Assessing the quantitative and qualitative effects of using mixed reality for operational decision making

Mark Dennison[1], Jerald Thomas[2], Theron Trout[1,3] and Evan Suma Rosenberg[2]

[1]U.S. Army Research Laboratory West, Playa Vista, CA
[2]University of Minnesota, Minneapolis, MN
[3]Stormfish Scientific, Chevy Chase, MD

Abstract

The emergence of next generation VR and AR devices like the Oculus Rift and Microsoft HoloLens has increased interest in using mixed reality (MR) for simulated training, enhancing command and control, and augmenting operator effectiveness at the tactical edge. It is thought that virtualizing mission relevant battlefield data, such as satellite imagery or body-worn sensor information, will allow commanders and analysts to retrieve, collaborate, and make decisions about such information more effectively than traditional methods, which may have cognitive and spatial constraints. However, there is currently little evidence in the scientific literature that using modern MR equipment provides any qualitative benefits or quantitative benefits, such as increased task engagement or improved decision accuracy. There are also no validated metrics in the field for comparing across display devices and tasks. In this paper, we surveyed potential metrics for assessing the usefulness of MR technologies, discuss how these data might be acquired in experimental and tactical scenarios, and pose issues in multi-user communication and collaboration. We also introduce the Mixed Reality Tactical Analysis Kit (MRTAK), which functions as an experimental platform to perform these assessments during collaborative mission planning and execution.

1 Background

The modern battlefield and Army operational environment is becoming more varied and dynamic, with a greater reliance on the integration of information from intelligent things/devices, agents, and systems. Information overload caused by multitasking and mission execution at standoff remain significant challenges in C3I scenarios. The integration of information for decision making and other mission command tasks is often still done using discrete and disparate systems (physical objects, computers, paper documents, etc.) that require a significant amount of resources and effort to bring into a unified space. Human interactions require shared cognitive models where interaction with systems must support and maintain this shared representation, stored information must persist the representation at a fundamental data-level, and the underlying network must allow these data and information to flow without hindrance between human collaborators and non-human agents (1; 2).

The emergence of Mixed Reality (MR) technologies has provided the potential for new methods for the Warfighter to access, consume, and interact with battlefield information. MR may serve as a unified platform for data ingestion, analysis, collaboration, and execution and also has the benefit of being customized based on the mission needs and requirements of each operator (see Figure 1). MR lies in the Reality-Virtuality Continuum between the physical and the digital world (3). Augmented Reality (AR) and Virtual Reality (VR) exist at the extremes of the MR spectrum, as shown in Figure 2. Where AR superimposes generated content over the real world, VR occludes the real world entirely to present something entirely fabricated. Each of these immersive technologies has benefits and drawbacks, but there has been limited research exploring what these are beyond speculation. Importantly, MR ensures a visual connection to the physical world, while utilizing elements that may be superimposed on reality or occluding it completely with a purely virtual rendering of information and objects. Thus, MR may serve as a medium to integrate data from sensors monitoring the real world, with the ability to perceive and reason on this information without many of the spatial and physical constraints of currently used C3I systems.

The recent increase in the ease of access to modern high-fidelity head-mounted displays has caused a resurgence in interest for using immersive technolo-

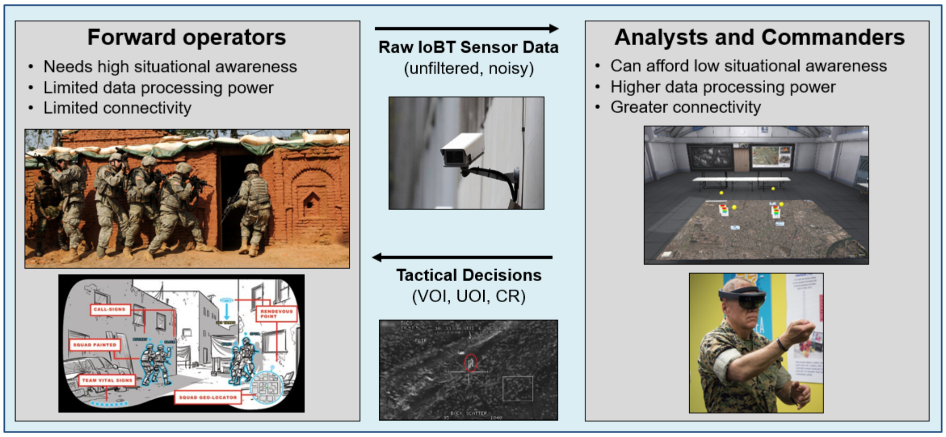

Figure 1: Information flow between forward operators and analysts using immersive display devices.

gies in the Private and public sectors. Consequently, the "cool factor" associated with VR and AR technologies has become a common reason for their adoption, with little empirical data backing this up. Similar issues have been found in adding gamified elements to training, which can be completely ineffective or simply less effective than less engaging traditional methods. With respect to command and control, prior work has shown that novel interfaces that support decision making have traditionally been challenging for users to understand and interpret (4). Additionally, there is little evidence in the literature of the quantitative benefits of using immersive technologies in operational decision-making, nor are there a set of tasks and validated measures for assessing optimality. Here, we review the limited current literature that have attempted to address this issue with respect to evaluation and communication in MR, and discuss how the MRTAK project seeks to build upon this work as a sandbox for immersive C3I research.

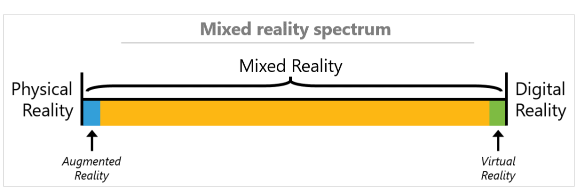

Figure 2: Mixed reality technology spectrum. Adapted from Milgram (3)

1.1 Evaluation of Immersive Technologies

Recently, this area of research has been referred to as "Immersive Analytics" (5). Chandler and colleagues suggest five major topic areas: 1) What paradigms are enabled by immersive technologies and how do we evaluate them over other traditional mediums and each other?, 2) Do these technologies provide a more holistic way of looking at data that contains 3D spatial and abstract information?, 3) What are the best interface "tricks" and affordances that change a user's perspective from an allocentric to egocentric view of the data?, 4) Do these technologies invalidate the literature on 2D data interaction?, and 5) What is the typical work-flow for examining data across domains and how do we develop generic platforms to support immersive analytics? Although each of these questions is important, here we focus on items one and two, which pose the more general question of how should we evaluate the effectiveness of immersive technologies and what data is necessary to perform this assessment.

Uses for MR Technology

One area where immersive technologies have been used extensively is for simulation and training on real-world tasks. For example, a study by Donalek and colleagues (6) reported that in a way-point 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 desk-top monitor. Moran and colleagues (7) created an immersive virtual environment where Twitter data was overlaid atop real geography to improve the experience for analysts. The authors claimed that this MR environment enhanced situational awareness, cognition, and that pattern and visual analytics were more efficient than on traditional 2D displays. A study by Dan & Reiner (8) measured performance differences among subjects who had to complete a paper folding task after viewing information on a 2D desktop monitor or through an augmented environment. Subjects showed a higher cognitive load index when learning in 2D vs 3D, as measured by the ratio of frontal theta power over parietal alpha power. This indicated that information transfer was significantly easier when the data was viewed in an MR environment. Other work has shown that the perception of one's virtual body and hands is also a critical feature when performing cognitively demanding tasks, such as memorization, when done in a virtual environment (9). This decreased cognitive load may be related to the fact that humans are "biologically optimized" to perceive in 3D (6). McIntire and colleagues (10; 11) reported that use of a 3D stereoscopic display increased task performance by roughly 60% on average. Recently, it was reported immersive AR was found to be better when manipulating data that required spatial perception and interactions with a high degree of freedom (such as tangible user interfaces), but users were generally faster on the desktop if the task was familiar (12). Generally, though, these studies provide limited empirical evidence for which immersive mediums (VR, AR, MR) are best for improving user decision making across content domains, and many do not use similar or easily comparable metrics.

Evaluation Metrics

The issue of evaluation metrics is of critical importance. In a seminal study by Borsci and colleagues (13), the authors conducted a review of all existing studies that performed an assessment between an immersive technology, such as AR and VR, and a traditional, such as a desktop monitor, or between different immersive devices. They list nine evaluation criteria used previously in the field: 1) Cognitive skills 2) visuo-spatial abilities 3) Levels of trust/acceptance of VR/MR tools and motivation in use, 4) Participants attitude, 5) Previous experience, 6) motion sickness, 7) physiological Reactions (attention shift, cognitive load, stress) 8)Level of presence and engagement, and 9) Technical aspects and tools. Studies also reported using pre-training assessments and demographic measures, task performance assessments, varying experimental conditions, and assessing post-training criteria, typically through questionnaires. Across the literature surveyed, the degree of overlap varied significantly and the authors felt that researchers focused solely on assessing time and task errors that they failed to measure for critical effects such as motion sickness, decay and recall of skills after immersion, the level of trust or acceptance of the device, and the users' prior attitude, skills, and experience with similar technologies.

Some studies have shown the importance of the issues brought up by Borsci. For example, it was found that reported VR system usability was correlated strongly with a user's level of trust in that system (14; 15). These criteria were assessed through validated metrics such as the System Usability Scale (16) and Trust in Technology questionnaires (17). Neurophysiological surveys have also been shown to correlate strongly with performance in immersive environments. Davison (18) showed that Performance on the Trail-Making Task A (TMT-A) (19), a task considered to assess motor speed, was found to be significantly related to other measures which also assessed speed, such as the time taken to complete parking simulator levels and the time taken to place virtual objects around a room. Measures of executive function, such as TMT B performance, was found to be significantly related to performance on both of these spatial location tasks. Dennison and colleagues found that motion sickness caused by immersion in a virtual environment (VE) greatly impacted the duration to which participants elected to remain in the VE and complete decision making tasks (20; 21; 22). Collectively, these studies demonstrate the need to assess not only the psycho-physiological profiles of intended users — as measured through questionnaires and pretask assessments — but also the potential benefit of monitoring these states during real time use of immersive technologies, when possible.

1.2 Evaluation of Multi-User Interaction

The majority of scenarios in which MR can be applied have multiple users. These users can be operating with the same or different immersive technology, can be colocated or connected remotely over a network, and may have access to the same information or only pieces of it (information symmetry). A task incorporating multiple participants will be affected, to some degree, by the communication behaviors among participants. Consequently, researchers must consider these dynamics when determining metrics that assess the effectiveness of an immersive technology for an entire scenario or for component tasks. We examined literature from the field of computer mediated communication (CMC) because interpersonal communication, as mediated by virtual environments, is still a new concept within the VR, AR, and MR fields.

It is first important to consider the research regarding the efficacy of various CMC modalities. How does an observing analyst best share important information to a squad at the tactical edge? When bandwidth on a tactical network is limited, or the likelihood of unwanted third-party observation, interception, or tampering of communications is high, what is the level of fidelity required to effectively execute command and control to complete the mission? These questions, which are not tied to any specific technology, are critical factors when assessing how an immersive medium might help or hinder individuals in a decision-making scenario.

Currently, there are many theories and models regarding CMC (23). Despite this, it is difficult to find metrics or evaluation frameworks that provide empirical evidence on the efficacy of these CMC modalities. To the best of our knowledge, existing studies examine only specific CMC modalities and compare them strictly against face-to-face communication methods. Across these studies, task performance has been used as the key metric for determining CMC efficacy (24; 25). It is also important to note that, generally, the only independent variable in these studies is the CMC modality. Furthermore, obtaining empirical evidence about task performance becomes increasingly difficult when multiple people are performing the task together (23). Factors such as the participants' relationship (26), amount of trust in their teammate (25), and the ability of each individual to perform their designated portion of the task (27) can individually and collectively be extremely hard to control. These factors must be considered both when designing experimental measures and when evaluating immersive systems. If a participant does not trust their counterpart as a valid source of information or they do not believe that they can competently complete the task, it will likely have a significant effect on task performance.

After examining the CMC literature, we have a compiled a list of suggestions for conducting MR user studies with multiple local or remote users. First, it is important to create a rigorous study design that controls for confounding factors. If the experiment is using task performance to measure both technology efficacy and communication efficacy, make sure to include an appropriate number of permutations within the study design to control for order effects. Second, consider the participants' relationship as a factor. A group of friends and a group of strangers will behave and communicate differently, effecting the way in which they execute the task and how they value different performance outcomes. Third, include appropriate questionnaires to tap into specific measures of interest, rather than asking overly generalizable or vague questions, such as "How did you like the communication system?". A questionnaire determining a participant's level of trust in another participant can be adapted from Rotter (28) and a questionnaire determining participants' level of rapport with one another can be adapted from Puccinelli and Tickle-Degnen (29), as examples.

2 The Mixed Reality Tactical Assessment Kit

The U.S Army Research Lab and industry partner Stormfish Scientific have built a collaborative mixed reality infrastructure, called the Mixed Reality Tactical Assessment Kit (MRTAK). The goal of MRTAK is to allow researchers to perform controlled studies evaluating how immersive technologies compare against traditional systems in single and multi-user operational decision making tasks. MRTAK also will allow researchers to test and evaluate different network and data management control frameworks. Currently, we are collaborating with academic partners at the University of Minnesota, USC Institute for Creative Technologies, and University of California Irvine.

The DICE Network

One of the key components of MRTAK is the Defense Integrated Collaborative Environment (DICE) network (30), developed at ARL with Stormfish Scientific. DICE hosts a confidential private network where collaborative MR services are hosted, and local and remote clients can connect to these services through a secure VPN. A diagram of DICE is shown in Figure 3. This network was designed to meet rigorous Department of Defense and Army Standards and uses policy based security so that access can be controlled at multiple-levels of granularity. Thus, DICE allows for controlled experimentation of how normal and degraded network conditions, where bandwidth may be extremely limited, affect different aspects of multi-user collaboration. For example, consider a situation where one user is providing navigation to another using complex spatial markers rendered over an AR display. If the network were to be strained to the point that image data could no longer be sent from the edge, researchers could test how teams could communicate that critical information over alternative channels until bandwidth was restored.

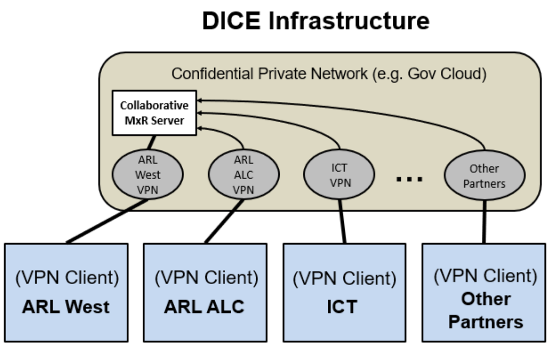

Figure 3: Overview of the DICE Network structure.

Sensor Connectivity and Machine Learning

MRTAK is also equipped with a fully synchronized data source management system. This system allows for ingestion of sensor data from local users or from external devices such as Internet of Battlefield Things (IoBT) sensors, and is integrated through the DICE network. With this system, researchers are able record data from all aspects of the collaborative decision making process in real time. Moreover, these data can even be viewed from within the immersive environment as a form of training or feedback. MRTAK allows for key data to be recorded at each step of the decision making process, and allows for experimenters to freely choose which display devices are used and whether or not the participants are local or remote. Similarly, communication among users via any modality (voice, symbology, tracks) can be recorded and used for later analysis. The underlying data framework also makes it easy to run machine learning applications on data generated from participants in the environment or on incoming information from external programs or sensors. Thus, models for value of information (31), information availability (32), or uncertainty (33) can be integrated and tested with respect to which display platform or tactical decision they are optimal for.

3 Conclusion

In conclusion, the current literature suggests that immersive technologies may provide a means of improving certain aspects of operational decision-making. Future work should aim to report more objective and precise measurements of task outcomes when comparing different immersive interfaces and, when possible, include comprehensive assessments of a user's background and experience with similar systems. Decision making tasks should be broken down into key processing steps and performance increases or decreases should be discussed with respect to these elements and with respect to the overall mission. Physiological sensors can also be used to track state information that may not be readily accessible through surveys or behavior, such as cognitive load or task engagement. Finally, future work should take special consideration when designing studies involving multiple participants and rigorously control for communication styles, prior relationships, and even cultural differences.

References

[1] J. Galegher, R. E. Kraut, and C. Egido, *Intellectual teamwork: Social and technological foundations of cooperative work*. Psychology Press, 2014.

[2] S.-Y. Tu and A. H. Sayed, "Distributed decision-making over adaptive networks.," *IEEE Trans. Signal Processing*, vol. 62, no. 5, pp. 1054–1069, 2014.

[3] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," in *Telemanipulator and telepresence technologies*, vol. 2351, pp. 282–293, International Society for Optics and Photonics, 1995.

[4] M. A. Livingston, S. Russell, J. W. Decker, E. Leadbetter, and A. Gilliam, "Cedars: Combined exploratory data analysis recommender system," in *Large Data Analysis and Visualization (LDAV), 2015 IEEE 5th Symposium on*, pp. 139–140, IEEE, 2015.

[5] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson, "Immersive Analytics," in *2015 Big Data Visual Analytics, BDVA 2015*, pp. 1–8, IEEE, sep 2015.

[6] C. Donalek, S. G. Djorgovski, A. Cioc, A. Wang, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, *et al.*, "Immersive and collaborative data visualization using virtual reality platforms," in *Big Data (Big Data), 2014 IEEE International Conference on*, pp. 609–614, IEEE, 2014.

[7] A. Moran, V. Gadepally, M. Hubbell, and J. Kepner, "Improving big data visual analytics with interactive virtual reality," in *High Performance Extreme Computing Conference (HPEC), 2015 IEEE*, pp. 1–6, IEEE, 2015.

[8] A. Dan and M. Reiner, "Eeg-based cognitive load of processing events in 3d virtual worlds is lower than processing events in 2d displays," *International Journal of Psychophysiology*, vol. 122, pp. 75–84, 2017.

[9] A. Steed, Y. Pan, F. Zisch, and W. Steptoe, "The impact of a self-avatar on cognitive load in immersive virtual reality," in *Virtual Reality (VR), 2016 IEEE*, pp. 67–76, IEEE, 2016.

[10] J. P. McIntire, P. R. Havig, and E. E. Geiselman, "What is 3D good for? A review of human performance on stereoscopic 3D displays," p. 83830X, 2012.

[11] J. P. McIntire and K. K. Liggett, "The (possible) utility of stereoscopic 3D displays for information visualization: The good, the bad, and the ugly," *2014 IEEE VIS International Workshop on 3DVis (3DVis)*, 2014.

[12] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister, "The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality?," vol. 24, pp. 457–467, jan 2017.

[13] S. Borsci, G. Lawson, and S. Broome, "Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance," vol. 67, pp. 17–26, feb 2015.

[14] D. Salanitri, C. Hare, S. Borsci, G. Lawson, S. Sharples, and B. Water Fi Eld, "Relationship between trust and usability in virtual environments: An ongoing study," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9169, pp. 49–59, Springer, Cham, 2015.

[15] S. Borsci, G. Lawson, B. Jha, M. Burges, and D. Salanitri, "Effectiveness of a multidevice 3D virtual environment application to train car service maintenance procedures," *Virtual Reality*, vol. 20, pp. 41–55, mar 2016.

[16] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[17] D. H. Mcknight, M. Carter, J. B. Thatcher, and P. F. Clay, "Trust in a specific technology: An investigation of its components and measures," *ACM Transactions on Management Information Systems (TMIS)*, vol. 2, no. 2, p. 12, 2011.

[18] S. M. C. Davison, C. Deeprose, and S. Terbeck, "A comparison of immersive virtual reality with traditional neuropsychological measures in the assessment of executive functions," vol. 30, pp. 1–11, apr 2017.

[19] R. M. Reitan and D. Wolfson, *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*, vol. 4. Reitan Neuropsychology, 1985.

[20] M. S. Dennison, A. Z. Wisti, and M. D'Zmura, "Use of physiological signals to predict cybersickness," *Displays*, vol. 44, pp. 42–52, 2016.

[21] M. S. Dennison and M. D'Zmura, "Cybersickness without the wobble: Experimental results speak against postural instability theory," *Applied ergonomics*, vol. 58, pp. 215–223, 2017.

[22] M. Dennison and M. D'Zmura, "Effects of unexpected visual motion on postural sway and motion sickness," *Applied ergonomics*, vol. 71, pp. 9–16, 2018.

[23] J. B. Walther, "Theories of computer-mediated communication and interpersonal relations," *The handbook of interpersonal communication*, vol. 4, pp. 443–479, 2011.

[24] M. Alavi, "Computer-mediated collaborative learning: An empirical evaluation," *MIS quarterly*, pp. 159–174, 1994.

[25] N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright, "Effects of four computer-mediated communications channels on trust development," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 135–140, ACM, 2002.

[26] J. B. Walther, "Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction," *Communication research*, vol. 23, no. 1, pp. 3–43, 1996.

[27] G. Bubaš, "Computer mediated communication theories and phenomena: Factors that influence collaboration over the internet," in *3rd CARNet users conference, Zagreb, Hungary*, Citeseer, 2001.

[28] J. B. Rotter, “A new scale for the measurement of interpersonal trust,” *Journal of personality*, vol. 35, no. 4, pp. 651–665, 1967.

[29] N. M. Puccinelli and L. Tickle-Degnen, “Knowing too much about others: Moderators of the relationship between eavesdropping and rapport in social interaction,” *Journal of Nonverbal Behavior*, vol. 28, no. 4, pp. 223–243, 2004.

[30] T. T. Trout, S. Russell, A. Harrison, M. Dennison, R. Spicer, E. S. Rosenberg, and J. Thomas, “Collaborative mixed reality (mxr) and networked decision making,” in *Next-Generation Analyst VI*, vol. 10653, p. 106530N, International Society for Optics and Photonics, 2018.

[31] J. R. Michaelis, “Value of information driven content management in mixed reality infrastructures,” in *Next-Generation Analyst VI*, vol. 10653, p. 106530O, International Society for Optics and Photonics, 2018.

[32] L. R. Marusich, J. Z. Bakdash, E. Onal, M. S. Yu, J. Schaffer, J. O’Donovan, T. Höllerer, N. Buchler, and C. Gonzalez, “Effects of information availability on command-and-control decision making: performance, trust, and situation awareness,” *Human factors*, vol. 58, no. 2, pp. 301–321, 2016.

[33] A. Raglin and A. Harrison, “Reasoning with an uncertainty of information measure: decision making for military and non-military applications,” in *Next-Generation Analyst VI*, vol. 10653, p. 1065308, International Society for Optics and Photonics, 2018.