Using an Industry-ready AR HMD on a Real Maintenance Task: AR Benefits Performance on Certain Task Steps More Than Others

Andrew Pringle*‡, Abraham G. Campbell‡, Stefanie Hutka†, Alberto Torrasso†, Colin Couper†, Fabian Strunden†, Jan Bajana†, Kamil Jastząb†, Ralph Croly†, Rob Quigley†, Ross McKiernan†, Paul Sweeney† & Mark T. Keane‡

University College Dublin‡, DAQRI† & Trinity College Dublin*

Figure 1: Left and right panels display images of a 3D virtual motor as seen through the DAQRI Smart Helmet HMD (centre panel, alongside the real physical motor). The left panel shows the work instruction for a sub-task, task step 4; requiring a user to insert a "feeler gauge" into a cavity (air gap) on the motor to measure its thickness while the right panel shows the instructions for a sub-task, task step 9, requiring a user to mount a nut on the motor shaft in order to allow them to turn it using a torque wrench.

ABSTRACT

This paper presents a novel evaluation of an industry-ready HMD for delivering AR work instructions in a real-life, industrial procedure for novice users. A user study was performed to examine the potential benefits and limitations of a dynamic 3D virtual model and AR text instructions, delivered through an optical see through HMD, for training users in a new industry procedure (i.e., Yaw Motor Servicing of a wind turbine). Measures of task accuracy and completion time were used to evaluate the performance of one group of mechanical engineering students performing this procedure for the first time guided by AR compared to a second group performing it using a tablet-delivered instruction manual. Results showed AR improved accuracy but not speed of task completion. AR significantly increased accuracy on one specific task-step in the procedure, namely measurement of a thin air gap (see figure 1, left panel), but also showed limitations with other task-steps not benefitting or even being slowed down by AR (see figure 1, right panel). Findings speak to the importance of incorporating an analysis at the level of individual task steps in order to fully evaluate AR work instructions.

Keywords: Augmented Reality, providing instructions, maintenance, workpiece, head-mounted displays.

Index Terms: H.5.2 [Information interfaces and Presentation]: Multimedia Information Systems- Artificial, augmented and virtual realities; H 5.2 [Information interfaces and Presentation]: User Interfaces- Training, help and documentation

* email: pringlea@tcd.ie

1 INTRODUCTION

Industries currently face a problem with training employees in industrial maintenance and assembly procedures [3, 6]. One proposed solution to tackle this is to deliver work instructions for these procedures using augmented reality [1, 3, 6]. Prior work has evaluated the potential benefits of different AR technologies, including head-mounted displays (HMDs), Tablet AR and in-situ projection, compared to current non-AR methods of work instruction delivery, such as 2D paper or tablet-based instruction manuals [2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15]. While some of these evaluations have been conducted on industry training tasks, we know of no current work that has evaluated the benefits of an industry-ready HMD for delivering AR on a real industry task. So, the current paper addresses this gap in the literature by conducting a user study to examine the potential benefits and limitations of AR work instructions delivered using an industry-ready HMD, the DAQRI Smart Helmet, on a real-world industrial task; namely, the maintenance procedure for a wind turbine Yaw motor.

Further, the majority of evaluations of AR work instructions have only examined the benefits of AR at the level of an entire procedure involving multiple task steps. Only a few prior evaluations have examined the potential benefits of AR at the more fine-grained level of individual task steps, with different steps involving different operations (e.g., measuring air gaps, turning bolts). The results of this work have shown that some task-steps are better supported by AR than others [6, 13]. In designing AR work instructions for industrial settings, it is critical to identify which task-steps are likely to benefit most from AR in order to help determine where best to apply AR to improve training outcomes. The present paper adds to the literature by incorporating an analysis evaluating AR instructions at the level of individual task steps.

2 INDUSTRY PROCEDURE: YAW MOTOR SERVICE

Wind turbine Yaw Motors must be serviced annually in the form of a maintenance procedure conducted by a technician with a background in mechanical engineering. Currently the procedure is performed within industry (Siemens) by one mechanical engineer guided by a pdf instruction manual delivered on an iPad tablet or on paper. The instruction manual consists of eleven task-steps (see supplementary material for the full set of eleven task steps).

3 INDUSTRY READY AR HMD

The HMD used to deliver AR in the current user study is designed specifically for augmenting worker capability in industry and ergonomically suited for industrial environments, having integration with common workplace software environments and tools, being portable and designed to function outdoors.

3.1 DAQRI Smart Helmet Developer Edition (DSH DE)

The DSH DE (Figure 1, central panel) is a head mounted display (HMD) for delivering high-resolution 3D AR content. The field of view is 44 degrees per eye and the weight of the DSH DE with the battery installed is 1500g. The DSH DE has a hands-free user interface with a cursor termed “the reticle”, controlled by the user moving their head, to navigate through the interface and with the user “dwelling” on components for a few seconds to select them. This method of interface control is known as “gaze and dwell”.

3.2 AR application

We applied an AR application to the maintenance procedure. This application, “Yaw Brake Service” was developed by DAQRI from the original Siemens pdf instruction manual for the procedure. Two AR components of this application were delivered to participants in the current user study; (1) text instructions registered in body space that move with user’s head movements and (2) a 3D virtual model of the Yaw Motor registered in world space that remains in a fixed location in the physical environment and around which the user can move. Examples of AR components (1) and (2) are shown in figure 2.

3.3 AR-instruction during Maintenance Procedure

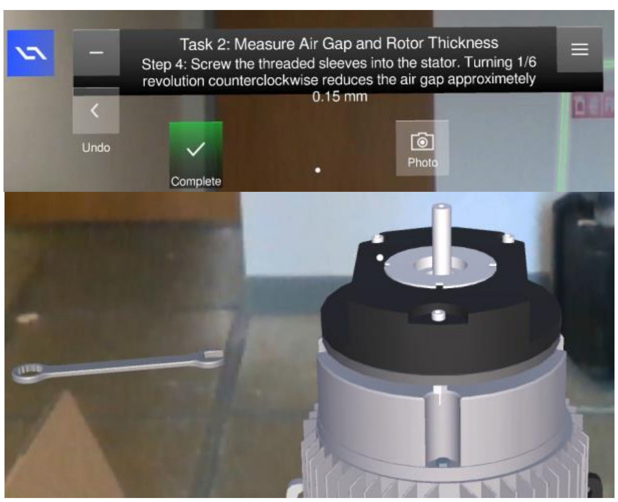

Figure 2: Shows AR text instructions (top panel) and the dynamic 3D virtual model (bottom panel) as seen through the DAQRI Smart Helmet for task step 6 “screw sleeves”.

The maintenance procedure using the “Yaw Brake Service” AR application proceeds as follows. The AR application is run from the HMD operating system menu. The user then scans an origin marker to register the location of the 3D virtual model in world-space. At this point the user must move the 3D virtual model of the Yaw Motor into their field of view using the reticle. In the user study this action was performed by the first author (Pringle) so as the 3D model was positioned in approximately the same location for all participants. The participant then puts on the HMD and using the reticle progresses to the first step of the procedure by gazing and dwelling on a green start button. In each step, users are presented with both text instructions and a 3D virtual model of the Yaw motor (as shown in figure 2). On the 3D virtual model dynamic animations are displayed, for example showing the direction that bolts should be turned, locations where measurements should be made and tools to perform these actions. Once participants have followed all instructions in a task-step they proceed to the next step by gazing and dwelling on the green “complete” button. In this way, participants proceed through the 11 steps of the maintenance procedure until they see a green text notification stating, “assembly complete”.

4 USER STUDY

A user study was designed and conducted to compare performance on the maintenance procedure between (i) a group of participants who performed it while wearing the optical see through AR HMD, guided by AR work instructions (AR-instruction) and (ii) a group of participants who performed it guided by the current industry instruction manual for this procedure (tablet-instruction, control condition).

Mechanical engineering students were recruited for the user study rather than students from other disciplines (e.g. computer science) because (a) they were the student group that best represented potential future end-users of AR work instructions for this procedure and (b) because the majority (31 out of 36) reported previous experience repairing or servicing mechanical systems, suggesting some experience with the tools and operations required on the maintenance procedure. It was decided to recruit mechanical engineering students rather than qualified practicing mechanical engineers for ethical reasons specifically, it was easier to get ethical clearance to include them in the study. However, students evidently do not represent current real-world users of AR for this procedure (see limitations section 6.1).

4.1 Methods

4.1.1 Participants

Thirty-six mechanical engineering students (Age: M= 22.44, SD = 5.50, all male) were recruited using opportunity sampling from the following Universities in Ireland; University College Dublin, Dundalk Institute of Technology, Trinity College Dublin and Dublin Institute of Technology. The majority of students in classes where we advertised for participants were male which perhaps explains the lack of females volunteering to take part. Thirty-two participants were native English speakers and four were non-native English speakers. Due to the language barrier, non-native speakers reported that they could not fully understand the instructions for the procedure and hence the four non-native speakers, N =2 in each instruction condition were excluded from all subsequent analyses. The University College Dublin Ethics Review Board approved this study. Informed written consent was obtained from all participants prior to participation. The information sheet and consent form informed participants that video recordings would be made of their hands working on the motor while they completed the procedure and that these would be used for research purposes only and not shared externally.

4.1.2 Design

Participants were randomly assigned to two groups; the AR-instruction condition (N = 18) and the tablet- instruction condition

(N = 18). An important component of the design was that we treated task step as a repeated measure, with participants across both groups performing every task in the same order. The order of task steps remained constant across participants to ensure that the task made logical sense (e.g., they would only be instructed to adjust an air gap after first measuring it to check it needed adjusting). The full user study thus had a mixed between-subject, repeated measures design enabling us to compare task accuracy and completion time metrics between work instruction conditions (AR and tablet-instructed) at the level of each individual task step.

4.1.3 Materials

A Yaw Motor on loan from Siemens was mounted to a workbench using four bolts (see left panel of figure 3). The set of tools that users would be required to use during the procedure were positioned on the left side of the workbench and consisted of a torque wrench, hook wrench, two wrenches, two screwdrivers, an allen key, a sliding gauge, a feeler gauge and a special nut. In the AR-instruction condition, a marker for tracking AR content was placed on the right side of the workbench (right of motor in figure 3, left panel). In the tablet-instruction condition, the iPad with AR instructions was instead positioned on the right side of the workbench (see figure 3, right panel).

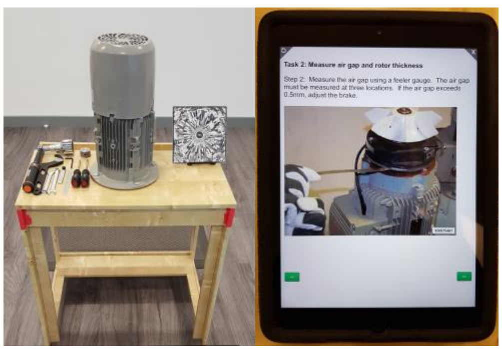

Figure 3: Yaw Motor mounted on bench (left panel), Tablet instruction manual delivered on an iPad (right panel)

4.1.4 Control condition

The current pdf-based industry instruction manual was modified slightly for use as the control condition in the current user study. The modifications consisted of implementing each task step from the pdf manual as a separate screen delivered to participants using Qualtrics software (see right panel figure 3 for an example). Each screen displayed both text instructions for each step, along with photographs of the procedure to be performed. Both text instructions and photographs were taken directly from the industry pdf instruction manual but increased in size in order to optimise clarity. Green buttons were added below instructions to allow participants to easily move back and forth between instruction screens. By ensuring each screen showed only one task-step we made it easier to identify exactly which one participants were working on which was critical to enable the scoring of task completion time. We believe these modifications are valid because they should maximize the clarity of instructions and be easy for participants to navigate through, while not changing the content of the current instruction manual used by industry engineers to perform the procedure.

4.1.5 Pilot testing of work instructions

Before running the user study, it was important to develop a detailed understanding of where errors could be made during the procedure. One of the authors (Torrasso), a qualified engineer, worked with Siemens and observed the procedure being performed in an industry training centre. This author briefed the experimenter (Pringle) running the study, demonstrating the correct tasks to perform on the motor (e.g., the correct positions where measurements should be made) at each of the task-steps. From this, a coding scheme that comprehensively categorized errors was developed, and used to score Yaw Motor servicing task accuracy (see section 4.2.1).

On four of the eleven task steps, instructions did not exactly correspond to the physical procedure that must be performed on the motor. For example, on task step 10, instructions tell the user to use a hook wrench to adjust the shaft of the motor when in practice, the motor shaft is actually adjusted by turning bolts on the top. Due to these discrepancies it was decided that the experimenter would verbally deliver assists on these four task steps to allow participants to perform the core tasks required in each task step (e.g., turning the motor shaft). These assists were identifiably worded for all participants in both instruction-conditions in order to ensure they did not bias the performance of one over another. Pilot testing with two male computer science students helped determine the user study length; 1 hour for the AR-instruction condition and 50 minutes for the tablet-instruction condition. Participants were tested individually with the experimenter present. The user study phases are now described.

4.1.6 Technology familiarization phase

All participants had never previously used the AR Yaw brake service application and only one had previously used the HMD. Participants in the AR-instruction condition used two AR applications to familiarize themselves with the HMD and AR. They were instructed on how to fit the HMD on their head and then guided through the use of the reticle to open a first AR application involving interaction with a 3D virtual model of a turbine and a second application which mirrored the combination of text instructions, 3D virtual model and gaze and dwell interface controls in the Yaw brake service work instructions. In the tablet-instruction condition, participants were familiarized with the iPad interface, and the use of the green forward and back controls to navigate through instructions.

4.1.7 Maintenance task procedure

Participants were asked to perform a motor servicing task on the motor following instructions in the HMD/iPad application and using tools on the workbench. They were asked to work as quickly and accurately as possible, to inspect both the visuals and written HMD/iPad instructions provided at each step carefully, and to ask for assistance if unsure at any point. The experimenter then opened the Yaw Brake service application, tracked the marker and ensured that text instructions and the 3D virtual motor were visible on the right of the physical motor. Participants put on the HMD/opened the iPad app and selected the green start button within either the AR or iPad application and started the procedure following instructions until completion of all task-steps.

4.1.8 Post-task data collection

Following completion of the procedure, participants completed four questionnaires: (1) NASA Task load index (TLX), (2) System Usability Scale, (3) Confidence in successful completion of Yaw motor servicing and (4) Social acceptance of the work instruction technology. For brevity, analyses of self-report measure data are not included in this paper but briefly, results showed that AR instruction reduced user task workload (mental but not physical load), increased usability and learnability, confidence and social acceptance relative to tablet-instruction.

4.2 Objective Measures of Task Performance

4.2.1 Task accuracy

The coding scheme to record performance on the task was used to measure task accuracy. It includes specific information on the text instructions delivered to participants on this task step, the possible errors that could be made and a score for each error (see supplementary material for details). To illustrate, task step 4 instructs participants to "measure air gap using a feeler gauge. The air gap must be measured at three locations. If the air gap exceeds 0.5mm adjust the brake". A photo of the correct procedure for this step is shown in the top left panel of figure 4 while the other three panels show possible errors on this task step. Possible errors that could be made on this task step involve (a) using the wrong tool to measure the air gap (e.g., the sliding gauge), (b) measuring the incorrect air gap, (c) using an incorrect feeler gauge measure (i.e. not 0.5 or 0.4mm), (d) failing to measure the air gap at three locations along the circumference of the motor and (e) tightening or loosening the bolts on top of the motor to alter the air gap. These errors matter because failing to measure the correct air gap with the correct tool and measure could result in the decision not to adjust the brake when in fact it does need to be adjusted. It is important to measure the air gap at three locations because it is not necessarily uniform across the entire motor circumference. Finally, loosening or tightening the bolts on top in this step when not instructed to do so, could lead to bolts being over-loosened on the following step, step 5. Each of these errors received a score of 1.

Some errors on other task steps were deemed less serious. For example on task-step 2, participants were instructed to "remove the fan using two screwdrivers or other suitable tools". The fan could be removed using one screwdriver or with the hands, so participants could still successfully perform the core task even if they deviated from the standard procedure. Such errors received a score of 0.5. Some participants missed entire task steps. In such instances they received the maximum error score for the task steps missed. Errors made at each task step are summed (e.g., for the feeler gauge step this would be Σ [a, b, c, d, e]) to give a total error score for that step. The total error score for the entire procedure was the sum of errors across all eleven task steps.

4.2.2 Task completion time

The time taken to complete each task step was extracted from video recordings of participants completing the procedure. For the tablet-instruction condition, completion time for each task step was calculated from the point at which the green forward button was pressed to begin the current task step to when the same button was pressed to progress to the next task-step. For the AR-instruction condition, completion time was calculated from the point when an audible bleep sound could be heard signaling the appearance of AR content for a new task step to when a second bleep sound was heard signaling progress to the next task step. Total completion time for the entire procedure was the sum of the completion time across all eleven task steps.

4.3 Hypotheses

Based on related user studies that have compared task performance between AR technology and paper/tablet instruction manuals [2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15] and those that examined an interaction between the type of work instruction delivered and task-step performed [6, 13] we formulated the following hypotheses: fewer errors will be made on the procedure in the AR instruction compared to tablet-instruction condition (hypothesis 1); completion time will be faster in the AR compared to tablet-instruction condition (hypothesis 2);

Figure 4. The far-left panel shows the correct procedure for performing task step 4 "measure air gap". The second from the left panel shows error (a) use of the wrong tool, the second from the right panel shows error (b) measuring incorrect air gap and the far-right panel shows error (c) using an incorrect feeler gauge measure, 0.04 mm in this case.

there will be an interaction between instruction condition and task step, with the group following AR instruction making fewer errors (hypothesis 3) and completing the procedure faster (hypothesis 4) compared to the group following tablet-instruction on some but not all task steps.

5 RESULTS

Differences between the group guided by AR-instruction and the group guided by tablet-instruction were examined on two metrics of task performance: (1) task accuracy, defined as a lack of errors, and (2) speed of task completion. For each metric, group differences were examined both at the level of the entire task and for each of the eleven task-steps. We then report qualitative information on specific types of errors made by the tablet-instruction group that are made less often by the AR-instruction group to reveal more information on why AR benefitted or hindered users. Finally, we include some participant quotes.

5.1 Task accuracy analysis

A 2 (Instruction condition: AR-instruction, Tablet-instruction) X 11 (Task-step: 11 distinct steps) mixed ANOVA was run on number of errors made, with instruction-condition entered as the between-subjects factor and task step as the repeated measures factor. Instruction condition revealed a significant main effect on the number of errors made ($F_{(1, 30)} = 6.13$, $p = .02$). A lower mean number of errors were made for AR-instruction ($M = 0.38$, $SD = 0.21$) compared to tablet-instruction ($M = 0.57$, $SD = 0.23$), representing a 33% improvement from AR. Task step also revealed a significant main effect on the number of errors made ($F_{(10, 300)} = 10.97$, $p < .001$) showing more errors were made on certain task steps than others.

There was a significant Instruction condition X Task-Step interaction ($F_{(10, 300)} = 4.03$, $p = .003$), indicating that the number of errors made at the level of individual task steps differed between AR and tablet-instruction. This interaction was explored further using post-hoc independent *t*-tests with a Bonferroni correction ($\alpha = .0045$) to examine group differences in the number of errors made at each of the eleven task steps. Means and their 95% confidence intervals at each task step for each instruction condition are shown in figure 5. *T*-tests revealed that a lower mean number of errors were made on task step 4 "measure the air gap using a feeler gauge" in the AR-instruction ($M = 0.50$, $SD = 0.73$) compared to the tablet-instruction condition ($M = 1.56$, $SD = 1.09$, $t\ (30) = 3.23$, $p = .002$), representing a 68% improvement from AR. There were no significant differences in the number of errors made between AR and tablet-instruction on any of the other ten task steps.

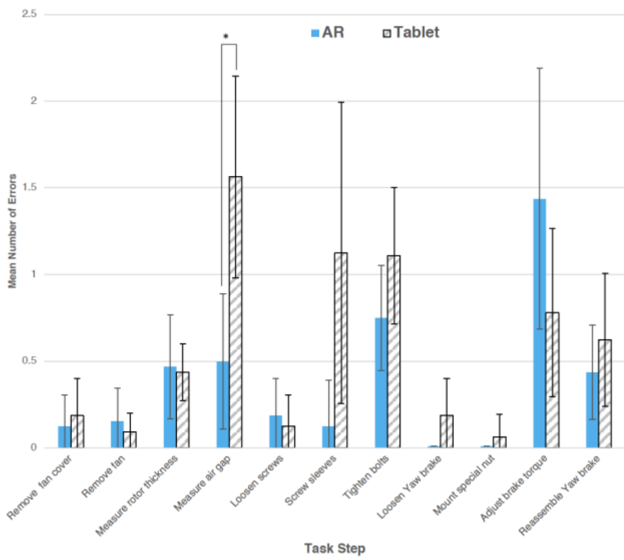

Figure 5. Mean number of errors made by participants, and their associated 95% confidence intervals, at each task-step in each instruction condition (AR-instruction, tablet-instruction). The star (*) identifies where differences between conditions were statistically significant after applying the Bonferroni correction ($p < .0045$).

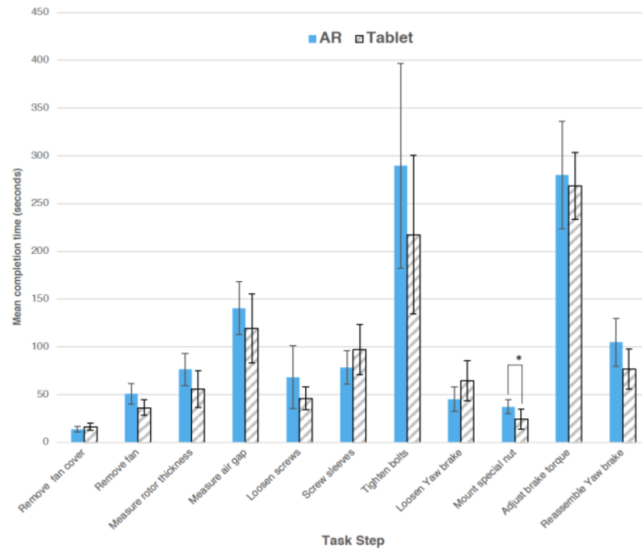

Figure 6. Mean time taken for participants to complete each task step, and their associated 95% confidence intervals, in each instruction condition (AR-instruction, tablet-instruction). The star (*) identifies where differences between conditions were statistically significant after applying the Bonferroni correction ($p < .0045$).

5.2 Task completion time

Task completion time data were first pre-processed, by imputing condition mean completion times for two participants (one in each condition) who missed a task step and a $\log_{10}$ transform was applied to the raw time data to correct for positive skew.

A 2 (Instruction condition: AR-instruction, Tablet-instruction) X 11 (Task-step: 11 distinct steps) mixed ANOVA was run on $\log_{10}$ transformed task completion time, with instruction-condition entered as the between-subjects factor and task step, the repeated measures factor. Instruction condition revealed a marginally significant main effect on task completion time ($F_{(1, 30)} = 3.74$, $p = .06$). The mean time to complete a task step on the Yaw Motor Servicing task was slower for the AR-instruction condition (M= 108 seconds, SD = 33 seconds) compared to the tablet-instruction condition (M= 93 seconds, SD = 27 seconds), representing a 16% slowdown for AR. Task step revealed a significant main effect on task completion time ($F_{(10, 300)} = 2.95$, $p < .001$) showing participants took longer completing certain task steps than others.

There was a significant Instruction condition X Task step interaction ($F_{(10, 300)} = 4.03$, $p = .002$), indicating that time to complete individual task steps differed between AR and tablet-instruction. This interaction was explored further using post-hoc independent t-tests with a Bonferroni correction ($\alpha = .0045$) to examine differences in the number of errors made at each of the eleven task steps. Means and their 95% confidence intervals at each task step for each instruction condition are shown in figure 6. T-tests revealed that task step 9 "mount special nut" was completed more slowly by the group using AR (M= 37 seconds, SD = 14 seconds) compared to tablet-instruction (M= 24 seconds, SD = 19 seconds, t (30)= 1.56, $p = .002$), representing a 54% slowdown from AR. There were no significant differences in task completion time between AR and tablet-instruction on any of the other ten task steps.

5.3 Qualitative analyses and Quotes

Qualitative analyses focus on the type of errors made on task step 4 "measure the air gap with a feeler gauge" given this was the task step on which AR reduced errors relative to tablet-instruction. The types of errors made by more than one participant on this task step are detailed in section 4.2.1 and figure 4. AR-instruction had the biggest impact on reducing errors resulting from (d) failing to measure the air gap at three locations along the circumference of the motor (3 participants using AR made this error vs. 13 participants using tablet-instruction). Using AR-instruction, zero participants (b) measured the incorrect air gap or (a) used the wrong tool while 5 participants made error (b) and 4 made error (a) using tablet-instruction. Three participants (c) used an incorrect feeler gauge measure while following AR-instruction compared to 5 following tablet-instruction.

Participants using the AR HMD reported higher physical load relative to those using the tablet, possibly due to its weight. One stated "the helmet got a bit heavy at the end" and another that "the helmet was not comfortable to wear with glasses due to the helmet pressing on them". Evidently the weight of the HMD was a drawback even for the short (≈ 20 min) maintenance procedure.

6 CONCLUSION AND DISCUSSION

This paper presents results of the first known user study to evaluate the potential benefits and limitations of AR work instructions delivered using an industry-ready HMD on a real-world industrial task; Yaw Motor Servicing. Relative to a traditional method of delivering work instructions using a tablet-based instruction manual, HMD delivered AR benefitted users by reducing errors made (improving accuracy) but did not enable users to complete the procedure more quickly.

Users following AR work instructions made 33% less errors on the maintenance procedure compared to those following tablet-instruction (supporting hypothesis 1) and mirroring benefits reported in prior work using prototype HMDs or custom designed in-situ systems for delivering AR [2, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15]. However, the study also revealed limits on the benefits of AR by comparison to the current tablet-based manual; namely, that it slows users down, by 16% (failing to support hypothesis 2). While findings based on the measure of task completion time were not as statistically robust as those on task accuracy, there was no evidence that AR improved speed of task completion. This finding contrasts with the prior reported benefits of AR in reducing task completion time [7, 8, 11] although some prior work has reported that AR delivered via an HMD can slow performance [5]. It is important to discuss this finding alongside the benefits of AR HMD work instructions since it may impact on industry adoption of AR.

A second key aim of the paper was to focus on the types of task, defined as task-steps within the industry procedure, that are best supported by AR work instructions. Findings support hypothesis 3, that AR-instruction only results in less task errors relative to tablet-based instruction on certain task steps. Specifically, AR-instruction resulted in a statistically significant 68% improvement in performance on task step 4 requiring measurement of a thin air gap (see figure 1, left panel). Qualitative analysis revealed that AR helped users realize they should measure the air gap in three locations along the circumference of the Yaw motor, helped them find the correct air gap and identify the correct tool to measure it. Prior user studies focusing on the individual task steps benefitting from AR-instruction found similar findings, with AR assisting users in performing exact placement tasks [6, 13].

The findings on task completion time showed the converse, that AR slowed performance on some task steps more than others relative to the tablet instruction manual. Specifically, completion time on step 9 requiring mounting a special nut on the Yaw motor (see figure 1, right panel) was slowed by 54% using AR. The reasons for this are less clear but could reflect a discrepancy between AR instructions, that showed screwing the nut onto the motor, and the physical action a user had to perform that involved fitting the nut into a grove in the motor; an action that would be hindered by screwing. This point is returned to in the limitations (section 6.1). Clearly, findings from this user study speak to the importance of incorporating an analysis at the level of individual task steps in order to fully evaluate AR work instructions.

6.1 Limitations

Contributions from this work should be considered in the context of several limitations. As mentioned, there were some discrepancies on certain task steps between work instructions, both AR and tablet, and the actual maintenance procedure on the Yaw motor. This is not uncommon within industry. The decision was made to keep these discrepancies in the tablet-instructions and mirror them in AR as they represent real world instructions currently given to trainees. However, some discrepancies could have potentially impacted more negatively on performance in the AR compared to tablet-instruction condition, precisely because the animations in AR instructions could have been misleading. This may explain participants slower performance when using AR to mount the special nut on step 9. A second limitation concerns the quality of the baseline instruction manual against which to evaluate AR with participants in the tablet-instruction condition commenting on a lack of clarity in the photos. Future work should follow the example of Funk *et al.* [4] in using a standardized well-designed instruction manual against which to evaluate AR. A final limitation is that we did not include real world users of the work instructions. Future user studies should evaluate AR technologies with those who perform the procedure in industry.

Acknowledgments

The authors wish to thank DAQRI, Siemens and the University staff and students who supported this work. This work was supported by a Science Foundation Ireland (SFI) Industry Fellowship, Grant #16/IFA/4331.

References

[1] R. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, Vol. 6* (4): 355–385. 1997.

[2] T. P Caudell and David W Mizell. 1992. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *System Sciences, Proceedings of the Twenty-Fifth Hawaii International Conference on*, *Vol. 2. IEEE*: 659–669. 1992.

[3] M. Funk, A. Bächler, L. Bächler, T. Kosch, T. Heidenreich, and A. Schmidt. Working with Augmented Reality? A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace. In: *Proceedings of the 10th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '17* (Rhodes, Greece, 21-23 June), pages 222-229, 2017.

[4] M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt. A benchmark for interactive augmented reality instructions for assembly tasks. In: *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia. MUM '15* (Linz, Austria, 30 Nov-2 Dec), pages, 253–257, 2015.

[5] M. Funk, T. Kosch, and A. Schmidt. Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. In: *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany, Sept 12-16), pages 934–939, 2016.

[6] N. Gavish, T. Gutierrez, S. Webel, J. Rodrıguez, M. Peveri, U. Bockholt, and F. Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, *23(6*): 778–798. 2015.

[7] S. J. Henderson and S.K. Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. *Proceedings of the International Symposium on Mixed and Augmented Reality, ISMAR '09* (Orlando, Florida, 19-22 Oct), pages, 135–144, 2009.

[8] S. J. Henderson and S.K. Feiner. Augmented reality in the psychomotor phase of a procedural task. *Proceedings of the International Symposium on Mixed and Augmented Reality, ISMAR '11* (Basel, Switzerland, Oct. 26-29), pages 191–200, 2011.

[9] L. Hou, X. Wang, L. Bernold, and P.E.D. Love. Using Animated Augmented Reality to Cognitively Guide Assembly. *Journal of Computing in Civil Engineering*, 27 (5): 439-451. 2013.

[10] M. R. Marner, A. Irlitti, and B. H. Thomas. Improving procedural task performance with augmented reality annotations. *Proceedings of the International Symposium on Mixed and Augmented Reality, ISMAR '13* (Adelaide, Australia, 1-4 Oct), pages, 39–48, 2013.

[11] M. Nakanishi, M. Ozeki, T. Akasaka and Y. Okada. Human factor requirements for Applying Augmented reality to manuals in actual work situations. *IEEE International Conference on Systems, Man and Cybernetics* (*ISIC '07*, Montreal, Canada, 7-10 Oct), pages 2650–2655, 2007.

[12] N. Pathomaree and S. Charoenseang, S. Augmented reality for skill transfer in assembly task. *IEEE International Workshop on Robot and Human Interactive Communication, ROMAN '05*, (Nashville, TN, 13-15 Aug), pages 1–25, 2005.

[13] T. Richardson, S. Gilbert, J. Holub, F. Thompson, A. MacAllister, R. Radkowski, E. Winer, P. Davies and S. Terry. Fusing self-reported and sensor data from mixed-reality training. Interservice/Industry Training, Simulation, and Education Conference, I/ITSEC '14 (Orlando, FL, 1-4 Dec), pages 1–12, 2014.

[14] B. Schwerdtfeger and G. Klinker. Supporting order picking with augmented reality. In: *Proceedings of International Symposium on Mixed and Augmented Reality, ISMAR '08*, (Cambridge, UK, 15-18 Sept), pages 39–48, 2008.

[15] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03* (Ft. Lauderdale, FL, 5-10 April), pages 73–80, 2003