Experimental Evaluation of Augmented Reality in Object Assembly Task

Arthur Tang, Charles Owen, Frank Biocca, Weimin Mou
M.I.N.D. Labs, M.E.T. Lab, Michigan State University.
tangkwo1@msu.edu, cbowen@cse.msu.edu, biocca@msu.edu, mouw@msu.edu

Abstract

This study evaluated the effectiveness of spatially overlaid instructions using augmented reality (AR) in an assembly task comparing with other traditional media. Results indicate that overlaying 3D instructions on the workspace reduce error rate by 82%, particularly cumulative errors. Measurement of mental effort also suggests some of the mental workload is offloaded to the computer.

1. Introduction

The purpose of this study is to explore the effectiveness of AR as an instructional medium. This study has sought to provide three contributions for our understanding of computer-human interaction in AR environment: (1) determine if the use of spatially overlaid information can improve human performance in assembly tasks comparing to other media, (2) provide theoretical basis for how AR interfaces might provide cognitive support and augmentation, and (3) determine potential weakness in current AR interface design methodologies.

2. Methodology

Three hypotheses were generated. **H1:** Spatially overlaid media will significantly reduce the amount of time to complete an assembly task. **H2:** Spatially overlaid media will significantly reduce the amount of errors during object assembly. **H3:** Spatially overlaid media will significantly reduce the cognitive load of an assembly task.

The experiment uses a between subject design. There was one independent variable, type of instructional media used, with four levels of treatment: (1) printed manual, (2) CAI on LCD monitor, (3) CAI on see-through HMD, and (4) spatially registered AR. The dependent variables included time of completion, error rates, and perceived mental workload. 75 participants from an undergraduate class volunteered to participate in the study. Subjects were required to complete an assembly task consisted of 56 steps using the specific medium as per the treatment. An assembly task made up of Duplo was used in the experiment to minimize bias towards population with expertise in certain knowledge related to a specific domain, and for task generalization so the result is applicable to general assembly tasks rather than assembly tasks in specific domains. To facilitate hands-free task engaged operation, subjects in treatment 2, 3, and 4 used voice command to control the instructions. A human agent interpreted the voice commands and controlled the instruction accordingly (with reaction time within a second) to ensure maximum accuracy on the voice recognition task. Three measurements were collected: completion time, number of errors in the final assembly, and a subjective measurement of mental workload using the NASA Task Load Index (NASA TLX) [1]. Two classes of errors were defined. Dependent error is an error related to another error made previously, and independent error is an isolated error that does not relate to a previous step.

This experiment was designed to compare the effectiveness of spatially overlaid information with traditional media. We controlled some variables between different treatments that affect task performance, but not within the objectives of this study. In treatment 3 and 4, instructions were presented to the subjects through a see-through HMD. Performance is expected to be reduced due to a reduced luminosity, field-of-view, and discomfort caused by the HMD. To control these factors, subjects in all treatments were required to wear the HMD during operation so that these variables remain constant among treatments. In treatment 4, subjects were required to perform a display calibration procedure that is considered to be challenging for untrained user, and can induce fatigue and mental workload factors that affect task performance. To control these factors, participants in all treatment were required to perform the calibration procedure so that these variables remain constant among treatments.

3. Results

Table 3.1 illustrates the mean time of completion, mean number of errors, and mean score of the NASA TLX rating for each treatment. The descriptive statistics reveal that treatment 4 (AR) has significantly lower error rates. They also indicate that the majority of errors in treatment 4 are independent errors, whereas treatment 1, 2 and 3 exhibit a majority proportion of dependent errors. It shows that subjects in treatment 1 have the highest mental workload, whereas subjects in treatment 4 have the lowest mental workload. It also shows that treatment 2, 3, and 4 have a shorter completion time then treatment 1.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

Treatment	1	2	3	4
Completion time (minutes)	14.4	11.4	11.1	10.9
Total Error	9.4	8.4	9.5	1.6
Dependent Error	7.2	6.2	7.1	0.2
Independent Err.	2.2	2.3	2.4	1.4
NASA TLX	13.3/20	12.2/20	11.0/20	10.0/20

3.1 Effect of Medium on Completion Time

A one-way Analysis of Variance (ANOVA) was conducted on the effect of medium on completion time. The effect of completion time depending on medium is statistically significant, $F(3, 71) = 3.75$, $p = .015$. Post-hoc comparisons further show that there are statistically significant effects between treatment 1 and 4 ($p = .020$). The effects between treatment 1 and 2 and treatment 1 and 3 trend toward significance ($p = .090$). But there is no significant effect between treatment 2 and 3 ($p = 1.000$), treatment 2 and 4 ($p = 1.000$), and treatment 3 and 4 ($p = 1.000$). The results of the ANOVA analyses show that treatment 2, 3 and 4 have a significant shorter completion time comparing with treatment condition 1, but there is no statistically significant effect between treatment 2, 3 and 4. Therefore, hypothesis H1 is not supported.

3.2 Effect of Medium on Accuracy

Effect of Total Error. A one-way ANOVA was conducted on the effect of medium on total error. The effect is statistically significant, $F(3, 71) = 4.41$, $p = .007$. Post-hoc comparisons further show that there are statistical significant effects between treatment 1 and 4 ($p = .019$) and treatment 3 and 4 ($p = .012$). The effect between treatment 2 and 4 trends toward significance ($p = .073$). But there is no significant effect between treatment 1 and 2 ($p = 1.000$), treatment 1 and 3 ($p = 1.000$), and treatment 2 and 3 ($p = 1.000$). The results of the ANOVA analyses show that treatment 4 has significantly fewer total errors comparing with treatment 1, 2 and 3. However, there is no statistically significant effect between treatment 1, 2 and 3.

Effect of Dependent Error. A one-way ANOVA was conducted on the effect of medium on dependent error. The effect is not statistically significant, $F(3, 71) = 4.68$, $p = .005$. Post-hoc comparisons further show that there are statistical significant effects between treatment 1 and 4 ($p = .017$) and treatments 3 and 4 ($p = .009$). The effect between treatment 2 and 4 trends toward significance ($p = .070$). But there is no significant effect between treatment 1 and 2 ($p = 1.000$), treatment 1 and 3 ($p = 1.000$), and treatment 2 and 3 ($p = 1.000$). The results of the ANOVA analyses show that treatment 4 has significantly fewer dependent errors comparing with treatment 1, 2 and 3. However, there is no statistically significant effect between treatment 1, 2 and 3.

Effect of Independent Error. A one-way ANOVA was conducted on the effect of instructional medium on independent error. The effect of independent error depending on treatments is not statistically significant, $F(3, 71) = .967$, $p = .413$.

Subjects in treatment 4 had a significantly higher accuracy and hypothesis H2 is supported. The number of dependent errors in treatment 4 was much lower than the other 3 treatments. This may be due to the fact that spatially overlaid medium eliminates some dependency among procedural steps. It was observed that the rate of subjects correcting mistakes made in previous steps in treatment 4 was much lower than other 3 treatments. This observation is coherent with a phenomenon called attention tunneling: attention is focused on the area cued at the cost of other areas. Yeh, et al. reported that "cueing aided the target detection task for expected targets but drew attention away from the presence of unexpected targets in the environment" [2]. Attention tunneling can reduce user performance and generate hazardous scenarios. Yeh et al. recommended designer of such cueing systems more carefully evaluate operator reliance on automation.

3.3 Effect of Medium on Mental Workload

A one-way ANOVA was conducted on the effect of medium on the NASA TLX rating of mental workload. The effect was statistically significant, $F(3, 71) = 6.26$, $p = .001$. Therefore, hypothesis H3 is supported.

Determining position and orientation from pictorial diagram drawn from the author's perspective is a primitively hard task. Human beings tend to approximate position and orientation using fixations and landmarks already in place. By spatially overlaying instructions, AR reduces the cognitive workload to determine position and orientation at the workspace from the instructional media.

4. Conclusions

The results of this study support that the feature of overlaying and registering information on the workspace in a spatially meaningful way in AR improves human performance and relieves some of the user mental workload. However, the limitations in the current display and tracking technologies are the biggest obstacles preventing AR from being realistic in practical uses. A full length article about this study is available at "http://io.midlab.msu.edu/ARAssembly.pdf".

5. References

1. Hart, S.. *Background Description and Application of the NASA Task Load Index (TLX).* in *Department of Defense Human Engineering Technical Advisory Group Workshop on Workload*. 1987. Newport, RI.
2. Yeh, M. and Wickens, C. (2000). *Attention and Trust Biases in the Design of Augmented Reality Displays*. Report ARL-00-3/FED-LAB-00-1, Aviation Research Lab, University of Illinois, Urbana-Champaign: Savoy, IL.