

Augmented Instructions: Analysis of Performance and Efficiency of Assembly Tasks

Eleanor Smith [✉]¹, [0000-0001-6869-836X], Gordon Semple², Dorothy Evans¹, Kenneth McRae¹, Paul Blackwell¹, [0000-0001-9740-7971]

¹ Department of Design, Manufacturing and Engineering Management,
University of Strathclyde, Glasgow, UK
eleanor.smith@strath.ac.uk

² Booth Welsh, Irvine, UK

Abstract. Augmented Reality (AR) technology makes it possible to present information in the user's line of sight, right at the point of use. This brings the capability to visualise complex information for industrial maintenance applications in an effective manner, which typically rely on paper instructions and tacit knowledge developed over time. Existing research in AR instruction manuals has already shown its potential to reduce the time taken to complete assembly tasks, as well as improving accuracy [1-3]. In this study, the outcomes of several aspects of AR instructions are explored and their effects on the chosen Key Performance Indicators (KPIs) of task completion time, error rate, cognitive effort and usability are assessed. A standardised AR assembly task is also described for performance comparison, and a novel AR experimental tool is presented, which takes advantage of the flexibility of internet connected peripherals, to explore various different aspects of AR app design to isolate their effects. Results of the experiments are given here, providing insight into the most effective way of delivering information and promoting interaction between user and computer, in terms of user performance and acceptance.

Keywords: Augmented Reality, Industry 4.0, usability study.

1 Introduction

Augmented Reality (AR) uses visual technology to enhance the user's view of the real world with additional digital information, overlaid onto the user's line of sight. In the world of manufacturing, this can have a range of applications; from visualising designed objects in their real-world location to recognising and highlighting key points on production / control equipment, to displaying live sensor data at the point of use. In this paper, we focus on AR as an alternative to paper instruction manuals in the case of industrial maintenance and assembly tasks. Several research works demonstrate the potential benefits of this approach. Prominent figures in AR research, such as Funk, Erko-yuncu and Sanna have all presented studies claiming the benefits of AR guided assembly including faster task completion and fewer errors [4-6]. On the other hand, research such as that carried out by Zaldívar-Colado et al suggests Augmented or Mixed Reality

can, in fact, slow down assembly [7], suggesting that AR is not unanimously beneficial to the assembly process. Instead, performance depends very much on the particular design aspects of an AR system. So the question is raised: what makes an effective AR instruction guidance system? And how can we best display information such that operators can easily follow these instructions?

This paper introduces an AR application developed using standard web technologies and AR libraries which can be used to compare user performance when various aspects of app design are changed. A standardised AR assembly task is also proposed, to allow performance comparison between different instruction types. Finally, we present the results of experiments to determine the most effective way of presenting information in AR.

2 Background

AR has many possible uses in industrial maintenance, from view sharing with remote experts [8], to displaying live sensor data [9]. This paper focuses on the most common use – assembly guidance [1]. In this application, AR acts as an alternative to traditional paper-based instruction manuals, delivering task guidance directly in the user’s line of sight in the form of text annotations, 2D/3D models, animations or videos [10].

AR can be applied in a number of different forms including:

- Spatial AR systems - which use a series of projectors to highlight relevant locations
- Wearable AR devices - which use optical techniques to combine the user’s view of the world with virtual content
- Commercially available smartphones - used alongside video to display virtual content overlaid onto a live camera feed.

In this paper, we focus mainly on the latter option. The two topics of primary importance in this study are user acceptance and user performance when using the AR system.

Funk’s use of spatial AR projection in automotive assembly showed a decrease in time taken to complete each assembly, as well as a reduction in errors [4]. Erkoyuncu et al present the ARAUM (Augmented Reality Authoring for Maintenance) system which demonstrates a reduction in time to perform maintenance tasks [5]. Sanna et al also demonstrated time and error rate reductions in both skilled and unskilled users following AR instructions to repair laptops [6]. Golanski et al [11] demonstrated a successful implementation of mobile-based AR applications to support aircraft maintenance technicians. Though some reported the device and interface to be cumbersome, overall it was well received. These results are backed up by Aromaa et al [12] whose mobile-based AR platform was widely accepted in the global manufacturing company where they were tested, as measured by the QUIS (Questionnaire for User Interaction Satisfaction), SUS (System Usability Scale) and TAM (Technology Acceptance Model).

Conversely, Aschenbrenner et al [13] found using tablet-based AR applications to repair a switch cabinet did not improve results in terms of time taken or cognitive effort involved. The work of Zaldívar-Colado et al supports this [7], as they too found Mixed Reality may have a detrimental effect on assembly performance.

The common theme in all these studies is that they present a single system and evaluate it in terms of usability, or performance measures. Sometimes they are compared to a ‘control’ condition of paper instructions, sometimes no comparison is provided, and in many cases, only a simple comparison of means is used to judge success or failure of a system, rather than a thorough statistical analysis. Each of these systems differs hugely from one another. When we attempt to compare performance between systems, even where consistent measures are available it is impossible to tell whether differences are due to a particular feature of the AR system, the task chosen, or variation in the underlying populations sampled.

Not only is there a need for rigorous, statistically sound comparison between AR and non-AR instructions, there is also a value in attempting to isolate different aspects of AR guidance to identify which have the greatest effect on AR usability and performance ratings.

3 Effect Size Meta-Analysis

A meta-analysis of AR usability studies in the field of industrial maintenance tasks was carried out, to identify typical effect sizes in this field of research. Papers were identified from SCOPUS, Engineering Village and IEEE Xplore libraries and screened according to the PICOC Framework [14] criteria outlined in Table 1 below. Any papers using AR to guide humans during industrial maintenance included for analysis, while AR for training purposes, teleoperation of robots or medical uses were out of scope. Technical advances in hardware or software were not considered either, as the focus was on new applications of AR. There also needed to be a measurable indication of performance, and comparison against traditional instructions.

Table 1. PICOC Framework

	Inclusion	Exclusion
Population	Industrial maintenance task Human Operators Application	Training applications only Medical application Robotic control/teleoperation
Intervention	Utilisation of AR	Utilisation only of VR
Comparison	Paper/PDF manuals	
Outcomes	Time to complete operation Number of errors User experience	Hardware/software improvement Time to develop an application
Context	Industrial environments	Consumer environment

508 unique records were identified from the search criteria: title or abstract contains (“Augmented Reality” OR “Mixed Reality”) AND (“Maintenance” OR “Repair”) published since 2013. Once records which were irrelevant, or presented no usable data on

user performance were filtered out, 6 remained [6, 7, 15-18]. All 6 presented data on the time to complete the specified operation for both an AR instruction and paper instruction. Using Cohen’s D calculation [19], the effect size was calculated for each study. The results of these calculations are presented in Figure 1 below, alongside Cohen’s standard estimates for small, medium and large effect sizes.

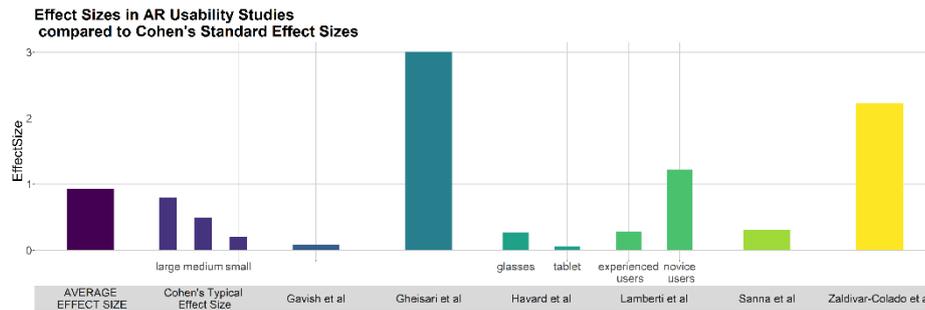


Fig. 1. Effect sizes in AR usability studies

It is clear, that the average effect size in this field (0.871) is closest to Cohen’s estimate of large effect size (0.8). This suggests it may be possible to detect statistically significant differences in performance, even in relatively small datasets.

4 Methodology

The aim of this work was to identify the effect of different aspects of the AR application design on user performance. First, definitions are required for “good performance” or “effectiveness” during the assembly process. In the papers reviewed in Section 2, the most frequently used measures of performance were Task Completion Time, Error Count, Cognitive Effort (which will be measured using the NASA-TLX scale [20]) and Usability (measured using the System Usability Scale [21]) – these will form the KPIs (Key Performance Indicators) for this work. The 2 independent variables selected for this stage of experimentation are Interaction Method (i.e. how user’s control and interact with virtual content) and Display Mode (i.e. what format the instructions are shown in). The levels of each factor are shown in Table 2 below.

Table 2. Independent variables changed in experiments

<i>Interaction Method</i>	<i>Native</i>	Users interact with content using the device's native interaction method to navigate instructions i.e. touch screen for mobile
	<i>Voice</i>	Users interact with content using voice commands to navigate instructions
<i>Display Mode</i>	<i>CAD</i>	Instructions conveyed to user as 3D models showing where actions should be performed
	<i>Text Annotation</i>	Instructions conveyed to the user by concise written instructions, linked to the relevant location by an arrow
	<i>In-situ Video</i>	Instructions conveyed to the user by videos of the action to be performed, projected over the relevant location

To explore the effects of these factors, an experiment was designed such that each user performed a short assembly task using Lego bricks, guided by a different version of the AR application (Section 4.1). Participants for the experiment were from a convenience sample and consisted mainly (95.8%) of university students, comprising 20 males and 4 females. Performance was then compared across the different versions to identify the factors and levels which produce the best assembly performance. As these two factors are closely related, it is likely that there will be an interaction between them, so it is necessary for a design of experiments to explore every possible combination of factors. Therefore, a fully factorial design is selected.

In addition to performing assembly tasks using Lego bricks, each participant was also asked to complete a similar task using paper instructions, to provide a baseline measurement of their assembly performance. In order to reduce the effect of individual variation on the basis of assembly performance, the results using the AR instructions were then subtracted from the baseline paper results to determine the difference. This will be denoted in the results section by Δ . The order the tasks were performed in was alternated to cancel out possible order effects.

Task time and error rate were recorded through the AR applications, or for paper versions through the investigator recording them. The cognitive effort was measured via questionnaires taken after each assembly, using the NASA-TLX scale [22], a widely recognised measure of cognitive effort. In this case, the raw TLX score was used (i.e. without weightings) as this has been shown to be either more or equally sensitive as the weighted version and much faster for participants to complete [20]. Usability was also determined using a post-experiment questionnaire, based on the System Usability Scale – a fast tool for assessing the ease and enjoyment associated with using IT-based systems [21].

4.1 System Design

The software created for the experiments was written in NodeJS and executed on a Raspberry Pi [23]. This allowed peripheral devices such as smartphones or other internet connected devices to be used for interfacing with the instructions which are controlled on a central webserver. To provide real-time interaction with the assembly instructions, WebSockets [24] were used to update the users display with the appropriate instruction for the task they were required to complete. This architecture was selected as it is not only easily expandable to multiple operators working on a common task, but also for the integration with IoT devices and a wider Industry 4.0 architecture, which could provide additional information.

Three display types were created for investigation:

- CAD Models – a 3D representation of the parts was displayed in the location and orientation of the intended placement.
- Annotations – written instructions were placed in the 3D environment describing the action required whilst indicating the location.
- Video – a pre-recorded video was activated to show the user where to place the parts.

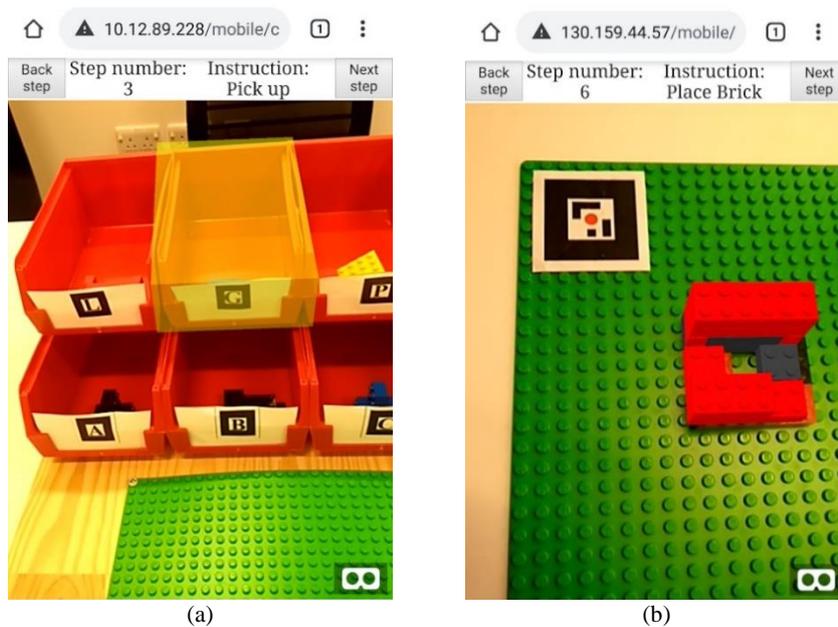


Fig. 2. View of instructions showing the location of (a) the selection, and (b) the placement of Lego bricks using virtual CAD models

4.2 Standard Assembly Task

The suggested standard assembly task for these experiments takes inspiration from Funk et al's work [4], in which participants were asked to build a section of a wall using

Lego bricks. Like Funk's example, the tasks used in the present work use Lego bricks since it is safe, familiar and low-cost. However certain changes have been made. When building Funk's wall, single bricks are added one at a time to an existing partially assembled wall, and there are a limited number of orientations which are possible to place the brick in.

The task presented here instead uses pre-made subassemblies of 1-5 bricks, and the orientation in which they are placed is very important to the next steps. This increased complexity is a better test of AR's capability to display objects in 3D in the existing environment.

The experimental situation was arranged as below in Figure 3 below, with a different Lego subassembly placed in each tray, and markers to indicate to both the human operator and the AR app, where the bricks should be selected from.

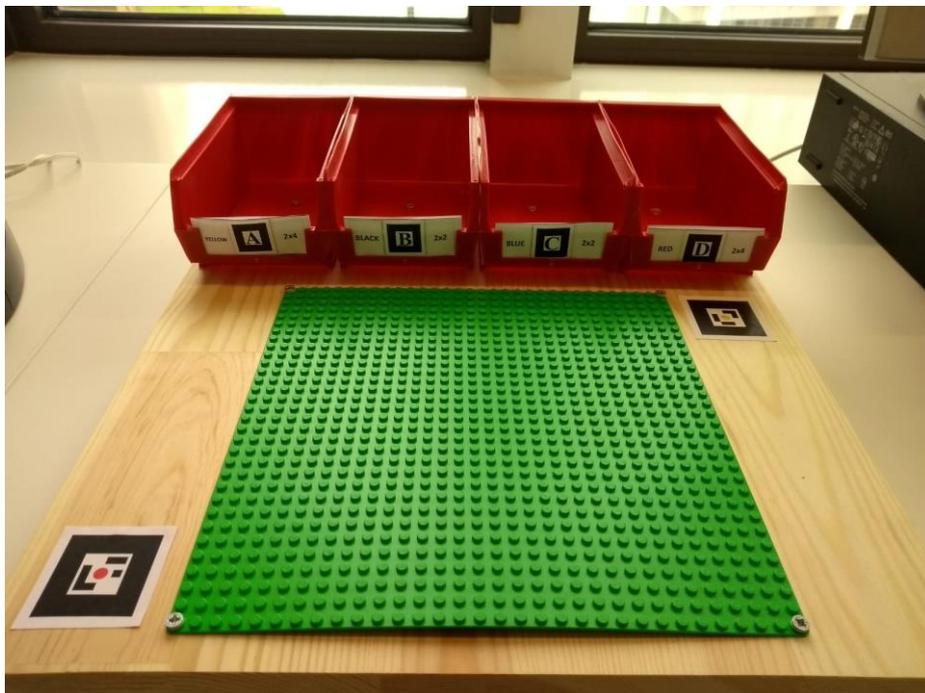


Fig. 3. Experimental Set-Up

Time and error rate data from the first step of each task was discarded, to discount loading time for the AR web app pages, as this is largely dependent on internet connection rather than the measured factors.

5 Results

5.1 AR vs Paper

First, user performance is compared between the AR application to the use of paper instructions, using a paired samples t-test - a statistical test used to investigate differences between conditions in a repeated measures experiment [25].

Table 3. Comparison of performance using AR and paper instructions

	TLX Score		Task Time (secs)		Error Rate (#)		System Usability	
	<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>
AR	1.845	1.040	157.0	44.90	0.5000	0.8341	39.96	2.694
Paper	1.593	0.8202	52.91	18.78	0.2917	0.6241	36.71	3.114

The results of the paired t-test show that statistically significant differences at the 90% confidence level or higher exist between the two conditions for Cognitive Effort (NASA-TLX Score), Task Time, and Usability. From Figure 4 below, we can see that while the users perform tasks slower while using the AR mode, they find the system more usable. This is reflected in some of the additional comments recorded from participants, who cited the AR system as being more enjoyable to use.

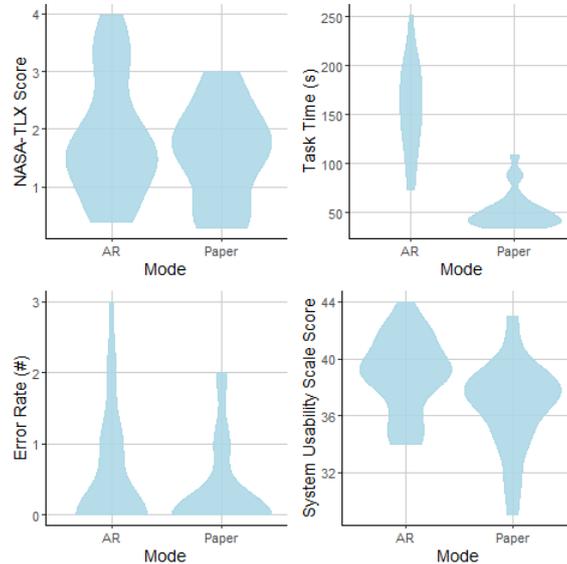


Fig. 4. Comparison of performance with AR instructions to paper instructions

Participants performed more poorly using AR instructions than paper across all but one of the measured variables. This differs from previous studies (described in Section 2),

and may be due to the length of the task – Funk et al [4], for example, found that the benefits of AR were greater the more steps each task had. However, the AR condition did still rate more highly on the system usability scale than paper, and enjoyment and ease of use are important factors to consider when implementing new technology.

5.2 AR Treatments

Further investigation into which AR factors lead to the best user performance during the assembly tasks was then undertaken. Table 4. summarises the key statistics generated.

Table 4. Data Summary

		Δ TLX Score		Δ Task Time (secs)		Δ Error Rate (#)		Δ System Usability	
		<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>
Interaction	<i>Native</i>	93.11	58.43	3.472	6.796	0.500	1.382	7.708	10.08
	<i>Voice</i>	115.1	43.81	2.847	10.31	-0.083	0.669	5.000	17.81
Display	<i>CAD</i>	107.3	49.35	2.500	10.33	-0.125	0.641	7.500	8.763
	<i>Text</i>	133.0	45.74	6.979	9.074	0.875	1.356	12.19	10.81
	<i>Video</i>	71.99	46.10	0.000	4.714	-0.125	0.991	-0.625	19.35

As there are multiple independent variables at multiple levels, an ANOVA (Analysis Of Variance) method was chosen to analyse the data. Using the ANOVA test, it was demonstrated that Display Mode has a significant effect on Δ Task Completion Time (95% level, Figure 5) and Δ Error Count (90% level, Figure 6).

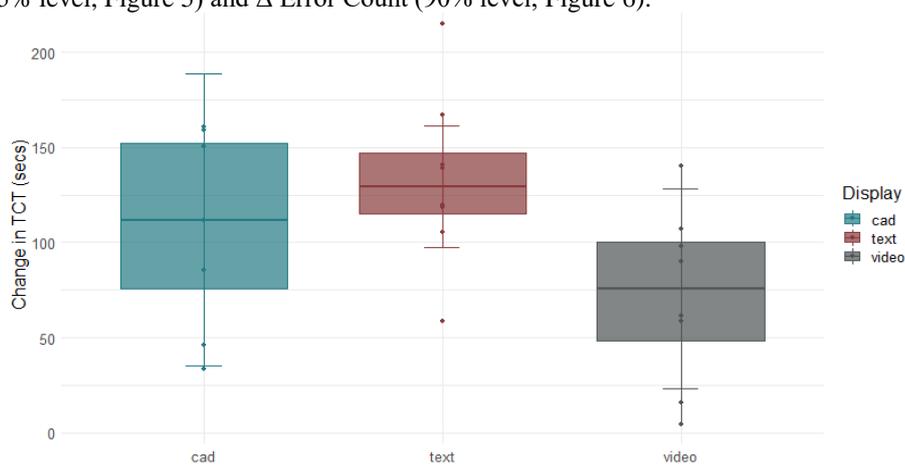


Fig. 5. Effect of display modes on total task completion time

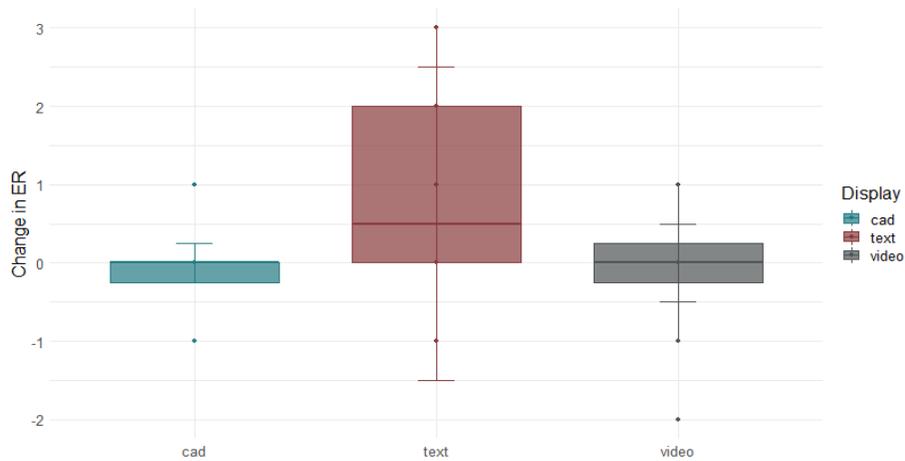


Fig. 6. Effect of display modes on the total error count

Participants using in-situ video-based AR instructions completed the task in an average of 131 seconds, 26% faster than text-based instructions and 21% faster than those using CAD models superimposed onto the environment.

Using Tukey's Honestly Significant Difference test [26, 27], the greatest performance difference lies between using video to display instructions and text-based annotations. In-situ video was found to produce significantly fewer errors than text-based annotations

There were too few female participants to draw statistically significant conclusions about the effect of gender on assembly performance using AR instructions. Nevertheless gender-disaggregated data is presented below, in the hope that it may be of use to future researchers performing meta-analyses on the subject.

Table 5. Gender disaggregated performance data (Task Time)

	Paper		AR	
	<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>
Female	95.5 s	59.2 s	167 s	56.7 s
Male	56.3 s	15.0 s	181 s	45.8 s
Overall	62.8 s	31.4 s	178 s	48.1 s

6 Conclusions

Although performance when using paper instructions was better than AR for task time, error count and cognitive effort, users reported finding the AR system easier and more enjoyable to use than its paper counterpart, and this is reflected in the system usability scores.

More significantly, it is shown that the way in which users interact with AR content has little effect on any of the performance measures studied. The practical implication of this is that:

- a. those implementing AR systems can allow users the option of choosing their preferred interaction method, safe in the knowledge that it will not significantly affect their performance; or
- b. this decision is best made on other factors such as background noise levels, and whether or not the user is required to wear gloves or other safety devices which may impair their operation of a touchscreen.

The way in which information is displayed in AR, however, does have a significant effect on performance:

- Using in-situ videos to demonstrate assembly tasks lead to significantly fewer errors than annotated models.
- Video also offered a slight improvement upon CAD models.
- Display mode had a significant effect on-task time - video condition again allows faster task completion than either annotations or CAD models.

This suggests that in-situ videos within an AR environment are a very effective way to display procedural instructions for assembly tasks.

Acknowledgements: This work forms part of a PhD project under the Renewable Engine INTERREG VA programme, which aims to encourage cross-border research and collaboration between the border regions of Scotland, Northern Ireland and the Republic of Ireland. This project, in particular, is a collaboration between the Advanced Forming Research Centre (AFRC) at the University of Strathclyde, and the industrial partner Booth Welsh.

References

1. R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi, "A systematic review of augmented reality applications in maintenance," *Robotics and Computer-Integrated Manufacturing*, Review vol. 49, pp. 215-228, 2018, doi: 10.1016/j.rcim.2017.06.002.
2. M. Funk, T. Kosch, R. Kettner, O. Korn, and A. Schmidt, "motionEAP: An Overview of 4 Years of Combining Industrial Assembly with Augmented Reality for Industry 4.0," in *i-KNOW '16*, Graz, 2016.
3. T. Kosch, Y. Abdelrahman, M. Funk, and A. Schmidt, "One size does not fit all - Challenges of providing interactive worker assistance in industrial settings," in *UbiComp/ISWC 2017* -

Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 2017, pp. 1006-1011, doi: 10.1145/3123024.3124395. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85030868685&doi=10.1145%2f3123024.3124395&partnerID=40&md5=8753b08cb92a1bca7c452d192389664d>

4. M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt, "A benchmark for interactive augmented reality instructions for assembly tasks," presented at the Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia, Linz, Austria, 2015.
5. J. A. Erkoyuncu, I. F. del Amo, M. Dalle Mura, R. Roy, and G. Dini, "Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality," *CIRP Annals*, vol. 66, no. 1, pp. 465-468, // 2017, doi: <https://doi.org/10.1016/j.cirp.2017.04.006>.
6. A. Sanna, F. Manuri, F. Lamberti, G. Paravati, and P. Pezzolla, "Using handheld devices to support augmented reality-based maintenance and assembly tasks," in *Consumer Electronics (ICCE), 2015 IEEE International Conference on*, 2015: IEEE, pp. 178-179.
7. U. Zaldivar-Colado, S. Garbaya, P. Tamayo-Serrano, X. Zaldivar-Colado, and P. Blazevic, "A mixed reality for virtual assembly," 2017: IEEE, doi: 10.1109/roman.2017.8172385. [Online]. Available: <https://dx.doi.org/10.1109/ROMAN.2017.8172385>
8. R. Masoni *et al.*, "Supporting Remote Maintenance in Industry 4.0 through Augmented Reality," *Procedia Manufacturing*, Article vol. 11, pp. 1296-1302, 2017, doi: 10.1016/j.promfg.2017.07.257.
9. M. F. Alam, S. Katsikas, O. Beltramello, and S. Hadjiefthymiades, "Augmented and virtual reality based monitoring and safety system: A prototype IoT platform," *Journal of Network and Computer Applications*, Article vol. 89, pp. 109-119, 2017, doi: 10.1016/j.jnca.2017.03.022.
10. I. F. del Amo, J. A. Erkoyuncu, R. Roy, R. Palmarini, and D. Onoufriou, "A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications," *Computers in Industry*, vol. 103, pp. 47-71, 2018.
11. P. Golanski, M. Perz-Osowska, and M. Szczekala, "A Demonstration Model of a Mobile Expert System with Augmented Reality User Interface Supporting M-28 Aircraft Maintenance," *KONBiN*, vol. 31, no. 1, p. 23, 2014.
12. S. Aromaa, A. Väättänen, M. Hakkarainen, and E. Kaasinen. *User experience and user acceptance of an augmented reality based knowledge-sharing solution in industrial maintenance work*, *Advances in Intelligent Systems and Computing*, vol. 607, pp. 145-156, 2018.
13. D. Aschenbrenner, M. E. Latoschik, and K. Schillingz, "Industrial maintenance with augmented reality: Two case studies," in *22nd ACM Conference on Virtual Reality Software and Technology, VRST 2016, November 2, 2016 - November 4, 2016*, Munich, Germany, 2016, vol. 02-04-November-2016: Association for Computing Machinery, in Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST, pp. 341-342, doi: 10.1145/2993369.2996305. [Online]. Available: <http://dx.doi.org/10.1145/2993369.2996305>
14. A. Booth, A. Sutton, and D. Papaioannou, *Systematic approaches to a successful literature review*. Sage, 2016.
15. V. Havard, D. Baudry, X. Savatier, B. Jeanne, A. Louis, and B. Mazari. *Augmented industrial maintenance (AIM): A case study for evaluating and comparing with paper and video media supports*, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9768, pp. 302-320, 2016.

16. M. Gheisari, G. Williams, B. N. Walker, and J. Irizarry, "Locating building components in a facility using augmented reality vs. paper-based methods: A user-centered experimental comparison," in *Computing in Civil and Building Engineering - Proceedings of the 2014 International Conference on Computing in Civil and Building Engineering*, 2014, pp. 850-857, doi: 10.1061/9780784413616.106. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84934286807&doi=10.1061%2f9780784413616.106&partnerID=40&md5=72be0986a0438d7268040a8f6ec9c63e>
17. N. Gavish *et al.*, "Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks," *Interactive Learning Environments*, Article vol. 23, no. 6, pp. 778-798, 2015, doi: 10.1080/10494820.2013.815221.
18. F. Lamberti, F. Manuri, A. Sanna, G. Paravati, P. Pezzolla, and P. Montuschi, "Challenges, opportunities, and future trends of emerging techniques for augmented reality-based maintenance," *IEEE Transactions on Emerging Topics in Computing*, Article vol. 2, no. 4, pp. 411-421, 2014, Art no. 7024955, doi: 10.1109/TETC.2014.2368833.
19. J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.
20. S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904-908, 2006, doi: 10.1177/154193120605000909.
21. J. Brooke, "SUS-A quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4-7, 1996.
22. S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," 1988.
23. T. V. Guinard D, *Building the Web of Things: With examples in Node.js and Raspberry Pi*. Greenwich, CT, USA: Manning Publications Co, 2016.
24. V. Pimentel and B. G. Nickerson, "Communicating and Displaying Real-Time Data with WebSocket," *IEEE Internet Computing*, vol. 16, no. 4, pp. 45-53, 2012, doi: 10.1109/mic.2012.64.
25. A. M. Colman, "related scores t test," ed: Oxford University Press, 2015.
26. H. Abdi and L. J. Williams, "Tukey's honestly significant difference (HSD) test," *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage, pp. 1-5, 2010.
27. H. Abdi and L. J. Williams, "Tukey's honestly significant difference (HSD) test."