Evaluation of Immersive Interfaces for Tactical Decision Support

Mark Dennison[1]([✉]), Mark Mittrick[2], John Richardson[2],
Theron Trout[2], Adrienne Raglin[2], Eric Heilman[2],
and Timothy Hanratty[2]

[1] U.S. Army Research Laboratory West, Playa Vista, CA 90094, USA
mark.s.dennison.civ@mail.mil
[2] U.S. Army Research Laboratory, Adelphi, MD 20783, USA
{mark.r.mittrick.civ, john.t.richardson7.civ,
theron.t.trout.ctr, adrienne.raglin2.civ,
timothy.p.hanratty.civ}@mail.mil

Abstract. The benefits and limitations of immersive technologies in military decision-making are not well understood. Here, we describe the framework of an experiment which seeks to empirically determine the effects of immersive and non-immersive technology on decision-making. In this experiment, users are shown tactical spatial information about a building layout and told they must decide which of three pre-determined breach points is optimal for maximizing mission success and minimizing risk to the ground team. To ensure observable effects are related to immersion and not simply perception of depth, we deploy a between-subjects design with three viewing conditions: data shown in 2D on a desktop display, data shown in 3D on desktop display, and data shown in 3D in a head-mounted display (HMD). Dependent variables include decision accuracy, time to task completion, decision confidence, and score on the System Usability Scale. In the VR version of the experiment, full telemetry is captured to track when and for how long users interacted with specific information in the scenario environment. Pilot results suggest that tracking these metrics will allow for intricate comparison of decision-making behaviors between display types.

Keywords: Virtual reality · Mixed reality · Decision-making · User experience

1 Introduction

Future combat environments will require military analysts to utilize a common operating environment that integrates information from a complex network of interacting intelligent systems (human, robotic, and networked sensors) to support the decision-making of commanders. The integration of information originating from heterogeneous sensors and intelligence sources often requires the use of discrete systems, such as multiple computers, laptops, tablets, or even physical objects on a map. Such setups typically require significant physical and cyber resources to bring information into a unified space where collaborative decision-making can occur more effectively. Immobile command and control structures are especially at significant risk from hostile attacks.

Discovering new ways of accessing, consuming, and interacting with information from the battlefield is necessary to expedite military operational effectiveness. Mixed reality (MR) is one technology that may enable this. Two promising applications of MR may be data ingestion, visualization and collaborative analysis at standoff from the tactical edge and the ability for users to modify their perception of reality depending on their personal preferences, areas of expertise, or other mission requirements. Despite this, there is currently a limited understanding of how, when, and where using mixed reality provides explicit benefit over more traditional display systems. For example, does viewing spatial information in an immersive HMD result in faster or more accurate decisions? More so, the existing literature is dominated by studies that only compare non-immersive and immersive systems at a surface level, where reported metrics are often qualitative.

In this paper, we discuss the design of a decision-making experiment which seeks to measure how users interact with tactical information when it is displayed in 2D, 3D, and in an immersive virtual environment. We describe the technical aspects of AURORA-VR, a research system developed at ARL which enables precise tracking and analysis of military tasks across mixed reality. Finally, results from pilot data from the development team are discussed.

Existing literature outside of the military domain has shown that virtual and augmented reality may provide benefit to decision-making and understanding in certain scenarios. For example, Moran [1] showed enhanced situational awareness, cognition, data pattern detection, and that visual analytics were more efficient in an immersive system compared to a traditional 2D system. McIntire and colleagues [3, 4] showed that for most subjects, using a stereoscopic display increased performance by roughly sixty percent. However, this survey covered a broad range of tasks and devices, which makes true comparison difficult. A study by Dan and Reiner [4] measured differences in performance on simple tasks, such as a paper folding, after training occurred in 2D as viewed on a desktop monitor versus in 3D when viewed in augmented reality. The authors reported that subjects in the augmented training condition demonstrated significantly less cognitive load when learning the folds, as measured by power spectral changes in electroencephalographic (EEG) recordings. This indicated that information transfer was significantly easier when the data was viewed in an augmented environment. This decreased cognitive load may be related to the suggestion that humans are "biologically optimized" to perceive and learn in 3D [5].

Research by Donalek et al. [5] reported that in a waypoint drawing task, subjects who viewed the environment in an Oculus Rift HMD performed with less distance and angle errors than those who viewed the environment on a 2D desktop monitor. Moran et al. [1] created an immersive virtual environment where Twitter data was overlaid atop real geography to improve the experience for analysts. The authors claimed that this augmented environment enhanced situational awareness, cognition, and that pattern and visual analytics were more efficient than on traditional 2D displays. However, metrics enabling precise comparison across display types are reported with limited detail.

1.1   Immersive Environments and Decision-Making

Decision-making tasks, such as battlefield intelligence analysis or optimal route planning for tactical operations, are two areas where virtual or augmented reality may enable Warfighters to better perform their roles. For example, individuals who manipulated visualized social media data in a fully-immersive and motion-tracked virtual environment reported that they learned more about the data than when it was viewed in a traditional setting [6]. Often, though, subjective metrics may not be externally valid. It has also been shown that overlaying virtual information may dramatically improve performance and task engagement [7, 8]. MR can also be used to overlay the virtual hands from an expert user to a novice, guiding them through a complex scenario remotely [9]. In a military context, an analyst in VR at a forward operating base could assist a solider at the tactical edge by highlighting known enemy positions and display that information on a tablet or through an augmented reality HUD integrated with their helmet.

Prior work has shown that augmenting data displays may provide some benefit to understanding and decision-making. In a recent study by [10], the authors compared the effectiveness of immersive AR (HoloLens) and tablet VR to traditional desktop use by measuring completion time and error in a point cloud estimation tasks. The immersive AR environment was reported as best for tasks involving spatial perception and interactions with a high degree of freedom, but subjects were generally faster on the desktop where interactivity was already familiar to them. As suggested by Bellgardt et al. [11], it is critical that mixed reality systems are designed in such a way that integration with existing workflows is seamless.

Finally, from a military perspective, the time and resource cost associated with travel across the battlefield, construction and deployment of command and analyst tents, and upkeep to meet constantly changing mission demands make fluid collaboration across the battlefield challenging. Virtualizing some elements of mission command and intelligence operations may help to reduce this difficulty. For example, Fairchild and colleagues built a VR telepresence system which allowed scientists in Germany and the U.K. to collaborate remotely and in real-time on data from a Mars mission [12]. The VR system tracked gaze, facial expressions, and user positions to maximize the scientists' nonverbal communication over the wire. GraphiteVR, a project by Gardner and Sheaffer [13], allows multiple remote users to visualize high-dimensionality social media data and manipulate it in a shared virtual space.

2   Experiment Design

2.1   Study Objectives

The goal of this work is to determine how different display technologies and data visualization techniques affect decision-making in a tactical scenario. Although prior research [2, 3] has demonstrated some evidence that visualization of spatial information is easier when viewed through a stereoscopic-display, there is limited work showing such enhancement in tactical decision-making scenarios. For this study, the scenario will involve the perception and analysis of both spatial and non-spatial information

from a fictional military operation to breach a hostile building and secure a specific item. Information pertinent to execution of this operation will be presented either in 2D slices on a desktop monitor, as a full 3D model on a desktop monitor, or as a 3D model presented in virtual reality. These conditions will enable us to empirically determine how differences in display type and data visualization method affect the speed and confidence of decision-making, and the usability and comfort of the display medium.

The inclusion of the 3D model condition is essential because it allows us to determine if any changes in behavioral performance in the VR condition are truly because of immersive qualities produced by stereoscopy, and not simply because of the presence of depth information in the scene.

2.2   Questionnaire

The System Usability Scale (SUS) [14] a reliable, low-cost usability scale that can be used for global assessments of systems usability. It is a is a simple, ten-item scale giving a global view of subjective assessments of usability. The SUS is generally used after the respondent has had an opportunity to use the system being evaluated, but before any debriefing or discussion takes place. Respondents should be asked to record their immediate response to each item, rather than thinking about items for a long time.

2.3   Task and Stimuli

This experiment uses a between-subjects design, with three separate interfaces as the independent variable. In Condition 1 (2D Desktop), users will view task information as 2D images on a desktop monitor. In Condition 2 (3D Desktop), users will view task information as a 3D model rendered on a desktop monitor. In Condition 3, users will view task information as a 3D model displayed in virtual environment inside an Oculus Rift HMD. Screenshots of the task environment are shown in Fig. 1.

The task information displayed across all three conditions is spatial data from a fictional two-story building with various rooms, entrances, and objects. On the inside of the building, one room contains a bomb which the forward operators must safely navigate to for extraction. A visual legend is shown in the display in each condition which indicates whether or not a particular entrance of the building is either open (green marking), breachable by the forward operating team (yellow marking), or not breachable by the forward operating team (red marking). Across the outside of the building, three possible initial entry points are highlighted and given a text label. The user's task is to explore all of the information about the building and choose which initial breach point will maximize successful navigation to the target room, while minimizing risk to the forward operating team.

Users are given as much time as they would like to consider all options and explore the provided information. When ready, they must select which of the three initial breach points they think is most optimal be selecting its label on the experiment interface. Once selected, they will be prompted to select how confident they are in that decision, using a −2 to +2 Likert scale. Finally, they are told whether or not their selection was indeed the most optimal choice.
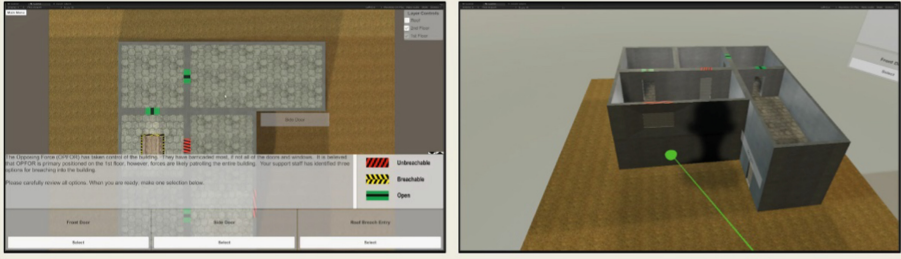
Fig. 1. Example scenes showing the scenario building from the 2D Desktop (left) and 3D Desktop/VR (right) experimental conditions. (Color figure online)

2.4   AURORA-VR

The virtual environment in all three experimental conditions is displayed through AURORA-VR, a research and development sandbox created by the Army Research Laboratory and Stormfish Scientific as the VR component of the AURORA (Accelerated User Reasoning: Operations, Research, and Analysis) project in the Battlefield Information Processing Branch. Real-time user interaction, position, and decision data are captured through the system and sent to Elasticsearch for online visualization in Kibana or offline analysis in MATLAB.

2.5   Data Analysis

Our hypotheses are motivated by prior research that has suggested it may be easier for humans to understand spatial information when observing a display that includes depth information (e.g. three dimensions or stereoscopy). Critically, we also expect that interacting with this spatial information through an immersive HMD will provide users with a better understanding of the layout and properties of the building. Therefore, we anticipate that actual and perceived task performance will be worse in the 2D Desktop condition, better in the 3D Desktop condition, and best in the VR condition.

To test this hypothesis, we will compare metrics recorded across experimental conditions. These data include the time spent on each of three introduction screens, the time spent exploring the virtual environment until a decision is made, the time spent deciding the level of decision confidence, the number of times the user switched which floor to display in the building, and how long the user spent observing each floor. Condition-specific metrics will also be analyzed to observe differences between subjects within that condition. In the 2D Desktop condition, metrics include the total variability of camera zooming and panning and the variability of these actions with respect to which floor was displayed. In the 3D Desktop condition, metrics include the total variability and magnitude of camera zooming and rotational panning (yaw, pitch) and the variability and magnitudes of these actions with respect to which floor was displayed. For the VR condition, position information about how the user moved around the virtual environment (forward-backward, right-left, up-down) will be analyzed over the duration of the experiment and with respect to which floor was viewed.

3 Pilot Results

Nine members of the research and development team pilot tested the environment. Data were acquired through custom software connecting Unity to Elasticsearch and offline analyses were carried out in MATLAB v2017b. Time dependent information were sampled at a rate of 5 Hz. Because a limited number of pilot data were acquired for the purpose of feasibility testing, statistical comparisons are not reported. All error bars represent one standard error (se) of the mean.

3.1 Questionnaires Data

The System Usability Scale is a ten-item scale giving a global view of subjective assessments of usability. Scores range from 0 to 100. For the 2D Display condition, the average score was 41.7 (se = 4.41). For the 3D Display condition, the average score was 48.3 (se = 0.83). For the VR Display condition, the average score was 48.8 (se = 1.25).

3.2 Behavioral Data

Across Condition Metrics
We first analyzed data to report metrics that were common to all three experimental conditions (2D, 3D, VR). Figure 2 shows that in the 2D and 3D conditions, one user chose the Side Entry point, and two users chose the Roof Entry point. For the VR condition, one user chose the Floor Entry point and another chose the Roof Entry point. As the optimal selection was the Roof Entry, performance was best for 2D and 3D displays.
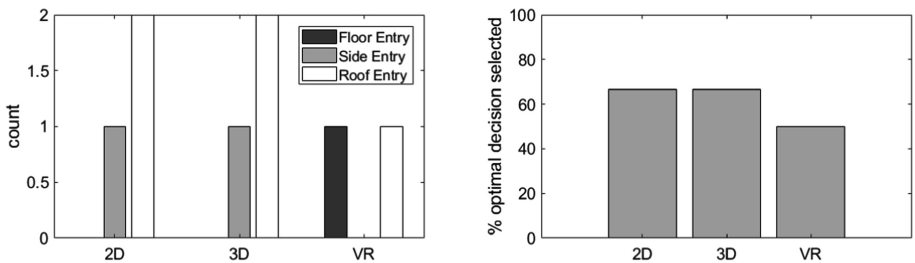

Fig. 2. Breach decision selection count (left) and percent optimal selection (right), with respect to display type.

The average time users spent reading over the introduction screens is shown in Fig. 3. Generally, the data trend to show that less time was spent on subsequent screens for all display types.
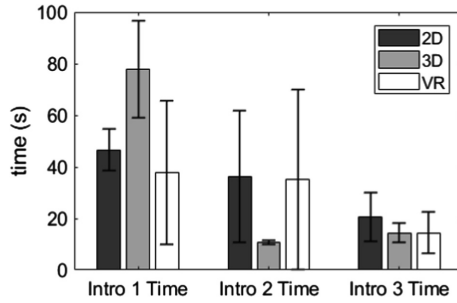
Fig. 3. Average time in seconds spent observing each intro screen, with respect to display type.

The average time users spent interacting with the breach environment information is shown below in Fig. 4 for each display type. The data trend to show that users spent more time examining information in the 3D than 2D condition, and the most time in the VR condition.
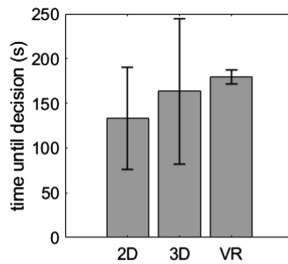

Fig. 4. Average time interacting with the VE until a decision was made, with respect to display type.

The left subplot of Fig. 5 shows that, on average, users trended to spend more time deciding how confident they were in their decision after viewing the breach environment in VR. Generally, the data trend to show that less time was spent on subsequent screens for all display types. The right sub-plot shows that generally users reported the same level of confidence across display types, with the highest variability in the 3D condition.
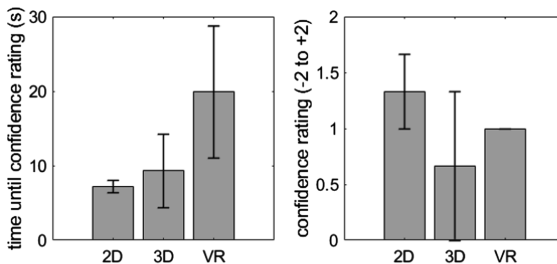

Fig. 5. Average time (left) and rating (right) for confidence input, with respect to display type.

On average, users trended to spend the most time looking at information on the first (ground) floor of the building, where the objective room was located (see Fig. 6). Generally, the data trend to show that less time was spent on subsequent screens for all display types. The right sub-plot of Fig. 6 shows that generally users reported the same level of confidence across display types, with the highest variability in the 3D condition.
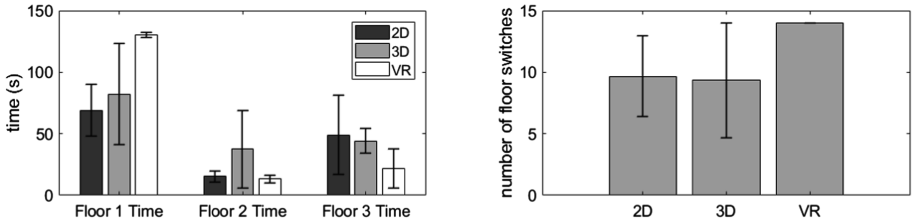

Fig. 6. Average time observing each floor (left) and total number of floor switches (right), with respect to display type.

2D Specific Metrics

In the 2D display condition, users zoomed in and out of the environment only when the 3rd Floor was being viewed (see Fig. 7). More left-right and up-down panning activity was observed when the 1st and 2nd floors were being viewed, as shown in the Fig. 7 center and right subplots.
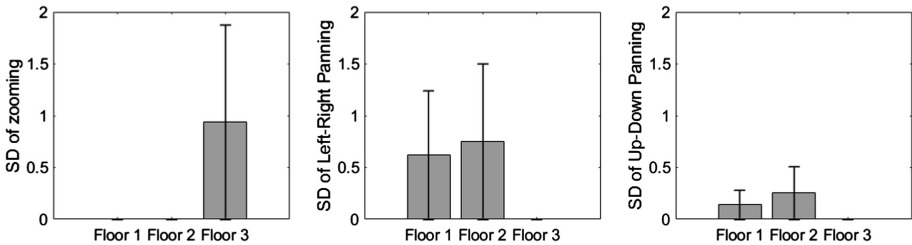

Fig. 7. Average variability in zooming activity (left), left-right camera panning (center), and up-down (right) camera panning activity, with respect to highest visible floor.

3D Specific Metrics

Users on average zoomed in and out of the 3D environment only while viewing the 1st and 2nd Floors of the building (see Fig. 8A). The data trend to show that users rotated the camera along the yaw direction more while viewing the 3rd Floor, as shown in Fig. 8B and D, whereas users pitched the camera up and down similarly while viewing all floors, as shown in Fig. 8C and E.
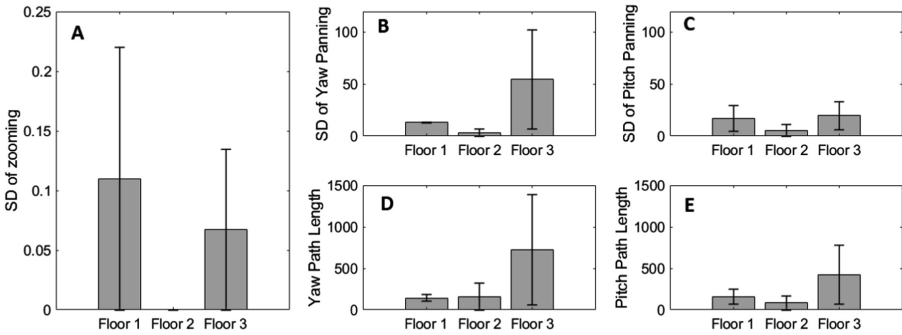
Fig. 8. Variability of zooming activity (A), panning in the yaw direction (B), and panning in the pitch direction (C), total path length in the yaw direction (D) and pitch direction (E).

VR Specific Metrics

Data from the VR display condition where analyzed with respect to mo variability in the cardinal directions. Figure 9A and D show that head movement in the left-right direction was greatest when users viewed the 1st Floor of the building. Figure 9B and E shows that users trended to have greater forward-back head movement variability and much greater distance traveled when viewing the 1st Floor. Figure 9C shows that the head movement variability in the up-down direction was similar when viewing all floors, but Fig. 9F shows a trend that users moved their head up and down more when viewing the 1st Floor of the building.
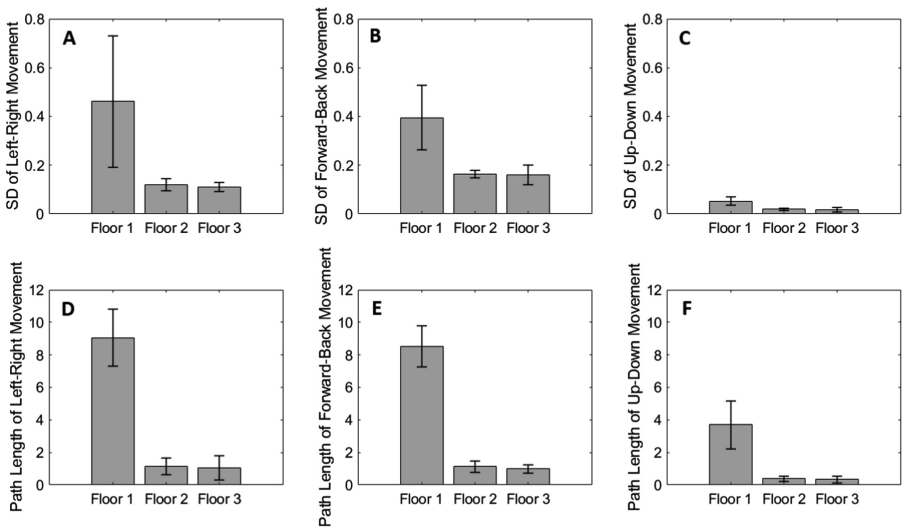

Fig. 9. Variability of head left-right movement (A), forward-backward movement (B), up-down movement (C), total path length of movement in the left-right (D), forward-backward (E), and up-down (F) directions.

4   Conclusions and Future Work

In this paper, we presented the design and pilot results of a study to examine qualitative and quantitative differences associated with viewing spatial tactical information rendered as 2D information on a flat-screen display, as 3D information rendered on a flat-screen display, and as 3D information rendered in an immersive virtual environment viewed through an Oculus Rift HMD. As a shift away from other work in the field that has primarily focused on high-level and broad comparisons between display types and tasks, our goal was to focus on extracting interaction and behavioral metrics that were common across all three experimental conditions and those that were unique to each. Critically, this allows for a much finer grain comparison of how users chose to interact and thus perceive information from the environment.

The inclusion of the 3D Display condition, namely where tactical information was rendered as a 3D model on a flat-screen display, is necessary to tease apart behavioral effects related to depth information from those potentially arising from the user being immersed in the virtual environment. This is important because although viewing the task environment in an HMD may be more enjoyable or captivating to the user, as has been previously reported in the literature (citation), VR may not provide any tangible benefit over simply viewing the same depth information on a traditional desktop display.

Individual differences in prior exposure to virtual reality technology and performance on spatial information are important to consider when doing any comparison between non-immersive and immersive systems. For this pilot study, all users had prior exposure to VR and thus may not represent the average population. The full version of this study will include the Visualization of Viewpoints and Spatial Orientation [15] pen-and-paper surveys to assess differences in 3D orientation skills of users before they interact with the VE. It is critical to ensure enough data are collected such that performance on these tests is not significantly different across condition groups.

We first analyzed the behavioral data with respect to metrics common across conditions. Users spent about the same time viewing the introduction screens for each display condition, with less time spent on subsequent screens (see Fig. 3). Figure 2 shows that generally most users chose the correct answer, which was entering the building from the roof. This answer was optimal as the floor entry was in line of sight of the enemy, and the side entry ended in a door that was not breachable by friendly forces, as denoted by symbology present in user interface for all display conditions. Users in each condition also took about the same amount of time to come to a decision about which breach point they thought was optimal, although the data trend to suggest users may have spent more time interacting with the environment when it was viewed in VR. Additionally, users showed the same trend when deliberating on the confidence rating associated with that choice (see Fig. 5), with data trending to show more time taken in the VR condition. The magnitude of confidence ratings ranged from neutral, to very confident, with the greatest variation from users in the 3D display condition. Lastly, the System Usability Scale results demonstrated that on average, users felt about the same for each of the display conditions, with perhaps slightly less favorable

scores for the 2D Display condition. Users likely had to perform more actions to view the same information, this could have led to feelings of frustration which might be the reason for the lower score.

The tactical information in this experiment's design was presented across three different floors, so interaction behavior was analyzed with respect to which floor the user was currently viewing. Figure 6 shows that across display conditions, most users spent most of their time observing the bottom floor of the building. This may have been because two of the breaching options were on this floor. The data also show that typically, users switched between floors about ten or so times during a run.

The 2D Display condition allowed the user to navigate the presented environment by clicking and dragging the camera left and right or up and down. It also allowed the user to zoom in and out of the building from a top-down perspective. Users in this condition showed the greatest zooming activity when floor three was presented (see Fig. 7). This is likely because floor three was visible at the start of the run, and users zoomed in and out to a view they preferred and then kept the camera at this distance for the rest of the run. Panning left and right appears to have occurred more than panning up and down, however more data are needed in a full study to confirm this statistically.

In the 3D Display condition, users could navigate the environment by zooming in and out of the 3D building model, or rotating the camera along the yaw and pitch axes. Here, we saw the most zooming activity when floors one and three were visible (see Fig. 8). Additionally, users may have moved the camera along the yaw axis more than they pitched the camera up and down. This makes sense as the camera was positioned at approximately 45° from the ground which may have already been an acceptable position for most users to view the 3D model.

For the VR Display condition, we tracked how users moved their head in 3D space around the virtual environment. Figure 9 shows that users primarily showed the most movement along the left-right and forward-backward directions. Minimal movement was shown in the up-down direction, which was likely caused by head bob from walking, or perhaps some tilting to look down into the building model. This matched our expectations as to completely view the environment, a user would have to walk around the horizontal plane.

Our results suggest the current experimental design may be useful for evaluating immersive and non-immersive interfaces for tactical decision-making. For the full experimental study, it is clear that a large number of users is necessary in each display condition to ensure that individual differences in VR experience, aptitude in spatial data navigation, and underlying bias are accounted for. We also plan to record four minutes of resting EEG collection, with two minutes of eyes-open and eyes-closed respectively. These EEG signals will be examined for time-frequency changes in power at individual and coherence across groups of electrodes. Prior research examining resting cortical activity has been predicative of differences in cognitive performance [16], visuo-motor learning [17], and related inversely to the default-mode network [18].

Finally, the current design allows for only one decision selection to be made at the end of the interaction. The full iteration of this study may benefit from users making multiple decisions in larger-scale tactical operation across multiple buildings. For example, the user could make a breach choice, receiving feedback in real time, and then continue making decisions until the final objective is reached. Here, their "score" could

be tracked based on optimality of decision selection, and then tallied at the end of the session. This may allow for a richer outcome measure with which to correlate the quantitative and qualitative metrics described in this paper.

References

1. Moran, A., Gadepally, V., Hubbell, M., Kepner, J.: Improving big data visual analytics with interactive virtual reality. In: 2015 IEEE High Performance Extreme Computing Conference (HPEC), pp. 1–6 (2015)
2. McIntire, J.P., Havig, P.R., Geiselman, E.E.: What is 3D good for? A review of human performance on stereoscopic 3D displays, p. 83830X (2012)
3. McIntire, J.P., Liggett, K.K.: The (possible) utility of stereoscopic 3D displays for information visualization: the good, the bad, and the ugly. In: 2014 IEEE VIS International Workshop on 3DVis (2014)
4. Dan, A., Reiner, M.: EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. Int. J. Psychophysiol. **122**, 75–84 (2016)
5. Donalek, C., et al.: Immersive and collaborative data visualization using virtual reality platforms. In: IEEE International Conference on Big Data, pp. 609–614, October 2014
6. Royston, S., DeFanti, C., Perlin, K.: A collaborative untethered virtual reality environment for interactive social network visualization, April 2016
7. Frank, J.A., Kapila, V.: Mixed-reality learning environments: integrating mobile interfaces with laboratory test-beds. Comput. Educ. **110**, 88–104 (2017)
8. Dascalu, M.-I., Shudayfat, E.A.: Mixed reality to support new learning paradigms. In: 2014 18th International Conference on System Theory, Control and Computing, pp. 692–697, October 2014
9. Huang, W., Alem, L., Tecchia, F., Duh, H.B.L.: Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration. J. Multimodal User Interfaces **12**, 77–89 (2017)
10. Bach, B., Sicat, R., Beyer, J., Cordeil, M., Pfister, H.: The hologram in my hand: how effective is interactive exploration of 3D visualizations in immersive tangible augmented reality? IEEE Trans. Vis. Comput. Graph. **24**(1), 457–467 (2017)
11. Bellgardt, M., Pick, S., Zielasko, D., Vierjahn, T., Weyers, B., Kuhlen, T.W.: Utilizing immersive virtual reality in everyday work. In: 2017 IEEE 3rd Workshop on Everyday Virtual Reality, pp. 1–4 (2017)
12. Fairchild, A.J., Campion, S.P., Garcia, A.S., Wolff, R., Fernando, T., Roberts, D.J.: A mixed reality telepresence system for collaborative space operation. IEEE Trans. Circuits Syst. Video Technol. **27**(4), 814–827 (2017)
13. Gardner, M.R., Sheaffer, W.W.: Systems to support co-creative collaboration in mixed-reality environments. In: Liu, D., Dede, C., Huang, R., Richards, J. (eds.) Virtual, Augmented, and Mixed Realities in Education. SCI, pp. 157–178. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-5490-7_9
14. Brooke, J.: SUS-a quick and dirty usability scale. Usability Eval. Ind. **189**, 4–7 (1996)
15. Kyritsis, M., Gulliver, S.R.: Gilford Zimmerman orientation survey: a validation. In: Proceedings of the 7th International Conference on Information, Communications and Signal Processing, ICICS 2009 (2009)
16. Manuel, A.L., Guggisberg, A.G., Thézé, R., Turri, F., Schnider, A.: Resting-state connectivity predicts visuo-motor skill learning. Neuroimage **176**, 446–453 (2018)

17. Wu, J., Srinivasan, R., Kaur, A., Cramer, S.C.: Resting-state cortical connectivity predicts motor skill acquisition. Neuroimage **91**, 84–90 (2014)
18. Scheeringa, R., Bastiaansen, M.C.M., Petersson, K.M., Oostenveld, R., Norris, D.G., Hagoort, P.: Frontal theta EEG activity correlates negatively with the default mode network in resting state. Int. J. Psychophysiol. **67**(3), 242–251 (2008)