EVALUATING THE USABILITY OF A NEXT-GENERATION HEADS-UP DISPLAY FOR FIREFIGHTERS IN A VIRTUAL ENVIRONMENT

Jason D. Saville[1], Randall D. Spain[1], Donia Slack[2], Edward Hill[2], and James Lester[1]
[1]North Carolina State University, Center for Educational Informatics, Raleigh, NC, USA
[2]RTI International, Research Triangle Park, NC, USA

Virtual environments provide a rich and immersive platform to investigate how interfaces should be designed to support public safety operations. Virtual environments can be used to simulate the tasks, demands, and conditions of emergency response events while providing a controlled and safe environment for testing and developing new user interface capabilities. This paper presents an evaluation of a prototype AR-based heads-up display (HUD) that was integrated in a desktop-based virtual environment that simulated an emergency response scenario. Members of the first responder community (N = 90) completed a series of emergency response tasks and responded to a series of usability and user experience questions and workload ratings. The study used a 2 (Condition: HUD vs. No HUD) x 2 (Task Load: Radio monitoring task vs. No radio monitoring task) mixed experimental design. Participants provided very positive feedback on the utility of the prototype HUD for supporting task performance. Results and future research plans are discussed.

INTRODUCTION

Public safety communication technologies are advancing at a rapid pace. Location-based sensors, smart buildings, and mission-critical voice technologies will soon afford first responders the ability to access new forms of information during emergency response. To maximize the value of this information, user interfaces must be designed to allow first responders to interact effectively with them and respond in an effective and timely manner, without inducing undue errors or excessive mental workload.

Technology development and testing for the first responder community requires significant time and resources (Morrison et al., 2021; Rao et al., 2014). Advances in virtual environment (VE) technologies such as desktop-based VEs, virtual reality (VR), mixed reality (MR), and augmented reality (AR) enable new opportunities to support user experience research for public safety technologies (Grandi et al., 2019; Spain et al., 2020; Suhail et al., 2019). By creating highly immersive scenarios, researchers can use VEs as a testbed to systematically investigate human-computer interaction principles in a controlled and safe environment, allowing for engaging interactions with prototype interfaces that can be rapidly tested and refined (Rebelo et al., 2012).

Prior work using VEs to support interface design research has investigated how data should be presented through HUDs to support operators' needs in high-stakes environments. For instance, Zaman et al. (2021) used VR to examine the utility of a prototype AR-based display to support military subterranean operations. The display provided navigation information to support wayfinding during a simulated search and rescue mission. Results showed the VR environment provided a valuable testbed for assessing the impact of the prototype AR-based display on participants' ratings of mental workload, usability, and situational awareness.

Grandi et al. (2021) also explored how VR could be used to support research on future interfaces. Utilizing a user-centered design approach, these researchers designed an AR-based HUD projected in VR to support law enforcement officers during traffic stops and firefighters during search and rescue tasks.

In this paper, we describe a study that investigated the usability of a prototype AR-based HUD with members of the first responder community. The prototype HUD was originally designed for a VR application but was converted into a desktop-based VE to facilitate remote testing during a pandemic. The prototype HUD provided first responders with navigation support, goal-based prompts, and additional forms of task-critical information as they completed a simulated search and rescue mission from a first-person perspective. While in the environment, users could interact with a simulated AR-based HUD to summon and dismiss different forms of information and data displays using keys on the keyboard.

In the following sections we describe the VE that was designed to support the study, the study procedures, and our study results. The goal was to identify features of the HUD that offered the most utility for completing the simulated search and rescue mission and determine whether the HUD reduced mental workload during periods of high task load. We hypothesized that the HUD would reduce mental workload compared to completing the set of simulated missions without the HUD and features such as navigational support would provide utility to end users for completing the simulated scenarios.

METHOD

Recruitment

We advertised the study to potential participants through a first responder mailing list and a blog post on a federal public safety communications technologies website. To be eligible, participants had to be at least 18 years of age, be able to comfortably complete desktop-based tasks for at least 90 minutes, and have previous experience working as a first responder or working with the first responder community. Participants were self-selected and received a $20 dollar gift card for completing the hour-long study. Participants completed the study remotely through a web browser on their own computer (laptop or desktop) at their own convenience.

Participants

Data were gathered from 90 participants over the course of a four-week period. The average age of participants was 33 years (range 21-70); 58% of the sample was 30 years of age or younger. Approximately 77% of participants identified as male (22% identified as female, 1% as non-binary). Among all participants, 64% reported having served as a first responder, with representatives from fire services (97%) and emergency medical services (3%). First responders' work experience ranged from presently being a first responder (90%), serving as a first responder in the past year (5%), serving within the last 5 years (2%), and serving more than five years ago (3%).

Measures

Graphical User Interface Questionnaire (GUI). We used an 18-item instrument designed by the research team to identify which features of the HUD provided the most value for completing the mission (Spain et al., 2020). Items included a mix of Likert-type items where higher scores represented greater HUD effectiveness, utility, or usability, and free response items that asked participants specific questions about their experiences with the HUD.

NASA Task Load Index (NASA TLX). Mental workload ratings were captured after each mission using the NASA TLX (Hart & Staveland, 1988). Participants self-reported the level of mental demand, physical demand, temporal demand, effort, performance, and frustration they experienced while completing the mission on an interval scale that ranged from 0 (*low*) to 20 (*high*).

Tasks and Procedures

The online study consisted of seven activities: (1) informed consent, (2) study introduction video, (3) controller tutorial, (4) radio monitoring task tutorial, (5) practice mission, (6) two performance missions, and (7) post-mission surveys.

After reading and signing the online consent form, participants were automatically directed to a short online video that provided an overview of the study goals. The video informed participants that the purpose of the study was to investigate how data should be presented through a prototype AR-based HUD to facilitate better task performance in stressful environments. They were informed that they would be assuming the role of a firefighter and completing a series of tutorials and several missions in a virtual environment. For each mission they were responsible for navigating through a metro station, finding a disabled train, and extracting any passengers that needed help. Participants were also told that they would hear a dispatcher stating radio call signs, and that their task, in addition to completing the mission objectives, was to respond to the dispatcher when they heard a target call sign. They were also informed that following the missions they would complete several online surveys. At the conclusion of the video, participants advanced to the desktop-based virtual environment which included menu options for accessing the study's tutorials and missions.

Next, participants completed two tutorials. The first tutorial provided guided instruction on how to use the keyboard and mouse to interact with objects, navigate the virtual environment, and interact with the prototype HUD to call and dismiss different information that would be helpful for completing each mission. The prototype AR-based HUD provided participants with navigational assistance, mission objective prompts, a chat window that provided text translations of radio communications, an edge detection feature that illuminated object outlines in low visibility conditions, a dynamic air display that provided information regarding the number of minutes of air remaining in the air tank, and a temperature gauge (Figure 1). Participants used their keyboard and mouse to navigate through the environment, interact with objects in the virtual environment, respond to the secondary radio monitoring task, and activate the HUD features.

Figure 1. Virtual scenario with prototype AR-based HUD.

The second tutorial provided instructions for completing the radio monitoring task. It also included a short practice session that required participants to monitor a series of dispatch calls, listen for when the sum of two numerical operator call signs would equal an even number (e.g., Battalion 3, 5th truck company), and respond correctly to three target radio calls in a row. Visual feedback was provided about the accuracy of each response.

Upon completing both tutorials, participants completed a practice mission (Mission 1) which allowed participants to familiarize themselves with using the keyboard controls, interacting with the HUD, and performing the radio monitoring task and mission requirements concurrently. All participants had access to the prototype AR-HUD during the practice mission and completed the radio monitoring task while performing the search and rescue mission.

The practice mission began with a short briefing that provided narrative background and instructions for completing the mission. Participants were encouraged to memorize the list of tasks that they needed to complete but were also informed that if they were unable to memorize the list of tasks, they would be able to summon a visual prompt through the HUD by pressing the appropriate key. The mission tasks included unlocking a catwalk gate, interacting with a non-player firefighter character to confirm power to the third rail was deactivated, locating and entering the disabled metro train, and safely egressing with any passenger that needed assistance to the nearest mezzanine exit.

After the mission briefing, participants began the mission. Participants were instructed to complete the mission and radio monitoring task as quickly and accurately as possible. Approximately every 15 (+/- 2) seconds, participants would hear a radio dispatch, at which point they would have approximately 12 seconds to evaluate if the sum of the call signs was an even number and respond if appropriate. There were approximately 15 minutes allotted for the mission, at which point the mission would end prematurely due to air tank depletion if participants had not finished all tasks.

After the practice mission, participants were provided visual feedback on how many objectives they completed, their performance accuracy on the radio monitoring task, and completion time. Participants rated the level of mental workload they experienced during the mission by completing an embedded version of the NASA TLX (Hart & Staveland, 1988), then returned to the main menu and advanced to Mission 2.

Mission 2 required participants to complete the same set mission tasks as the practice mission; however, the location of the train and the location of the injured passenger were altered. The scenario also included degraded visibility conditions which made wayfinding in the virtual environment more challenging. After Mission 2 ended, participants completed the NASA-TLX and then advanced to Mission 3. Mission 3 was unique compared to the practice mission and Mission 2 in that it included additional distractor trains and an extra task of turning off the power to the third rail at a designated power control box. After completing Mission 3 and the associated NASA-TLX, participants were provided with a hyperlink that they copied and pasted into a web browser to begin the post-mission surveys that were presented in Qualtrics. The surveys included a brief demographic questionnaire and the graphical user interface (GUI) questionnaire.

Design

The study followed a 2 (Condition: HUD vs. No-HUD) x 2 (Task Load: Radio monitoring task vs. No radio monitoring task) mixed experimental design in which participants were randomly assigned to either the HUD or no-HUD condition. Participants in the HUD condition (*n* = 45) had access to the prototype AR-based HUD while completing Missions 2 and 3; No-HUD condition participants (*n* = 45) did not have access to the HUD during these missions. Task load, represented as the radio monitoring task, served as a within-subject factor and was counterbalanced across Missions 2 and 3. This resulted in four groups: 1) access to HUD, monitoring task in M2 (*n* = 25); 2) no access to HUD, monitoring task in M2 (*n* = 19); 3) access to HUD, monitoring task in M3 (*n* = 20); and 4) no access to HUD, monitoring task in M3 (*n* = 26).

RESULTS

Graphical User Interface (GUI) Questionnaire

A summary of participant responses to selected items from the GUI questionnaire is presented below. Responses from both HUD and non-HUD conditions are combined for each item unless otherwise stated; for quantitative items, we combined responses only if group means were not significantly different ($p > .05$) when tested via one-way ANOVA. Since every individual used the HUD at least during Mission 1, we intended to display all items to everyone. Due to a survey logic error, not all participants responded to every item; thus, response rates are reported for each item.

How effective were the HUD elements for completing the mission? The first question in the GUI questionnaire asked participants to rate the effectiveness of HUD elements for completing the mission. Response options included 1 (*Not effective at all*), 2 (*Slightly effective*), 3 (*Moderately effective*), 4 (*Very effective*), and 5 (*Extremely effective*). Results showed that, overall, the informatics provided through the HUD offered a great deal of utility towards completing the mission, with average ratings for each HUD element centered around *Very effective* (*n* = 70; Figure 2).

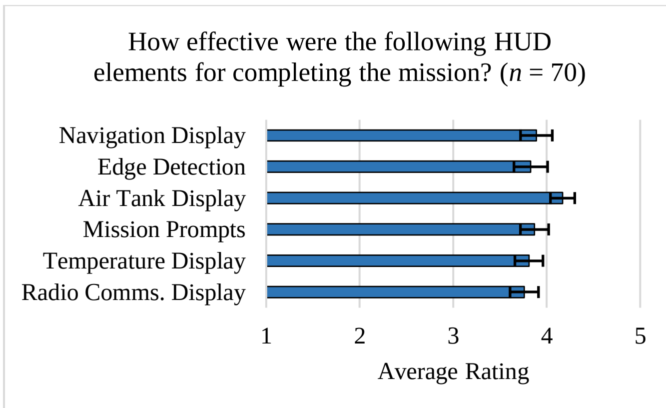

Figure 2. Ratings of HUD utility for mission completion.

Explored further, results of a repeated measures ANOVA was significant, indicating a significant difference between effectiveness ratings of HUD elements $F(5, 65)$, 3.66, $p < .01$, partial eta-squared = .22, observed power = .91. Follow-up pairwise comparisons revealed that the air tank display (M = 4.17, SD = 1.09) received significantly higher ratings than edge detection (M = 3.83, SD = 1.50, p = .03), temperature display (M = 3.81, SD = 1.25, $p < .01$), and radio display (M = 3.76, SD = 1.27, $p < .01$). There were no other significant differences between HUD features.

What HUD information was most helpful for completing the task? Why? A second question asked participants summarize what HUD information they found to be most helpful for completing the scenario and why. Responses were collected from participants (*n* = 63) and grouped into thematic categories for frequency analysis. The categories were navigation (*n* = 25), edge detection (*n* = 14), mission prompts (*n* = 8), temperature display (*n* = 4), air tank readings (*n* = 3), radio communication display (*n* = 1), and other (*n* = 8). Participants mentioned the navigation display most frequently as being most helpful for completing the task, followed by the edge detection feature (Figure 3).

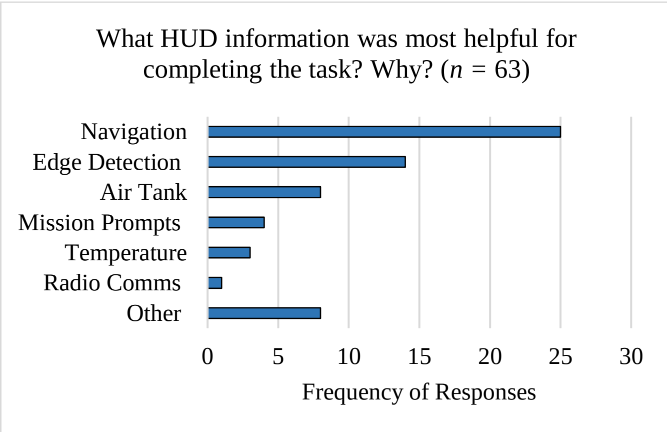

Figure 3. Most helpful HUD information for task completion.

Did you use the radio communication display to help complete the radio monitoring task? Why or why not? One item asked participants whether they made use of the radio communication display to help complete the radio monitoring task, as well as why or why not. Responses were gathered from (*n* = 32) participants in the HUD condition via a free-response format, where 19 participants reported having used the radio communication display and 13 reported not having used it. Some of the most frequent reasons for making use of the display included being able to clarify what was verbally said, especially when distracted or facing multiple stressors (*n* = 5) and having an alternative format to consume the same information was useful (*n* = 2). Some of the most frequent reasons for not making use of the display included a lack of need (*n* = 6) and too much going on already such that the display would be more distracting than helpful (*n* = 2).

HUD interaction preferences. Participants were asked to "*imagine that you are a firefighter using a helmet equipped with this HUD to complete a real mission. Which would work better to communicate/interact with the HUD, a 'hands-free' or a 'push-to-talk' speech interface or something else?"* Participants (*n* = 70) reported that they would most prefer a hands-free interface (58%), followed by push-to-talk (36%), while a minority of participants reported a preference for something else (6%). This suggests that user interface testing should explore the utility of using a hands-free voice interface for interacting with the HUD to determine if the hands-free component leads to increased usability, reduced errors and cognitive load, and improved mission performance compared to other methods of HUD interaction such as push-to-talk.

Can you describe the worst interaction you had with the system? What were you doing? How can this be improved? Participants were then asked to describe their worst interaction with the system, including what they were doing at the time and how the interaction could be improved. Of participants that responded (*n* = 70), 29 participants reported having no problems with the system. Six participants mentioned trouble with getting lost in the environment, and that prompts redirecting to the objective would be beneficial. Five participants mentioned troubles related to movement with the environment, including keyboard and mouse control functionality and artificial movement such as jumping. Four participants mentioned issues with multiple HUD elements presented simultaneously, whether this was due to screen sizes causing overlapping elements, elements blocking view of the environment, producing visual noise, or information overload. One suggestion included moving the chat window to the side of the screen, while another suggested moving the air tank display away from the center of the screen and including a color indicator when oxygen is low.

Which element of the HUD would be most important to improve? The following item asked participants which element of the HUD would be most important to improve. Of the total respondents (*N* = 65), the most frequent responses included the temperature display (*n* = 10; e.g., including temperature in degrees), radio communication (*n* = 9; e.g., including a small visual indicator when someone requires the user's response), mission prompts (*n* = 6), edge detection (*n* = 6), the oxygen tank/other vitals (*n* = 6), and navigation elements (*n* = 5). Other notable responses included overall visibility of the scenario past the HUD (*n* = 3), audio quality (*n* = 2), and communications with an ability to locate other first responders on the scene (*n* = 2).

Was the overall system user interface easy to understand and use? Participants also rated the overall usability of the AR-based HUD using a rating scale that ranged from 1 (*Extremely difficult*) to 5 (*Extremely easy*) to use. No significant differences (*p* > .05) between HUD (*M* = 4.22, *SD* = .97) or no-HUD (*M* = 4.51, *SD* = .82) conditions were found, allowing an examination of overall mean scores (*M* = 4.37, *SD* = .91).

Workload Ratings

Workload scores were calculated after each mission as the mean score of the six workload dimensions (range = 0 to 19). We conducted a series of between-subjects ANOVAs to analyze mental workload ratings under different task load conditions after Missions 2 and 3. Results showed that when participants performed the radio monitoring task during Mission 2, participants in the HUD condition reported lower levels of mental workload (*M* = 5.92, *SD* = 3.69) compared to participants in the No-HUD condition (*M* =8.57, *SD* = 4.15), *F*(1, 44), 5.22, *p* = .03, partial eta-squared = .11, observed power = .61. There was no significant difference in workload ratings between the HUD (*M* = 6.18, *SD* = 4.06) and No-HUD condition (*M* = 6.92, *SD* =3.22) when participants were not required to perform the secondary radio monitoring task, *F*(1, 43), .46, *p* = .50. Further analyses focusing on Mission 3 showed there were no significant differences in workload ratings between the HUD (*M* = 8.05, *SD* = 4.74) and No-HUD conditions (*M* = 6.52, *SD* = 3.80) when participants were required to perform the secondary monitoring task (*F*(1, 41), 1.37, *p* = .25) nor between HUD (*M* = 6.36, *SD* = 4.33) and No-HUD conditions (*M* = 8.57, *SD* = 5.48) when they did not (*F*(1, 42), 2.22, *p* = .14).

DISCUSSION

A primary goal of this study was to investigate the utility of a prototype AR-based HUD that was integrated into a VE that simulated a metro emergency response event. First responders participating in this study provided valuable insight into considerations for useful HUD design elements and features for emergency response. In particular, the air tank display was rated as being significantly more effective than some of the other HUD elements, and the navigation feature was rated and discussed as being the most helpful for completing the mission.

Another notable finding included a greater number of participants who reported a preference for a hands-free HUD design as opposed to a push-to-talk functionality or other design. However, one participant provided feedback in a separate survey section that a hands-free interface likely would not be sufficiently sensitive to firefighter commands in a loud environment. With this in mind, we suggest that further UI testing also examine how robust a given hands-free HUD design is at handling noisy environments.

Finally, we reasoned that access to critical information presented through the HUD would reduce the demands placed on participants by the environment and subsequently result in lower workload ratings. In fact, several participant responses noted the presence of certain elements allowed them to focus their attention elsewhere. However, results showed that the HUD only offered slight reductions in workload in Mission 2 when participants were required to perform the secondary radio monitoring task.

One limitation of this study is that participants in no-HUD conditions still had access to the HUD during the practice mission. It is possible that the removal of access to the HUD features may have inflated workload scores during Missions 2 for participants who did not have access to it. However, this possible result is likely inconsequential as mean values of workload were relatively low across all conditions and missions, with none reaching the 50% (10.0) possible score. The relatively low workload ratings suggest that the simulated tasks were not very demanding. The purposeful incorporation of additional demands, through either environmental elements, time pressure, or more immersive technologies such as head-mounted VR, may better facilitate an environment in which the impact of the prototype HUD can more easily be recognized.

While the shift from a head-mounted VR environment to a desktop-based environment was driven by safety considerations during a pandemic, it is important to note that there may be differences in preferences and mental workload ratings between these two formats. Therefore, future work should seek to gather data from first responders using VR technology.

Additionally, while participant responses show very favorable utility ratings for the HUDs, the positive reactions could be the result of response biases that were influenced by the novelty of the technology. Future research should continue to explore which HUD features provide the most utility for supporting emergency response performance by pairing user experience feedback with performance data. We plan to make use of additional data collected during the study to statistically explore the impact of access to HUD elements and task load on mission performance data.

CONCLUSION

A promising application of VEs for studying user experience is their use as a testbed for investigating how displays should be designed to support ease of use. HUDs offer a unique means for providing users with task-critical information. Understanding how new forms of information and data should be displayed and how users can interact and access this information is critical for designing intuitive user-centered technologies. Future research should continue to utilize the benefits of VEs to support rapid prototyping and evaluation of user interfaces for public safety operations.

ACKNOWLEDGEMENTS

The research described herein has been sponsored by the National Institute of Standards and Technology under grant number 60NANB18D148. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Grandi, J. G., Cao, Z., Ogren, M., & Kopper, R. (2021, March). Design and simulation of next-generation augmented reality user interfaces in virtual reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops* (*VRW)* (pp. 23-29). IEEE.

Grandi, J. G., Ogren, M., & Kopper, R. (2019, March). An approach to designing next generation user interfaces for public-safety organizations. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces* (*VR)* (pp. 944-945). IEEE.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*, (Vol. 52, pp. 139-183). North-Holland.

Morrison, K., Dawkins, S., Choong, Y. Y., Theofanos, M. F., Greene, K., & Furman, S. (2021). Current problems, future needs: Voices of first responders about communication technology. In *International Conference on Human-Computer Interaction* (pp. 381-399). Springer.

Rao, E., Remer, J., & Bauer, D. (2014). A model for development, transition and technology transfer leading to commercialization of security technology. In *2014 International Carnahan Conference on Security Technology (ICCST)* (pp. 1-5). IEEE.

Rebelo, F., Noriega, P., Duarte, E., & Soares, M. (2012). Using virtual reality to assess user experience. *Human Factors*, *54*(6), 964-982.

Spain, R., Saville, J., Liu, B., Slack, D., Hill, E., Holloway, J., Norsworthy, S., Mott, B., & Lester, J. (2020). Investigating a virtual reality-based emergency response scenario and intelligent user interface for first responders. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 2114–2118.

Suhail, M., Gainer, S., Haskins, J., Boyd, B., Laird, C., Huse, W., Eadara, S., & Jerald, J. (2019). Simulating a futuristic fire pump panel in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 1415-1416). IEEE.

Zaman, F., Drake, W., Intriligator, J., Gardony, A., Natick, M. A., & Rife, J. (2021). Investigating a virtual reality based subterranean scenario examining augmented reality implications for military operators. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *(*Vol. 65, No. 1) (pp. 1129-1133). SAGE Publications.