2018

Evaluating Augmented Reality for Space Telerobotics Training

John Karasinski

Evaluating Augmented Reality for Space Telerobotics Training

Date of Report: September 30th, 2018

Fellow:	**John Karasinski**
Advisor:	**Dr. Stephen K. Robison**
Institution:	**University of California, Davis**
Department:	**Mechanical and Aerospace Engineering**

Contents

1. Narrative

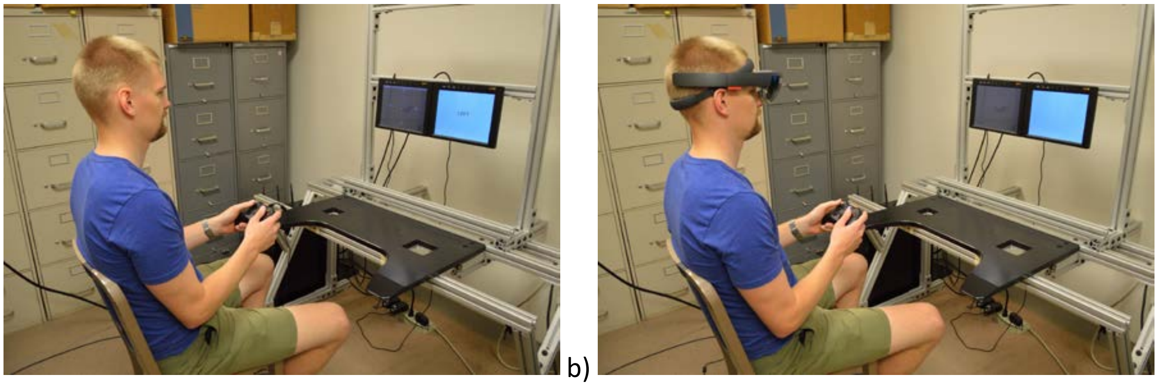
Introduction

Despite the extensive previous research into manual tracking, to the author's knowledge there exists no study in the literature addressing human performance or workload changes in manual tracking tasks between traditional computer monitors and mobile, augmented reality headsets. The aim of this study was to investigate the effect of several factors on human performance and workload in a three-axis manual tracking task. Recent advances in computing hardware have enabled a new generation of augmented reality stereoscopic devices, such as the Microsoft HoloLens, which have yet to be evaluated in the literature. The Microsoft HoloLens provides depth cues to the user which are not available with traditional 2D displays. If 3D displays can successfully improve performance and decrease workload in a manual tracking task, then it may also be valuable in the realm of robotics tasks. The human control of spacecraft, underwater, surface, and surgical robotics operations are all exciting areas of potential benefit. Each of these areas poses unique domain specific challenges, but the role of the human controller in robotic operations is often based on a simple three-axis tracking task.

We designed an experiment that investigated if the depth cue offered by 3D displays could improve performance and decrease workload. With the knowledge that We were also interested to see if a 2D display could gain the benefits of a depth cue by rotating the axis of the task such that the depth cue was more readily available. Research has shown that presenting three-dimensional information on a two-dimensional screen is not a simple task, and that the projection of the 3D information onto the 2D screen can cause large changes in the performance of the user [3]. In addition to these cues, we also investigated the effects of concurrent bandwidth feedback on task performance and workload as an alternate technique to improving performance. Concurrent bandwidth feedback alerts the operator when their real-time performance has drifted outside an acceptable, predefined window of performance. The use of feedback has been shown to improve performance in a wide variety of motor control tasks [4, 5]. We added this countermeasure as we have previously found it to be effective in similar tasks [6].

This study assessed the influence of display type (perspective vs. stereoscopic), relative display attitude (zero degrees vs. thirty degrees), and concurrent bandwidth feedback (with vs. without) on performance and workload. Objective performance was measured using the root-mean-square error (RMSE) of the depth (z) axis, and subjective performance was measured with the use of a questionnaire. Objective workload was measured using the response time to a secondary, two-choice task, and subjective workload was measured using the NASA-TLX [7]. It was hypothesized that:

1. Concurrent bandwidth feedback improves performance in the depth (z) axis for both display types and will decrease workload.
2. Stereoscopic augmented reality displays improve performance in the depth (z) axis, but do not affect workload.
3. Rotating the display improves performance in the depth (z) axis for both display types and will decrease workload.



a) b)

Figure 1: The simulator with an example participant in the a) 2D, LCD group, and b) 3D, HoloLens group.

A human-in-the-loop simulation was conducted using a fixed-base simulator, see Figure 1. The simulator consisted of two 10.4-inch LCD displays. The tracking task was completed on the left display, while the right display showed the two-choice task. For participants in the 3D group, the left LCD monitor was turned off, and the tracking task was instead displayed on the HoloLens. Subjects in both groups used the same Microsoft Xbox controller and control scheme to complete the task. Before entering the study, participants were randomly placed in a display group, and were then randomly placed into an order group (which consisted of Baseline-Feedback-Rotated, Feedback-Rotated-Baseline, and Rotated-Baseline-Feedback). This order group was created to remove any order effects that might arise due to training on a given display. The order of the designs was expected to be insignificant, but we will later discuss how this is not the case.

Three designs were presented to the subjects to evaluate: a baseline design, a color-based concurrent bandwidth feedback (CBF) design, and a rotated design. Figure 2 shows all three designs in the same error state. The three designs were very similar, having only minor differences between each other. The baseline design consists of a flat cross with a center target point and a green sphere error indicator. This indicator also casts a green, variable-length rod perpendicular to the plane of the cross, which allows for a visual estimation of the error in the z-axis. The x-axis is parallel with the horizontal cross, while the y-axis is parallel with the vertical cross. The color feedback display was identical to the baseline design in every way, with the additional of visual concurrent bandwidth feedback on the z-axis. When the absolute value of the error on the z-axis exceeded a fixed bandwidth, the color of both the spherical indicator and the cylindrical rod changed from green to red. When the absolute value of the error on the z-axis was lowered back below this fixed bandwidth, the indicator changed back to a green color. The rotated display was identical to the baseline design, but the relative attitude of the display was rotated about the y-axis by 30 degrees.

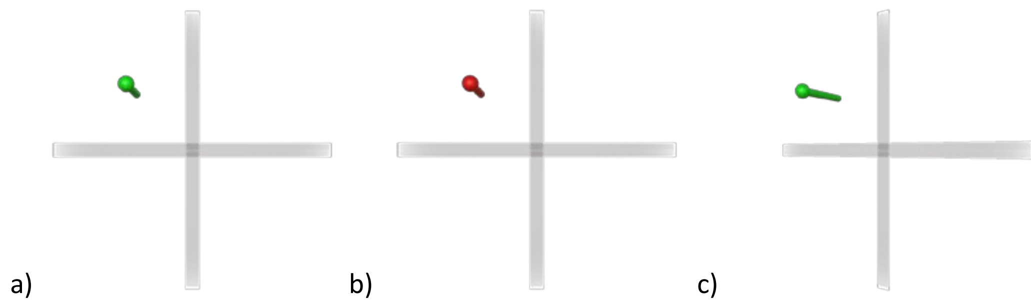


Figure 2: The three designs in the same error state. The designs are a) Baseline, b) Concurrent Bandwidth Feedback in the activated (red) state, and c) Rotated.

Results

A total of 24 participants (19 males, 5 females) were recruited in accordance with the University of California, Davis Internal Review Board (IRB), and participants signed a consent form and were not compensated. There were 12 participants in the 2D group, and 12 participants in the 3D group. Subjects were undergraduate and graduate students in the University of California, Davis College of Engineering. All participants had normal vision (no color blindness, eyesight correctable to 20/20 vision) and full motor control of their hands.

We conducted three-way mixed ANOVAs between display (2D or 3D), design (Baseline, Feedback, or Rotated), and starting design (Baseline, Feedback, or Rotated) with repeated measures on the design factor. When significant effects were observed, post hoc comparisons using the Tukey Honest Significance Difference (HSD) test with a Bonferroni adjustment were completed to investigate which pairs of the factor were significant. To remove learning and fatigue effects, each subject's best performing five trials of the ten trials in each design were averaged together to produce one average score for each subject and design.

The root-mean-square error (RMSE) of the z-axis was analyzed to understand the differences between the three designs and the two devices. The RMS of the disturbance signal was calculated and used to normalize the RMSE. Under this definition, an RMSE of 1 indicates performance no better than no input, and an RMSE greater than 1 indicates very poor performance. Results of the ANOVA on the z-axis RMSE showed significant effects for design ($F(2, 36)=84.92$, $p<.001$), device ($F(1, 18)=7.22$, $p<0.015$), and start design ($F(2, 18)=4.81$, $p<0.021$). The ANOVA also showed a significant interaction effect between design and starting design ($F(4, 36)=8.55$, $p<0.0001$), and a three-way interaction between design, device, and starting design ($F(4, 36)=5.57$, $p<0.002$). Further investigation into the effect of starting design using a Tukey HSD comparison with a Bonferroni correction showed significance differences between participants that started in the concurrent bandwidth feedback group and those in the baseline ($p<0.001$) or rotated ($p<0.001$) designs, but no difference between participants that started in the baseline and rotated designs ($p>.38$). The difference in performance along the depth axis between design, device, and starting design can be seen in Figure 3.

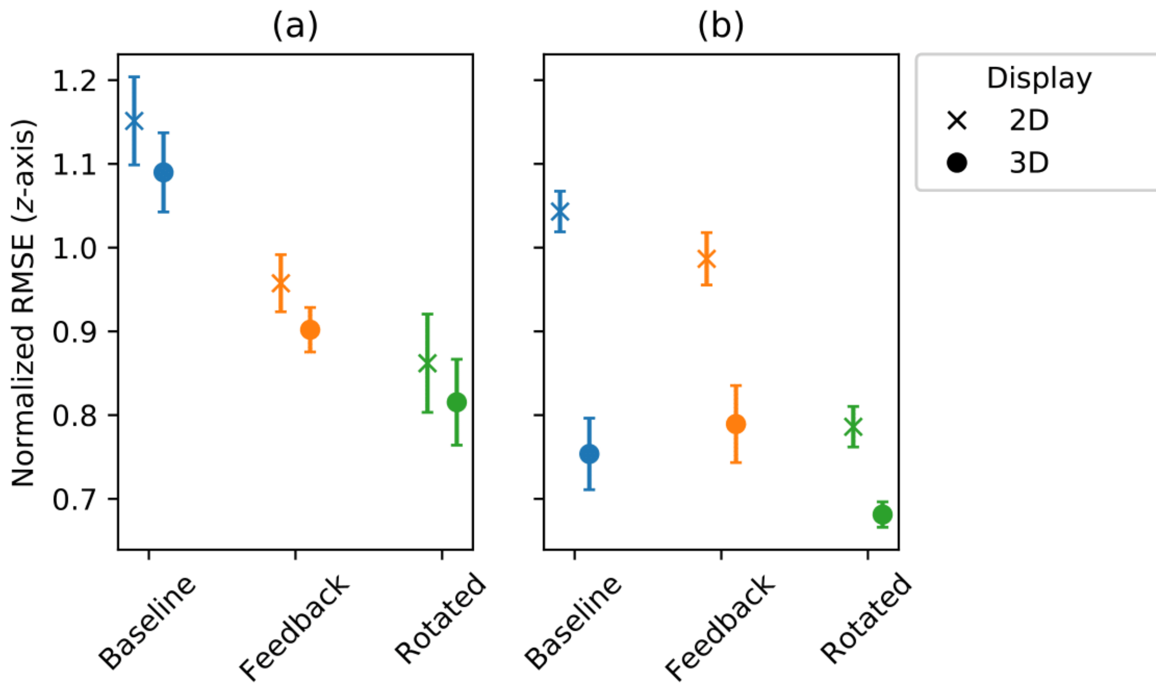


Figure 3: The normalized RMSE of the z-axis across subjects who a) did not start in the feedback condition and b) did start in the feedback condition. Error bars are the standard error of the mean.

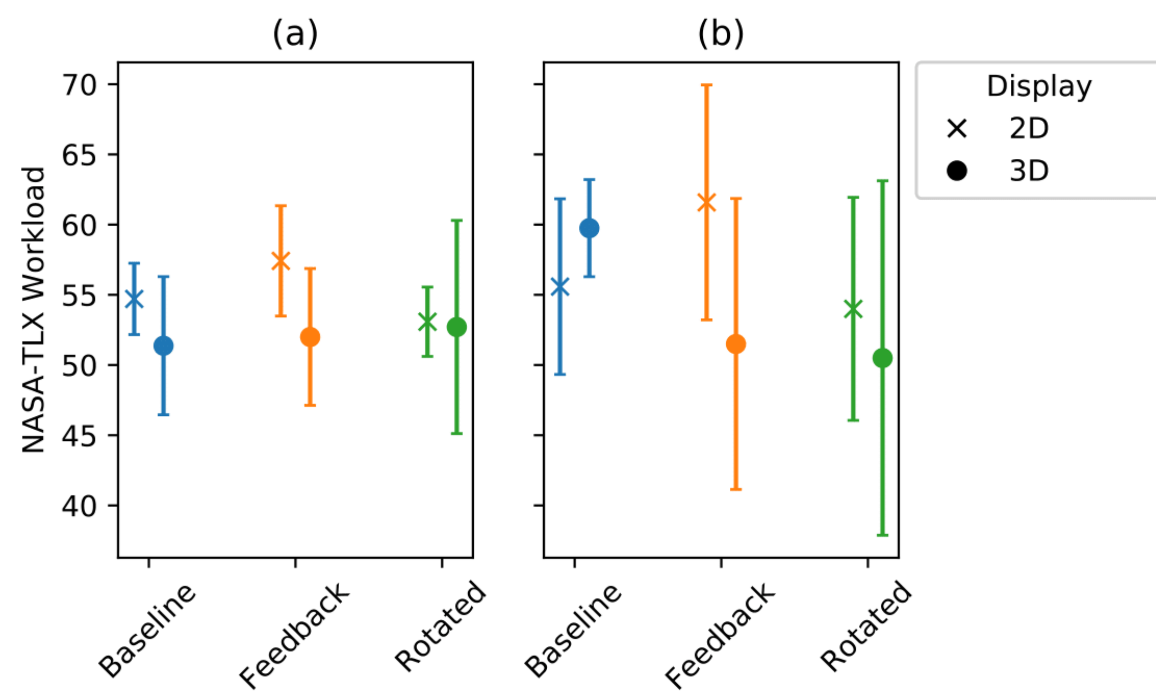


Figure 4: The mean NASA-TLX workload score across subjects who a) did not start in the feedback condition and b) did start in the feedback condition. Error bars are the standard error of the mean.

The NASA-TLX was used to measure differences in subjective workload, and the reaction time to the secondary task was used to measure differences in objective workload between design, device, and starting design. There were no significant effects, nor interaction effects, found in the ANOVA for design, device, or starting design for the NASA-TLX measurements, see Figure 4. There was a significant effect of design ($F(2, 36)=7.93$, $p<0.0014$) for the reaction time to the secondary task, though the magnitude of this effect was very small between designs and was not significant during Tukey HSD tests. In general, there were no significant effects found for workload measurements.

To summarize these results, there were significant effects found in the z-axis RMSE for design, with participants generally performing the best using the angled design, followed by the CBF design, and performing worst with the baseline design. There were significant effects found for the factors of device and starting design, though these must be interpreted carefully. Subjects that started in the design with concurrent bandwidth feedback performed better than participants that did not. An interesting effect of this exposure is that, after learning the task with CBF, participants continued to perform significantly better in the baseline condition than those participants that did not start in the CBF design.

Significance and impact

Subjects that started in the CBF design and that were wearing the HoloLens appear to have used the CBF to better learn the depth cue presented in the stereoscopic display. These participants continued to perform significantly better than participants who started with the CBF design but without the stereoscopic display when they continued to the baseline design. This indicates that even a brief exposure to the concurrent bandwidth feedback was enough to induce improved performance in the baseline design. Additionally, this also suggests that well trained participants could perform better using the stereoscopic display compared with the traditional display, but that participants who were still learning the task could not take advantage of the additional depth cues provided by the display.

Where might this lead?

This research suggests that early exposure to concurrent bandwidth feedback allows subjects to perform better on 3D tasks, especially when they can make use of a stereoscopic display. This is especially important in training and simulators, where trainers and developers can make use of this fact to improve performance quickly without an increase in workload.

2. How did the fellowship make a difference?

The Link Foundation Fellowship allowed me the freedom to work on the research I found the most exciting for my PhD, while being able to devote myself full time to research. Simply put, this allowed me to make great progress on my research. The Fellowship provided me with the ability to get my work to the point where it was accepted to the AIAA SciTech conference in San Diego, CA in January 2019. The fellowship also provided me the time and resources to successfully complete my qualifying examination.

3. Future Plans

In the near term, I am still working on my PhD. Given the results of this experiment, we plan to continue to evaluate human performance in our next experiment using NASA's Robotic Onboard Trainer (ROBoT). The ROBoT simulator is the same training tool that NASA astronauts use to train to use the robotic arm on the space station and provides a high-fidelity simulation with actual operational use. Given the interesting results we found with respect to training with concurrent bandwidth feedback, our next experiment will investigate the difference in performance between training with and without the feedback. Retention trials will investigate if improved performance can be sustained after the feedback has been removed.

4. Publications, Presentations, and Other Outputs

[1] Karasinski, J. and Robinson, S. K., "Evaluating Augmented Reality in a Three-Axis Manual Tracking Task," AIAA SciTech Conference, San Diego, CA, 7--11 January, 2019.

[2] Karasinski, J., "Qualifying Examination Presentation," 29 August, 2018.

5. References

[3] Kim, Won S., et al. "Quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking tasks." *IEEE Transactions on Systems, Man, and Cybernetics* 17.1 (1987): 61-72.

[4] Salmoni, Alan W., Richard A. Schmidt, and Charles B. Walter. "Knowledge of results and motor learning: a review and critical reappraisal." *Psychological bulletin* 95.3 (1984): 355.

[5] Sigrist, Roland, et al. "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review." *Psychonomic bulletin & review* 20.1 (2013): 21-53.

[6] Karasinski, John, et al. "Real-Time Performance Feedback in a Manually-Controlled Spacecraft Inspection Task." *AIAA Modeling and Simulation Technologies Conference*. 2017.

[7] Hart, Sandra G., and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." *Advances in psychology*. Vol. 52. North-Holland, 1988. 139-183.