# "IT'S MY WAY OR THE HIGHWAY"; Location Markup Prompts and Interaction Caveats in VR

*Bahar Memarian[1*], Christos Koulas[2] and Brian Fisher[3]*

*[1] University of Toronto, Canada*

*[2] MDA corporation, Canada*

*[3] Simon Fraser University, Canada*

## ABSTRACT

The use of augmented or virtual reality (AR/VR) wearables is becoming popular in command-and-control tasks such as location markup on a virtual map for remote training or collaboration purposes. Studies in this domain capitalize personalization through unique technological features to improve the situation awareness and performance of users, without probing onto the information requirements for functions such as markup that must be consistently presented across technologies and users' abilities. We aim to characterize and visualize location markup interactions in VR by proposing a systematic perspective to code and formulate mission prompts and identify variables that may contribute to the operator's credible task execution. We find that future research needs to focus on developing universal multimodal communication conventions so that they can be consistently used in virtual or real environments. This can in turn have a significant contribution to the users' appropriate skill training and error prevention.

**Keywords:** Virtual Reality, Command-and-Control, Map Location Markup, Interaction, Human Factors

# INTRODUCTION

Among other applications (ter Haar, 2005), there are interests to use augmented or virtual reality (AR/VR) wearables in the task force for remote training or collaboration where the display and user interaction are emulating a complex command-and-control task. To provide an example, a user wearing a VR headset is presented with a third-person view map of a city and is tasked to find and markup location(s) for a specific function (e.g., disarm or divert a threat) based on a prompt. The task is expected to concurrently happen in the real world at unfamiliar locations within a set time frame. Such task is completed in a terrain, which may be hostile (contain adversaries), through engaging a combination of assets: in-the-field UAVs, human team player(s), and remote human team player. The remote team player is expected to process and interact with the virtual map presented through the VR wearables to participate in simulation training or collaborate with the asset team player that is in the real-world terrain.

The premise for using wearable technology in the task force is to unlock design features that would benefit the user's experience and interaction (UX/UI) and ultimately performance outcomes both within their cognitive and ergonomic abilities. A good VR environment design like any other visual map interaction medium (e.g., mobile) is deemed to capture the necessary complexities of the real-world environment and take the user to where it wants to go objectively; safely, accurately, efficiently, easily, etc. In other words, the VR environment needs to, on one hand, accurately cue the realities of the real-world environment, and on the other hand, simplify and summarize the realities based on a set of key objectives so it does not exceed the cognitive and ergonomic abilities of the human user. Consequently, there are emerging concerns on whether to migrate visuals and interaction conventions from a traditional setting (e.g., 2D digital spaces seen on computers) to VR or develop entirely new display visualizations and interactions given that some technological features change.

Past studies in the literature often capitalize on personalization through unique and/or new technological features to improve the situation awareness and performance of users, without probing onto the functionalities of markup that must be consistently presented across technologies and user's abilities. For example, prior work in the literature has explored virtual interaction through the use of computer vision and pattern recognition to allow each user to interact with a map in their own way and activate new map layers or receive additional information around objects of interest differently (Bobrich and Otto, 2002). Prior work has also studied hand or foot gestures for interacting with large immersive AR maps placed on the floor or in the virtual domain (Lee and Hollerer, 2008; Kim and Lee, 2016; Austin *et al.*, 2020). Working with VR map developments may resemble the saying; "It's my way or the highway" where you, as either the user or researcher, need to adapt to a developed configuration or leave it behind. As a result, one may not be able to gain or study a VR interaction skill and transfer it to another context. In the literature, work on military theory has long focused on more high-level and intricate operational process flows (Athans, 1987; Builder, Bankes and Nordin, 1999). What we still find lacking is discussions and a set of conventions on how task force operations such as location markup should be communicated and interacted with technologies (e.g., VR) in a consistent manner that can still be used across technologies and their frequent upgrades. This is critical since the command and control domain, whether using technologies such as VR or not, is expected to be focused on information processing and synthesis, reducing uncertainty, and applying or adapting responses from experience or training (Pigeau and McCann, 2002; Ioerger and He, 2003).

We acknowledge that variations in technical specifications can contribute to the divide in VR map display designs and their way of supporting users during task completion. The immersion and field of vision achievable or presented, arrangement and ergonomics of the technology from head (glass versus head-mounted) to hands (gloves versus hand controllers), sensory feedback, and rendering capabilities all play a role in the diversity of user experiences and performances. Our emphasis, however, is on aspects of VR visual design and interaction communication that can be broadly transferred from one technology to another and enable possible future standardization of command-and-control tasks for the operators in VR. We wish to emphasize some of the current VR map location markup caveats that are either part of the realities of command-and-control scenarios or what the human users can understand when interacting within VR. Consequently, such considerations can be utilized when attempting to create a universal interaction in VR. Section 2 provides caveats for location markup in VR (prompts and map support), while section 3 details considerations for location markup support in VR.

## CHARACTERIZING LOCATION MARKUP IN VR

The core of VR location markup lies in working with a 3D (Orthogonal X-Y-Z) axis representation, where the user needs to specify the coordinates of the location through a combination of length and/or angular measures (e.g., in a cartesian or spherical coordinate system). A 3D axis, therefore, is the smallest building block for enabling such markup. This axis may be attached to the base of a map or objects within a map. Complex task force scenarios, however, require additional considerations that are not communicated with just a 3D axis. In this section, we provide a few important considerations around location markup in VR and attempt to classify prompts and the user's information processing more objectively. Each command-and-control task has a holistic scenario and constraints and a need to engage the assets virtually or directly in the field through a set of prompts as shown in Table 1. The scenario may be understood or recognized by the assets in more or fewer details and contain a set of instructions. For example, the human player assets may initially receive the following high-level information about the scenario they are put in:

- You are presented with a prompt and asked to work alone or collaborate with another team player [in the field] to mark a prescribed location around one or multiple adversary vans in the virtual map.
- Once you do so, an autonomous UAV is sent to that location to automatically track the adversary vans and disarm or divert them when needed.
- You will need to objectively find or subjectively infer the location of adversary vans based on the instructions presented to you through the prompt that may contain a set of rules such as the following: An adversary van (assuming unarmed) is most dangerous when moving and there pedestrians less than 10m of it with no barriers in between.

**Table 1.** Proposed classification of command-and-control prompts based on level of user engagement during markup along with example user stories.

| Mark point based on a | Example user stories |
|---|---|
| 1. Coordinate system form (least level of cognitive engagement) | Mark [+20m, +10m, +40m] or ~ [46m, +27°, +29°] with respect to North from the only adversary van present in the school parking lot |
| 2. Set of clues | Find the only adversary van present when it is parked at the North-East of the school parking lot, then mark +40m above the adversary van. |
| 3. Set of constraints | Mark +40m above the only adversary van present in the school parking lot but only when there are pedestrians within 10m of it AND when the adversary van is moving. |
| 4. Set of assumptions | Mark +40m above any moving van present that you assume or infer belongs to the adversary in the school parking lot. |
| 5. Combinations of 1 to 4 (highest level of cognitive engagement) | Mark +40m above the van that you assume or infer belongs to the adversary but only when it is parked at the North-East of the school parking lot. |

Executing tasks in a scenario such as location markup begins with a prompt presented to the user. The prompt's message may be simple or detailed, based on the number of variables, obstacles, or dynamic changes in the environment or actions involved and so there may exist variation in the user's level of engagement and information processing depending on the prompts given. In Table 1, we offer prompts and present example user stories for each.

We suggest that the least cognitively involved prompt would give the user a set coordinate location (e.g., [+20m, +10m, +40m]) to mark using the 3D axis that is in some form attached to a map. In such a task, the action of marking up is rather procedural and requires precision around using the 3D axis and matching markup with the coordinates. However, when the content of the prompt moves away from a fixed location coordinate data to either a set of clues, constraints, or assumptions, the user's engagement becomes more involved because uncertainty is introduced. We add that the prompt would be involving the user cognitively the most when a combination of different prompt styles (e.g., prompt 5 which may be a combination of prompts1 to 4 in Table 1) is used, as then both adaptability and managing uncertainty from the user side is needed.

Notice that for either prompt style presented in Table 1, there is a consistent relational message. That is location markup happens concerning a frame of reference, which is the adversary van. Moreover, the frame of reference may need to be in a certain orientation for task completion to be successful. For example, in prompt 2 in Table 1, we see that the user is instructed to find the adversary van when it is in the North-East of the school parking lot. This means the user needs to manipulate (e.g., rotate, resize, or shift) and align to the North to ensure that the van appears in the appropriate location. In prompt 4 in Table 1, on the other hand, the user is asked to mark any van without needing to worry about their orientation per se. Additionally, the frame of reference may be stationary or moving. This can be seen for prompts 3 to 4 in Table 1, where the user is asked to mark +40m above the moving adversary vans only. Furthermore, the prompt may be dynamic and dependent on the direction of the frame of reference which is not discussed in Table 1 (e.g., mark only when the car is moving in a certain direction or speed). If so, one should consider whether the frame of reference is moving the same or differently in 2D versus 3D. A UAV for example may have a different model of movement across axis planes depending on the technology and material used in its making.

As seen in Table 1, commands provided to or communicated between team players may contain different levels of objectivity and lead to the player's different interpretations and performances for the task. An important consideration is to not confuse the team player's misunderstanding of the prompts from the VR system. Not to neglect that compatibility in the awareness of agents (e.g., when different team players execute a unified prompt) plays a key role in task performance (Stanton et al., 2009). We may therefore need to prescribe a protocol for the team players to use and translate their understanding of a prompt into a set coding process. Putting the explored properties of coordinates and frames of references together, we could attempt to quantify location markup that is verbalized or prescribed in the qualitative prompt based on a set of conventions and variables (italicized and alphabetical letters used here as a placeholder) where each variable may be absent or present at different prompt instances:

- Clock convention: Mark a point at length *L* and at *T* O'clock of the object *B* oriented towards C, moving in direction *D*, at speed *E*, with additional rules *F*, *G*, etc.
- Cartesian convention: Mark a point at *X*, *Y*, *Z*, of the object *B* oriented towards *C*, moving in direction *D*, at speed *E*, with additional rules *F*, *G*, etc.
- Spherical convention: Mark a point at length L, *Theta_1*, *Theta_2* of the object *B* oriented towards C, moving in direction *D*, at speed *E*, with additional rules *F*, *G*, etc.
- Other conventions and combinations of conventions…

Coding a qualitative prompt into its quantitative physical characteristics can open the opportunity for the development of a universal convention for both the user and VR environment to understand and rely on. Yet, we should note that the prompt coding does not give a full picture of the operator's information processing. What then needs to be probed is the motivation and end goal of executing each prompt especially when a sparse overview around the scenario the team players are put in is provided. Mission completion whether completed solely by the remote operator, communicated with in the field team player, or integrated with algorithmic decision making and artificial intelligence needs to transparently answer the What-Why-How-When-Where questions when executing each action and consider alternative responses ready for extreme ends of failure/error and safety/success. We also need to caution that any convention should leave room for modifications and additions of new variables that represent the dynamics of the environment. One consideration is that the user stories presented in Table 1 may require the remote operator to seek further information from either the VR environment, team-players, or algorithms, changing the translation of the prompt from a one-line coding statement to lines of revised or updated quantitative codes.

## SUPPORTING LOCATION MARKUP IN VR

The design gaps explored in the previous section calls for the development of conventions and design ideas in VR to support the user with their task appropriately. An example of VR support design consideration is how to characterize the stages of the users' work. VR displays could support the users differently when: thinking/strategizing and navigating versus executing an action in the field. The composition of these stages may further change and be distributed over team members, time, tasks, and combinations of these. Looking more closely at the task of location markup, we find that a simple 3D axis representation may not be sufficient to characterize location markup in a complex task force environment. The 3D axis design, for example, may benefit from enabling markup across its six degrees of freedom (i.e., also pitch, yaw, roll) and further annotating the orientation, direction, and speed of frame of reference or interfering objects. An interaction consideration for the 3D axis is what convention should be deployed (e.g., hand controller buttons and motions versus simple cursor clicks) to enable the user to mark a point using the 3D axis. The number of steps taken (e.g., hand movements or clicks), the sensitivity, and the precision of markup are all relevant UI/UX considerations. The markup widget (i.e., improved 3D axis design) size needs to be realistic to the size of the objects if placed on the map and so should change in size when shifting/scaling. If overlaid on the map, the transparency of the widget is also important. The widget would benefit from supporting and presenting information in multiple prompt formats (e.g., cartesian and spherical axis simultaneously).

The type of feedback presented by the VR for the task of location markup can also play a role in the successful completion of the task. One type of feedback is verification where after the user marks up a location the display follows up by asking: "Are you sure X coordinates is what you meant?". The feedback may be textual, vocal, or better yet, overlaid onto the map to allow the user to check their work when interacting with the 3D axis. The display can also provide descriptive summaries based on historical data collected and present the user's trajectory of performance over time or against other users or team members completing a similar task. Last, the feedback provided by the display can be predictive using artificial intelligence. It would be interesting to identify whether feedback based on human or artificial or combinatory intelligence is most likely to improve user performance. The actions of the adversaries and/or assets in the field in the past would also contribute to feedback and operator's information processing (e.g., may correlate with both).

How the user gets to interact with the widget should also be further studied. For prompts such as prompt 2 in Table 1, the user needs to align the map with respect to the North to check to see whether the adversary is parked at the North-East of the VR space. Once the frame of reference (adversary van) is in the right location, the user then clicks on the frame of reference and a markup widget can appear on the top of the object (overlaid on the map) or outside the map. The user will then be using a convention enabled by the controllers (as a cursor or gesture) to mark +40m above the selected frame of reference. The VR environment can further provide a form of feedback to the user before the markup completion is confirmed. This example demonstrates that even for a relatively simple prompt in VR the interaction entails multiple steps where the sequence and type of actions and designs can be varied. Developing a markup widget, therefore, does not guarantee a unified markup experience. But rather, the labor work and procedures (e.g., touch, verbal, visual, textual) that the user needs to perform in VR via the widget or other designs to markup a location also need to be structured in a reliable way.

There may be some advantages and disadvantages with having widgets populated on map objects rather than having the user click on different objects and presenting only one widget outside the map. The advantage of having the widget on the object itself is that it is more intuitive and traceable in terms of checking where the current frame of reference is. The disadvantage is that when that map block is populated with too many objects, viewing, and interaction with the widget becomes difficult. Furthermore, the map layout is subject to variations (e.g., landscape or number of objects populated on the map) and so placing the widget on the map objects may act as a confound to the markup experience. Having one widget outside for the user to markup across different tasks and map objects, on the other hand, establishes a constant markup interaction experience. The user becomes used to marking a location within one set area, rather than different areas with variable map backgrounds. Not to neglect, the task of markup using a widget within the map area for air objects may be more difficult, and so it is worth examining the impact of having a widget inside or outside the map layout in future studies. Another advantage of having only one widget outside the map is that clutter is reduced. However, there needs to be a unique naming convention per object on the map or an alternative approach to make it easier for the user to track what is the current frame of reference and to prevent unintended errors.

A shared consideration for either widget convention is the markup shape. Here we have assumed the markup to be a point, but we need to acknowledge that a user in a task force scenario may also require marking one or multiple of a) points having different locations, b) lines having different lengths and thickness and locations, or c) regions having different points and lines. For example, multiple points may be marked to show safe zones, or a line can be drawn to show the direction of a threat. In summary, one could consider the degree to which the VR display can provide support to the user for interaction such as searching for or marking up a location on a digital map to lie on a spectrum. One end of the VR interaction support spectrum is manual. That is the user manipulates the map such as resizes, shifts, moves around such as in a table-top setting, and processes other typical map information available to infer and mark the location of interest for a prompt given. The other end of the VR interaction support spectrum is automatic. That is the user relays some details (either prompts or translated codes) and the display takes care of the information processing in the background and marks that location. The manual end of the spectrum benefits from allocating decision-making and processing entirely to the human user but runs the risk of errors due to the users' many manipulations of the map (e.g., rotation, resizing, etc.). The automatic end of the spectrum on the other hand outsources markup entirely to the display which may be more robust if developed appropriately, however, runs the risk of wrong location markup due to user's typos or wrong coding or communications if done vaguely or inconsistently. Each of the available VR map display designs on the market may lie somewhere on this spectrum and have certain flexibility to have their code modified and move along the spectrum. As such, it is important to lay design foundations for conventions and metrics consistently that can be used and communicated in command-and-control tasks irrespective of the technology used and where it lies in the VR's technological spectrum.

## CONCLUSION

We aimed to characterize and visualize location markup interactions in VR by proposing a systematic perspective to code and formulate mission prompts through a 5-level classification and noting variables that may contribute to the operator's credible location markup in VR through a 3D visual widget view. We find that future research needs to focus on developing universal multimodal communication conventions so that they can be consistently used by task force personnel over virtual and real environments. This can in turn have a significant contribution to the user's appropriate skill training and error prevention.

## ACKNOWLEDGMENTS

We would like to acknowledge the corresponding author Dr. Bahar Memarian for the design and description of the prompts and location markup design considerations and Dr, Brian Fisher for providing resources and guidance on this work. We would like to thank *Defence Research and Development Canada (DRDC)*'s *Innovation for Defence Excellence and Security (IDEaS)* program for supporting this work.

## REFERENCES

Athans, M. (1987) 'Command and control (C2) theory: A challenge to control science', *IEEE Transactions on automatic control*, 32(4), pp. 286–293. doi: 10.1109/TAC.1987.1104607.

Austin, C. *et al.* (2020) 'Elicitation study investigating hand and foot gesture interaction for immersive maps in augmented reality', *Cartography and Geographic Information Science*, 47(3), pp. 214–228. doi: 10.1080/15230406.2019.1696232.

Bobrich, J. and Otto, S. (2002) 'Augmented maps', *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(4), pp. 502–505.

Builder, C., Bankes, S. and Nordin, R. (1999) 'Command concepts: A theory derived from the practice of command and control'. Santa Monica, CA: Rand.

ter Haar, R. (2005) 'Virtual Reality in the Military: Present and Future', in *3rd Twente Student Conference on IT. University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.* Science.

Ioerger, T. and He, L. (2003) 'Modeling command and control in multi-agent systems'. 8th International Command and Control Research and Technology Symposium (ICCRTS).

Kim, M. and Lee, J. Y. (2016) 'Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality', *Multimedia Tools and Applications*, 75(23), pp. 16529–16550. doi: 10.1007/s11042-016-3355-9.

Lee, T. and Hollerer, T. (2008) 'Hybrid feature tracking and user interaction for markerless augmented reality', in *IEEE Virtual Reality Conference*. IEEE, pp. 145–152.

Pigeau, R. and McCann, C. (2002) 'Re-conceptualizing command and control'. Toronto Canada: Defence and Civil Institute of Environmental Medicine. Available at:

https://apps.dtic.mil/sti/pdfs/AD1005370.pdf.

Stanton, N. *et al.* (2009) 'Genotype and phenotype schemata as models of situation awareness in dynamic command and control teams', *International Journal of Industrial Ergonomics*, 39(3), pp. 480–489. doi: 10.1016/j.ergon.2008.10.003.