The Effectiveness of Gaze Guidance Lines in supporting JTAC's Attention Allocation

*Domenick Mifsud, *Chris Wickens, **Michael Maulbeck, **Peter Crane, *Francisco R. Ortega
*Colorado State University, **VRR, Inc.

We examined the utility of gaze guidance lines to facilitate the direction of attention between features on a minimap and their counterparts in a virtually rendered visual scene (Experiment 1) and between described features in the visual scene and their location on the minimap (Experiment 2); the experiment was designed to simulate a critical element of scene understanding for joint tactical attack controller (JTAC) communications with an attack pilot. In each experiment participants encountered 40 scenarios, half supported by a gaze guidance line and half without. The results of both experiments yielded a large and significant benefit of the gaze guidance line over the control condition, expressed in both speed and accuracy. The results of Experiment 2 however revealed some evidence that the gaze guidance lines led to an automation bias when the automation algorithm, creating the lines, misidentified a landmark in the visual scene.

INTRODUCTION

The role of the Marine Corps Joint Tactical Air Controller (JTAC) in close air support is to direct the actions of combat aircraft through descriptions of hostile targets as well as granting clearance to engage those targets. A critical component in the tasks of the JTAC is the ability to make comparisons between a map and the direct view of the battlefield. In recent research we have explored the use of the head mounted display (HMD) for the JTAC (Wickens, Mifsud, Rodriguez & Ortego, 2021). In both that and the current research, the direct view to battlefield needs to be compared with an electronic "minimap" positioned on the bottom of the field of view of the HMD. Such image comparisons may be required for example when a target object is designated electronically or by coordinate descriptions on the minimap, and then the JTAC must look at the real-world scene to locate that object and confirm its presence or identity. Alternatively, the JTAC must describe in more detail its appearance and relative location to an attack pilot as the aircraft approaches the target area. Moving attention in the opposite direction, the JTAC may see an important object and potential target (e.g., building or vehicle) in the forward view through the HMD, and wish to assess its location relative to other features depicted on the HMD-displayed minimap, perhaps to assure that the target has sufficient separation from friendly or neutral elements. Such relative location judgments can best be made by consulting the more top-down, non-compressed representation on the minimap, because the perceptual compression imposed on the real-world scene, as viewed from the low slant angle of the JTAC observation point, severely distorts the accuracy of location judgments (Wickens et al., 2021).

The task of attentionally "linking" two different objects (i.e., representation of the same target on the map and in the scene), in a cluttered scene as viewed from two different perspectives (i.e., the forward view of the scene and the higher angle view of the minimap) is challenging (Schreiber et al., 1998; Wickens & Hickox, 1999; Wickens, 2002). This can be particularly challenging if the target object is close to and looks similar to another object that is not the target. Consider for example an enemy building close to a similar looking structure that houses a school. Assuming that the structure with the school is the target can have disastrous consequences.

Such comparative judgments are consequential for any warfighter such as a UAV pilot or sensor operator, and, not just the JTAC, who needs to verify forward view information against that depicted on a top down electronic map display.

In order to facilitate performance on this difficult image comparison task, we have developed the AR concept of the **gaze guidance line** that draws a virtual line from the target representation on the minimap to its counterpart in the scene. The psychological basis for the benefits of the gaze guidance lines is provided by the **proximity compatibility principle** (Wickens & Carswell, 1995) which asserts that two objects that need to be mentally related or compared (e.g., confirming the common identity of the scene and mini-map object) should also be "linked" in the perceptual view. Such linkage can be achieved via a variety of techniques (e.g., close proximity in space); but in particular, by any physical line that connects them (see Wickens, McCarley & Gutzwiller, 2022), such as the line connecting two data points on a line graph (Schultz, 1961), the dashed line connecting verbal instructions with a component on the image of a piece of equipment to be assembled (Paas & Van Gog, 2009), or the line connecting the comment in a comment field to the relevant word or phase in the text. The basic psychology of attention behind this role of lines in attentional guidance is provided by Jolicoeur and Ingleton (1991), in which they explored response times in response to linking data points on differing line lengths. We believe that this concept could be applied to any task one needs to perceptually link two objects, such as mapping a moving target on a radar to the physical object or mapping a target from an external perspective (such as a drone) to the operator's perspective.

We sought to answer the question of: How well does this technique work to improve response times and accuracy when applied to real-world tasks encountered by JTACs? The current pair of experiments were designed to examine the degree of benefits offered by the gaze guidance lines (GGL) in (a) directing attention from the minimap to the outside scene (Experiment 1), and (b) directing attention in the opposite direction, from an object in the outside scene to its counterpart on the minimap (Experiment 2). In both cases, we hypothesized both statistically and practically significance of the GGLs to performance

EXPERIMENT 1: Directing attention from the Minimap to the scene

METHOD

Materials. Twenty scenes (scenarios) were created in virtual reality that depict five locations along the Front range of the Rocky Mountains in Colorado. These were carefully chosen to contain a mixture of terrain, some roads, and some buildings; that is, sufficient scene complexity to make the task a challenging one. Each of the five scenes was rendered in four images viewed from an observation point 100 feet off the ground, facing North, East, South, and West. Each of these 20 scenarios was rendered on a VR image to be viewed in the VIVE. Within this image was overlaid an AR minimap, depicting a 0.75-mile diameter circular rendering of the scene from a 45 degree slant (lookdown) angle. As shown in Figure 1, this map contained a "wedge" depicting the visual angle of the forward view of the scene through the VIVE. The minimap rotated with the direction of gaze, so that it would rotate as the participant moved their head and also would rotate between scenes depending on the initial N, E, S, W orientation. The minimap was positioned approximately 30 degrees below the center of the field of view, such that the upper rim of the minimap did not obscure (overlay) important features of the VIVE-rendered terrain (See Figure 1). The only difference between the GGL condition and the control condition was the inclusion of the GGL.

Two synthetic objects were then carefully placed on the scene terrain (to also appear in their corresponding locations in the minimap). One was the target object (e.g., a tall building, a water tower, etc.) and the other was a "foil". The foil was a similar object, placed close to the target in distance, and selected to be of similar in physical appearance. The foil was intended to create the ambiguity of target appearance that has provided a particular challenge for JTAC operations (personal observation by project SME). In order to add complexity to the scene in general, additional buildings and some synthetic roads were also added (again, each appearing in their corresponding locations in the minimap). The target was always cued on the minimap by an augmented reality red arrow pointing directly to it as seen in Figure 1. The yellow augmented reality gaze guidance line, when present as shown in Figure 1, connected the rendering of the target object in the minimap with its counterpart in the forward view.

Procedures. After completing the consent form and receiving instructions and four practice trials, each experimental trial began as the participant viewed a blank screen. When the participant indicated with a press on a trigger of the Xbox controller, both the scenario and minimap appeared in view, with the latter containing the arrow-designated target. The participant was then requested to center their gaze cursor directly on the target object in the forward scene. As soon as they were confident that they had selected the correct target, they pressed a trigger on the Xbox controller. This was immediately followed by the display of a red outline surrounding what the system inferred to be the participant-selected object. The time between the scene exposure and this first trigger press was recorded as the response time for the trial. If the participant confirmed the red box surrounding the item to be that which they intended, they clicked a second "confirm" trigger. In the very unusual case that it was not, the participant could refocus the head cursor on the intended target and repeat the process. Upon selection of the confirm response, the screen became blank, and the participant could initiate the subsequent trial when they chose.

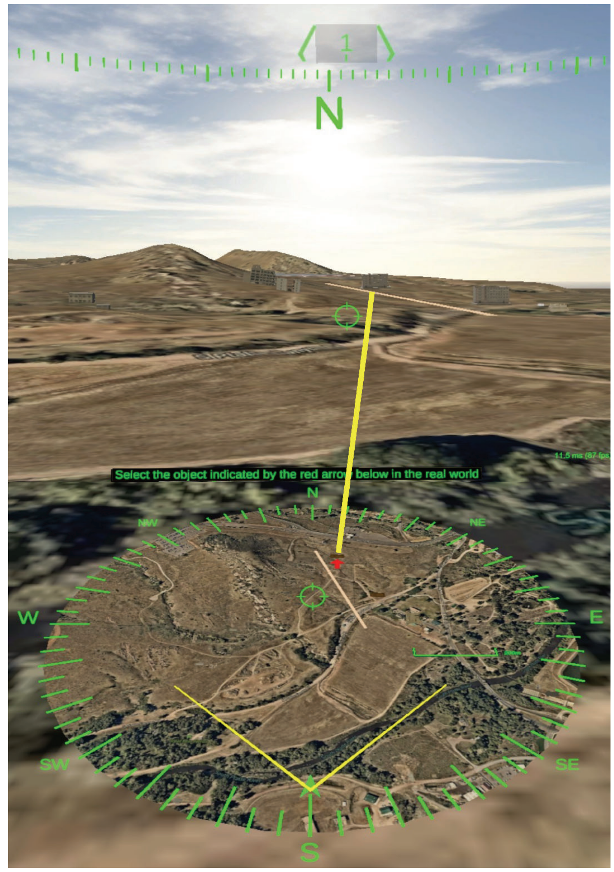

Figure 1. The participant's display on a north-oriented scene. The red arrow cues the target building on the minimap whose counterpart in the world can be seen just below the horizon. The reticle of the gaze pointer is seen just below and to the left of the building. The "foil" building can be seen to the right of the target building just down the straight highway. The quality in these images is degraded relative to their appearance in direct viewing of the VIVE. The green text displayed above the minimap simply reminded participants of the task.

Design. Participants completed 40 trials, with each scene represented twice. In one block of 20 trials, they performed in the control condition. In the other block, they performed with the gaze guidance line, connecting the two representations of the target by a yellow line as shown in Figure 1. Half of the participants received the GGL in the first block, and half received the control condition in the first block. Each scene was presented once in the control condition and once in the GGL condition. Instructions were provided before the first block, differing depending on the first condition encountered. Then short instructions were also issued prior to the second block introducing the other condition. Practice trials were completed before starting both conditions. Following completion of the experiment participants completed a short evaluation about their thoughts on the system.

Participants. 21 volunteers were each reimbursed with an $20 Amazon Gift card. Four of these were ROTC cadets.

RESULTS AND DISCUSSION: Experiment 1

A 2 (Guidance Lines vs control) x 2 (counterbalancing group) ANOVA was carried out on the data, the display factor was a repeated measures factor and counterbalancing group was varied between participants.

The response time data revealed a highly significant benefit for the GGL (M=1.25 sec, SE=0.07) condition over the control (M=4.6 sec, SE=0.22) condition, $F(1,19)=118.3$, $p<.001$. The accuracy data revealed a similar advantage of the GGL (M=99.0%, SE=0.47) condition over the control (M=84.3%, SE=1.78) condition, $F(1,19)=55.1$, $p<.001$. There was no significant effect of counterbalancing group for response time, $F(1,19)=0.15$, $p=.704$, or accuracy, $F(1,19)=2.69$, $p=.117$, nor a group x condition interaction for response time, $F(3,19)=1.89$, $p=.185$, or accuracy, $F(3,19)=1.75$, $p=.201$.

As apparent from the above statistics, the gaze guidance lines provided a very large benefit to this target matching task. In addition, it is noteworthy that the accuracy of the control condition was relatively high (84%). In this regard, considerable effort was made in designing the scenarios to assure that they were neither nearly impossible (so that the GGL benefit would be guaranteed) nor too easy, so that there would be a ceiling effect in the performance measures and hence no opportunity to see a benefit in the data.

One potential concern in Experiment 1 is that participants did not really need to identify the target object at all in the scene, but simply focus their attention on the end of the line. Participants were instructed to confirm the target identity as would a JTAC and not entirely rely upon the GGL, but we could not verify that all participants did so. This concern was addressed in Experiment 2.

EXPERIMENT 2: Directing attention from the scene to the Minimap

METHODS

In this experiment the GGL directed attention in the opposite direction from that used in Experiment 1. It simulated the task of a JTAC who observes a significant object in the scene and now must locate it on the minimap. For example, to assess an object's precise coordinates or ensure its separation from other neutral or friendly elements in the scene, which could not be clearly seen from the low slant angle of the scene observation point. The procedures and scenarios were identical to those employed in Experiment 2 with the important difference that a trial begin with the participant viewing, on an otherwise blank screen, a description of the target object. Four examples of these questions are to locate:

- The tall building southeast of the road at the base of the hills
- The building north of the straight road close to the mountain
- The shorter building at the end of the straight road just west of the small town
- The building on top of the farthest hill to the west.

The description of target object location and appearance was sufficient that with careful study, the participant could be expected to identify it with perfect accuracy. However, given the similarity of the foil (non-target) object, a too-hasty assessment in the control condition could lead them to attend to the foil instead, and hence select the foil, (or another object) on the minimap: an error. The red arrow in the Minimap was was removed.

On two of the twenty scenarios (10%), the gaze guidance lines mistakenly connected the described target and its position in the scene, to an incorrect object – the foil – on the minimap. This was done to simulate the fact that an AI generated cue, such as the GGL, depending on computer vision, may occasionally be in error (Yeh & Wickens, 2001). Participants were alerted to this possibility in their instructions with the wording: "**HOWEVER,** ensure that the objects being pointed to by the GGL are correct as the system will not always point to the correct object, hence there is a need to always check the raw image data". Incorporation of these two scenarios could reveal the extent to which participants did so, rather than succumbing to what is called the "automation bias" (Mosier, Skitka, Heers & Burdick, 1998; Parasuraman & Manzey, 2010).

Ten volunteers, who did not participate in Experiment 1, were, reimbursed for their participation as in Experiment 1. Three were ROTC cadets.

RESULTS AND DISCUSSION: Experiment 2

The response time data revealed a highly significant benefit for the GGL (M=2.69 sec, S=0.21) condition over the control (M=12.82 sec, SE=0.69) condition, $F(1,8)=88.62$, $p<.001$. The accuracy data revealed a similar advantage of the GGL (M=93.5%, SE=1.62) condition over the control (M=68.5%, SE=3.29) condition, $F(1,8)=70.4$, $p<.001$. There was no significant effect of counterbalancing group for response time, $F(1,8)=2.63$, $p=.144$, or accuracy, $F(1,8)=1.22$, $p=.302$, nor a group x condition interaction for response time, $F(3,8)=1.21$, $p=.304$, or accuracy, $F(3,8)=0.96$, $p=.356$.

A separate analysis examined, in the GGL condition, performance on those two trials in which the GGL was in error. These results revealed a mean response time of 3.3 seconds and accuracy of 57.14%.

The lower performance for both conditions as compared to their values in Experiment 1, indicated that the task in Experiment 2 was considerably more difficult. Such difficulty was both reflected in the need here, to remember the instructions (or to re-read them after the scene search started) and, in the control condition, to visually search for the described object, which was more difficult than simply searching for the red arrow in Experiment 1.

DISCUSSION

The goal of this research was to examine the effects that the addition of the gaze guidance line has on the search performance. In both experiments a significant advantage was observed in both response time and accuracy when using the gaze guidance line.

The gaze guidance line improved response time most significantly in Experiment 1 when participants went from searching the minimap to the scene. However, based on the differences in accuracy, Experiment 1 did not provide as challenging of a task as anticipated. Experiment 2 provided a much more difficult task, as evident in the accuracy differences between the experiments, and may provide a more realistic assessment of the positive impacts of the gaze guidance line in a real-world scenario.

In the second experiment, we also showed that the gaze guidance line may have some negative effects reflecting the automation bias that cannot be seen when only judging performance in terms of accuracy and response time across all trials. Regardless of the positive impacts on performance, any application of the gaze guidance line, especially in a time critical scenario, must bear consideration of the potential for users to blindly follow wrong instructions.

As mentioned before, many tasks, beyond those of the JTAC, rely on the operator mapping elements in two environments to one another. We feel that these results can extrapolate to most tasks that involve routing attention from one medium to another, as we have demonstrated the GGL's positive effects in a real-world scenario.

A note of caution, however, relates to the relatively high error rate observed on the two trials when the GGLs were incorrect. This indicates that many participants may not have been carefully examining the image to which the GGL was connecting, as would be anticipated for military personnel using this form of AI in combat. Notwithstanding this, other studies examining AR cueing have also revealed the tendency of participants, even experts to sometime over-rely on AI, and hence make mistakes when the AI itself is incorrect (Yeh & Wickens, 2001; Parasuraman & Manzey, 2010). This again is the well-known "automation bias".

CONCLUSIONS AND LIMITATIONS

In conclusion, the inclusion of the gaze guidance showed an improvement in both response time and accuracy in two search tasks; however, care should be taken when applying the gaze guidance line to critical tasks in which a human may assume that the automation has made no errors.

A big limitation behind our research was the small pool of ROTC cadets used. A larger group of ROTC cadets would have allowed us to better judge how the gaze guidance line assists those already somewhat familiar with the search task (due to ROTC training) in simulated naturalistic environments. In future work, it may be beneficial to examine the impact of other augmented reality cueing methods (Warden, Wickens, Mifsud, Ourada, Clegg & Ortega, 2022) as well as the impact of imperfect cueing.

ACKNOWLEDGEMENTS

We acknowledge the funding support from ONR N00014-19-C-2026, ONR N00014-21-1-2580, and VRR 19-C-2026-CSU-PHII. We would like to express thanks to Dutch Alessi from Virtual Reality Rehab, Inc.

This research was supported by the Office of Naval Research under grant numbers N00014-21-1-2949 and N00014-21-1-2580. Dr. Peter Squire was the scientific/ technical monitor.

REFERENCES

Hickox, J., C. & Wickens, C.D. (1999). Effects of elevation angle disparity, complexity, and feature type on relating out-of-cockpit field of view to an electronic cartographic map. Journal of Experimental Psychology: Applied, 5(3), 284-301.

Joliicoeur P. & Engleton, M. (1991) Size invariance in curve tracing. Memory & Cognition, 19, 21-36.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision-making and performance in high-tech cockpits. International Journal of Aviation Psychology, 8, 47–63.

Paas, F., & van Gog, T. (2009). Principles for designing effective and efficient training for complex cognitive skills. In F. Durso (Ed.), Reviews of human factors and ergonomics (Vol. 5, pp. 166–194). Santa Monica, CA: Human Factors and Ergonomics Society.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human factors, 52(3), 381-410.

Schreiber, B.T., Wickens, C.D., Renner, G. J., Alton, J., & Hickox, J.C. (1998). Navigational checking using 3D maps: The influence of elevation angle, azimuth, and foreshortening. Human Factors, 40(2), 209-223.

Waarden, A., Wickens, C. Mifsud, D., Ourada, S., Clegg, B, & Ortega, F. Visual Search in Augmented Reality: Effect of Target Cue Type and Location. *Proceedings, 2022 Annual Conference of the Human Factors & Ergonomics Society.* Sage.

Wickens, C.D., Mifsud, D., Rodriguez, R. & Ortego, F. (2021) Mitigating the costs of spatial transformations with a situation awareness augmented reality display: Assistance for the Joint Terminal Attack Controller. Human Factors, 63 3-17.

Wickens, C.D. (2002). Spatial awareness biases (Technical Report ARL-02-6/NASA-02-4). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C.D. McCarley, J.M. & Gutzwiller, R. (2022). Applied attention theory 2nd Edition. Boca-Raton, FL: CRC Press, Taylor & Francis.

Wickens, C.D., & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors, 37(3), 473-494.

Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. Human Factors, 43(3), 355-365.