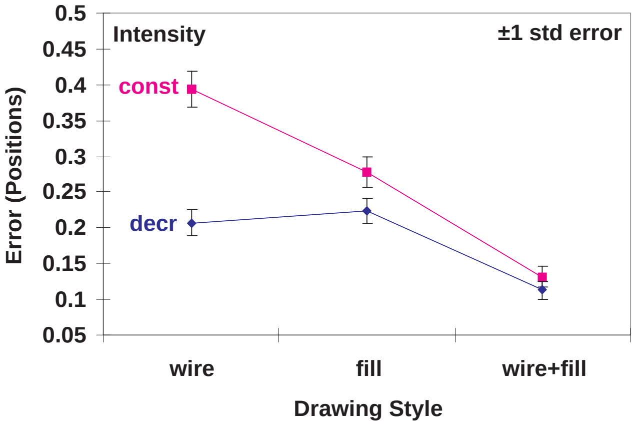

Figure 8. Drawing style by intensity (constant (□), decreasing (◇)) interaction on absolute error.

level tests inform the system design.

Our first experiment has given insight into how users perceive data presented in the system. The application of our results to human perception and thus our system design are straightforward. It is well-known that a consistent ground plane (a perspective constraint) is a powerful depth cue. However, we can now provide statistical backing for our fundamental hypothesis that graphical parameters can provide strong depth cues, albeit not physically realistic cues. We found that with the ground plane on the average error was .144 pos, whereas the with the ground plane off and the following settings:

- drawing style: "wire+fill"
- opacity: decreasing
- intensity: decreasing

the average error was .111 pos. The data thus suggest that we did find a set of graphical parameters as powerful as the presence of the ground plane constraint. This would indeed be a powerful statement, but requires further testing before we can say for sure whether this is our finding. As a secondary result, the fact that there was a main effect of repetition on response time but not on accuracy indicates that the subjects could quickly understand the semantic meaning of the encodings. This validates that BARS is performing at a level that is sufficient for users to consistently (but not always) identify the ordinal depth among three occluded objects.

There are several next steps available to us. Further perceptual-level testing will demonstrate whether these results extend to more complex scenes (with more layers of depth). We are currently designing a follow-up study that will use not just an ordinal depth metric, but an absolute distance metric. This study will task the user to move a virtual object into depth alignment with real objects. We are developing metrics to apply to the user's control of the object, such as the number of oscillations they use to place the object into position, that will give us insight into their confidence in the depth estimates they perceive through BARS. We are also considering ways in which to measure the user's subjective reaction to the system, as this is also an important aspect of the system's capabilities.

Once these results inform our future system design, we will move up to cognitive-level testing, in which we hope to have multiple users wear prototype systems in an urban environment. We can have users identify locations of objects relative to maps or to each other. We could have users retrieve objects from the environment. The metrics we plan to use will reflect the cognition required. Distance and response time will remain interesting measures, but now the absolute distance will become more important. We will be able to add directional measures as well, concomitant with the increased complexity of the task for a mobile user. Since our application is designed for a military context, we intend to design our cognitive-level tests in conjunction with military domain experts and have at least some of the subjects in our studies be active members of the military. This introduces the opportunity to measure system performance by comparing against current performance of dismounted warfighters in these tasks. This combined design and evaluation methodology will enable us to evaluate the Battlefield Augmented Reality System and its users.

References

[1] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads up display technology to manual manufacturing processes. In *Proceedings of Hawaii International Conference on System Sciences*, volume II, pages 659–669. IEEE Computer Society Press, Jan. 1992.

[2] Concepts Division, Marine Corps Combat Development Command. A concept for future military operations on urbanized terrain, July 1997.

[3] J. E. Cutting. How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments, and Computers*, 29(1):29–36, 1997.

[4] P. Edwards, D. Hawkes, D. Hill, D. Jewell, R. Spink, A. Strong, and M. Gleeson. Augmented reality in the stereo operating microscope for otolaryngology and neurological guidance. In *Medical Robotics and Computer Assisted Surgery*, Sept. 1995.

[5] S. R. Ellis and B. M. Menges. Localization of object in the near visual field. *Human Factors*, 40(3):415–431, Sept. 1998.

[6] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In *International Symposium on Wearable Computing (ISWC)*, pages 74–81, Oct. 1997.

[7] S. Feiner, B. MacIntyre, and D. Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):52–62, July 1993.

[8] C. Furmanski, R. Azuma, and M. Daily. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, pages 215–224, Sept. 2002.

[9] J. L. Gabbard, J. E. Swan II, D. Hix, M. Lanzagorta, M. A. Livingston, D. Brown, and S. Julier. Usability engineering: Domain analysis activities for augmented reality systems. In *Proceedings of SPIE (International Society for Optical Engineering)*, The Engineering Reality of Virtual Reality 2002, Jan. 2002.

[10] M. A. Livingston, J. L. G. J. Edward Swan II, T. H. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR2003)*, pages 56–65. IEEE, Oct. 2003.

[11] M. A. Livingston, L. J. Rosenblum, S. J. Julier, D. Brown, Y. Baillot, J. E. Swan II, J. L. Gabbard, and D. Hix. An augmented reality system for military operations in urban terrain. In *Interservice/Industry Training, Simulation, and Education Conference*, page 89, Dec. 2002.

[12] U. Neumann and A. Majoros. Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In *Proceedings of IEEE Virtual Reality Annual International Symposium*, pages 4–11, 1998.

[13] J. P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287–309, June 2000.

[14] J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence: Teleoperators and Virtual Environments*, 11(6):610–625, Dec. 2002.

[15] M. Rosenthal, A. State, J. Lee, G. Hirota, J. Ackerman, K. Keller, E. D. P. MD, M. Jiroutek, K. Muller, and H. Fuchs. Augmented reality guidance for needle biopsies: A randomized, controlled trial in phantoms. In *Lecture Notes in Computer Science: Medical Image Computing and Computer-Assisted Interventions (MICCAI)*, volume 2208, pages 240–248, Oct. 2001.

[16] A. State, D. T. Chen, C. Tector, A. Brandt, H. Chen, R. Ohbuchi, M. Bajura, and H. Fuchs. Case study: Observing a volume-rendered fetus within a pregnant patient. In *Proceedings of IEEE Visualization '94*, pages 364–368, 1994.

[17] A. State, M. A. Livingston, G. Hirota, W. F. Garrett, M. C. Whitton, E. D. Pisano MD, and H. Fuchs. Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies. In *SIGGRAPH 96 Conference Proceedings*, Annual Conference Series, pages 439–446. ACM SIGGRAPH, Addison Wesley, Aug. 1996.

[18] R. T. Surdick, E. T. Davis, R. A. King, and L. F. Hodges. The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence: Teleoperators and Virtual Environments*, 6(5):513–531, Oct. 1997.

[19] B. Thomas, B. Close, J. Donoghue, J. Squires, P. D. Bondi, M. Morris, and W. Piekarski. ARQuake: An outdoor/indoor augmented reality first person application. In *International Symposium on Wearable Computers*, pages 139–146, Oct. 2000.

[20] A. Webster, S. Feiner, B. MacIntyre, W. Massie, and T. Krueger. Augmented reality in architectural construction, inspection, and renovation. In *Proceedings of the Third ASCE Congress for Computing in Civil Engineering*, June 1996.