Applied Ergonomics 98 (2022) 103597

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

A head mounted augmented reality design practice for maintenance assembly: Toward meeting perceptual and cognitive needs of AR users

Dedy Ariansyah [a,*], John Ahmet Erkoyuncu [a], Iveta Eimontaite [b], Teegan Johnson [b], Anne-Marie Oostveen [b], Sarah Fletcher [b], Sarah Sharples [c]

[a] Centre of Digital Engineering and Manufacturing, School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK43 0AL United Kingdom
[b] Industrial Psychology and Human Factors Group, School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK43 0AL, UK
[c] Human Factors Research Group, Faculty of Engineering, University of Nottingham, University Park Nottingham, NG72RD, UK

ARTICLE INFO

Keywords:
Augmented reality
Head-mounted display
Eye gaze behaviours
Joint cognitive system
Usability
Multiple resource model

Head Mounted Display (HMD) based Augmented Reality (AR) is being increasingly used in manufacturing and maintenance. However, limited research has been done to understand user interaction with AR interfaces, which may lead to poor usability, risk of occupational hazards, and low acceptance of AR systems. This paper uses a theoretically-driven approach to interaction design to investigate the impact of different AR modalities in terms of information mode (i.e. video vs. 3D animation) and interaction modality (i.e. hand-gesture vs. voice command) on user performance, workload, eye gaze behaviours, and usability during a maintenance assembly task. The results show that different information modes have distinct impacts compared to paper-based maintenance, in particular, 3D animation led to a 14% improvement over the video instructions in task completion time. Moreover, insights from eye gaze behaviours such as number of fixations and transition between Areas of Interest (AOIs) revealed the differences in attention switching and task comprehension difficulty with the choice of AR modalities. While, subjective user perceptions highlight some ergonomic issues such as misguidance and over-reliance, which must be considered and addressed from the joint cognitive systems' (JCSs) perspective and in line with the predictions derived from the Multiple Resources Model.

1. Introduction

The main reason why augmented reality (AR) is so attractive to industry lies in its ability to improve a user's perception of their environment by overlaying relevant digital information on the real, physical world, and thereby, enable them to perform the task more effectively and efficiently (Nee et al., 2012). AR systems when working to their full potential will provide operators with support in coping with the complexity of dynamic manufacturing environments. However, prior to reaching this, they need to be designed in such a way as to facilitate effective working conditions as a joint cognitive system (JCS) (Jones et al., 2018). The underpinning philosophy of a JCS (Hollnagel, 2005) is a notion of working in partnership between people and systems, where all elements combine to work towards a common goal. Developing an understanding of what this partnership looks like requires research to understand the best way to provide and the suitability of interaction mechanisms.

Enriching the user's viewpoint with contextual information through AR has shown that many perceptual/cognitively demanding tasks can be done better, easier, and faster compared with traditional methods in various applications in design and manufacturing (Nee et al., 2012). Nevertheless, combining both digital and physical spaces in a way that enables users to perceive and comprehend both virtual and real objects simultaneously while doing one or more real world tasks is challenging (Gabbard et al., 2019). It is widely recognised that in order for AR to be an efficient and effective solution it requires: 1) powerful hardware to facilitate proper information overlay such as high resolution, sufficient field of view, brightness, and contrast, 2) reliable software to achieve accurate tracking, robustness, ease of calibration, authoring solution and content management tools, and 3) Human Factors (HFs) considerations in the design of the interface and interactions (Akçayir et al., 2016; Poushneh, 2018). Literature concerning AR applications in manufacturing and maintenance shows that some pertinent challenges to foster the adoption of AR are focused around hardware capabilities, robust tracking systems, and content creation (Bottani and Vignali, 2019; Nee et al., 2012; Palmarini et al., 2018). Although the need for HF studies is recognised in AR literature studies, user-based experiments examining different interface designs, information requirements,

* Corresponding author.
E-mail address: dedy.ariansyah@cranfield.ac.uk (D. Ariansyah).

https://doi.org/10.1016/j.apergo.2021.103597
Received 24 June 2021; Received in revised form 25 August 2021; Accepted 22 September 2021
Available online 28 September 2021

display types, etc. and their impacts on task performance, ergonomics, and safety are limited (Bottani and Vignali, 2019). A lack of research in this area leads to little understanding of 1) whether AR benefits are affected by hardware or interface factors (Livingston et al., 2005), 2) what factors may contribute to increased user task performance or productivity (Braly et al., 2019), 3) whether a certain type of modality is better suited than others for a particular type of task, as well as 4) safety-related issues and 5) the impact of long-term usage.

Although the benefits of AR for manufacturing and industrial maintenance have been widely demonstrated over traditional methods (Braly et al., 2019; Hoover et al., 2020; Kim et al., 2019), AR uptake in industry has been low because of unresolved usability issues (Schwerdtfeger et al., 2009). On one hand, failure to use the right AR interface could lead to suboptimal performance of AR systems, which reduces the appreciation of the technology for the investment. On the other hand, trial-and-error-style approaches to the implementation of AR systems due to lack of design practice (Rizzo et al., 2005) may result in task performance degradation (Schwerdtfeger et al., 2009), additional workload and distraction (Funk et al., 2017), as well as exposure to occupational hazards (Kim et al., 2019). As a consequence, AR may be perceived as unreliable, limiting user trust, and having negative safety implications (Endsley et al., 2017). For this reason, pushing forward AR technology adoption in industries would require new design practices to create an effective AR system from a Human Factors perspective.

Unfortunately, there is limited existing literature that provides design guidelines for addressing the usability aspects of AR implementation. One study proposed user-centred design principles for approaching the design of AR interfaces (Dünser et al., 2007). Nevertheless, as noted by the authors, the detailed implementation of the AR system is unclear and needs further refinement. A study by Endsley et al. (2017) presented a heuristic AR design guideline to be considered during the design process of a dynamic augmented experience. However, it is limited to a set of design instructions and does not provide any further detail on how it is applied and how these instructions correlate with the task performance metrics. Besides, there is a limited study in manufacturing and maintenance application that views AR design principles in the light of cognitive theory to predict the level of performance of multi-perceptual information and response selection/execution. As implication, the choices of modalities and the interaction are arbitrarily assigned and does not have grounded theory to dictate how and when the technology should be employed. In addition, AR implementation contexts and the input and output interfaces are diverse and therefore questions are raised as to whether design guidelines are transferable across AR devices and able to generate similar usability and effectiveness outcomes. Some recent surveys of AR use cases for industrial applications in maintenance and manufacturing/assembly related tasks reported that the Head-Mounted Display (HMD) is the most prominent AR device due to its mobility and hands-free feature (Dey et al., 2018; Palmarini et al., 2018). Of all the identified studies that used HMD, only a few have reported HF design considerations regarding AR interfaces and interaction into technology development. It is apparent that a lack of adoption of HF principles in AR technology development stems from the limited design practices regarding what is known to be effective for AR HMD system in specific use case assistance, and for certain groups of AR users.

This paper aims to provide a theoretically driven approach based on the Multiple Resource Model (MRM) (Wickens, 2002, 2008) and the JCS's perspective (Hollnagel, 2005) to the design practice for AR modalities. This study considers the *information mode* and *input interface* for the implementation on HMD. The reason for focusing on these two aspects is due to their criticality in ensuring effective information communication and progress through the task. Both factors are tied to the two main principles of JCS: designing for complexity by supporting functions of coping, through the identification of effective means of information provision, and support for time management through intuitive interaction interface mechanisms (Hollnagel, 2005). The application considered is in industrial assembly for manufacturing and maintenance, as these design principles have to be identified in specific domains and fields prior to generalisation across a broader context (Hollnagel, 2005). In addition, this study also recorded and analysed eye gaze behaviours and user perception to capture possible factors that contribute to a variation in performance and user safety, aspects that are rarely considered in the evaluation of AR usability in manufacturing and maintenance application. MRM was used to frame the selection of the AR modalities and to guide the analysis of the task performance difference in comparison with a paper-based manual. The use of MRM is considered appropriate in this study due to the composition of the model that is sufficiently granular to predict performance variance under different processing task, input/output modality, and types of stimuli. The theoretical basis of this model and JCS are described in the following section.

2. Theoretical background

The multiple resource model (MRM) proposes multiple attention resources that account for variance in time-sharing performance (Wickens, 2002, 2008). The MRM posits that attentional resources are limited, and their structure can be described by four categorical in which each dimension has two discrete levels: processing stages (visual/cognitive and responding), processing codes (verbal and spatial), input modality (visual and auditory), and responses (manual and verbal). According to the model, performance outcomes of concurrent multiple tasks, or a complex single task with multiple components are dictated by the degree to which common resources must be shared across tasks or task component and not merely the total resources invested in the task. From this perspective, using AR in manufacturing or maintenance application (e.g. to support manual assembly) should not impose cognitive challenges provided that task components that require perception and response selection/execution utilise different/separate pools of resources. MRM can also inform the prediction of the performance outcomes when selecting AR modalities such that a higher number of resources taxed while performing the task using AR will likely lead to higher cognitive efforts and suboptimal task performance.

Furthermore, task performance (mental workload) also appears to vary under resources compatibility concept (Wickens, 2008; Wickens et al., 1983). Wickens et al. (1983) propose the concept of Stimulus-Central Processing-Response (S–C-R) compatibility model that predicts the performance variations based on compatibility relation between modalities of input (auditory, visual or A, V) and output (manual, speech, or M, S), and codes of central processing (spatial versus verbal). In that, the performance benefits of A/S modalities will be most attained when they are applied for the verbal encoding, whereas those of V/M modalities will lead to best performance when they are associated with the spatial encoding. In this sense, the principle of combability model could also dictate performance trend in the assignment of AR modalities to the type of processing tasks in single-task conditions.

These models were used to frame the design practice for AR modalities selection in this study and predict the impact on performance differences. Nevertheless, the design of AR system should go beyond this paradigm if it is to be used effectively in the complex dynamic manufacturing environment. This means that user and the system should work collaboratively, and thus, the interface and the interaction between human agent and AR agent should be designed as a joint cognitive system (JCS). JCS is characterised by three principles (Hollnagel, 2005): (a) support for coping, (b) time management, and (c) predictability. The first principle enables strategies for coping rather than enforcing a particular strategy. The second enables support for time management to do the work. The predictability aspect provides predictions and anticipation for coping strategy. Applying these principles in a JCS ensures the ability to maintain control, both in terms of minimising the unexpected situation and being ready to respond when they occur (Hollnagel, 2005). This study applied JCS perspective to analyse user capabilities (i.e. coping strategy and time management) under the influence of different AR modalities and to examine to what extent both user and technology acting together can successfully carry out the required task. These insights are critically important because the efficacy of a joint cognitive system is only realised when the cognitive coupling between user and technical system is properly designed (Angeli, 2013). In this sense, there is a need for a good understanding of cognitive characteristics of users and the corresponding cognitive characteristics of the system. Otherwise, the unpredictable behaviour of each agent will result in incompatible integration.

3. Related work

It has been long established (Nichols and Patel, 2002) that the overall impact of technologies such as VR results from a combination of technology design, interface/environment design, task circumstances (e.g. task goals, length of period of use) and individual characteristics. A small number of articles have examined the fundamental perceptual needs of humans when using AR on HMD to support its usage in industrial applications. Kim et al. (2019) studied the effect of different HMD types (i.e. monocular vs. binocular), information mode (i.e. text vs. graphic), and information availability (i.e. always-on vs. on-demand) on a simulated warehouse job involving order picking and part assembly. They discovered that job performance and workload were more affected by the user interface design rather than by HMD type. They reported that using HMD with graphic-based instruction and always-on mode instruction led to improved job performance and reduced perceived workload compared to a paper pick list condition. The better performance obtained by graphic-based rather than text-based instructions can be attributed to the ease of information processing and a lower workload demand when confirming information on a screen.

Nevertheless, the use of HMD caused an increase in discomfort and has some safety implication in warehouse jobs such as being distracted while walking which may result in slips, trips, and fall-related injuries. This finding highlights the importance of UI design of AR for warehouse workers and the need for HMD improvements to provide comfort to AR users, especially when required to wear them during prolonged tasks.

As argued by Norros (2014), in addition to the need for the actor (operator) and the artefact (in this case the AR device) to be considered in design principles as detailed in Hollnagel's (2005) description of the joint cognitive system, it is also critical to consider the user environment. This is particularly evident for AR where the surrounding environment can not only impact the effectiveness of the presentation of information but additionally an understanding of environments is required to ensure the safety of the user during use. The effect of industrial backgrounds and text styles on the legibility of the text has also been investigated with respect to different HMD display types such as optical vs. video (Gattullo et al., 2015). In a real industrial setting, there are many scenarios in which background conditions affect the legibility, and thus, the perception of the augmented graphics on HMD. There was a statistically significant main effect of background and text styles on user's response time, but this did not affect error rates. The effect of the background on the legibility could be attributed to different luminance profiles of the background. The background that had neutral luminance improved the legibility. Meanwhile, the use of text billboards improved legibility on real industrial background and supports the need for a mandatory industry colour.

In regard to perceiving virtual and real information simultaneously, one study examined the negative effect of context switching (i.e. real to AR view) on user perception and eye strain wearing a monocular optical HMD (Gabbard et al., 2019). Since the object of interest in the real world can occur at a range of distances, an optical HMD with a fixed focal length requires users to continuously switch both accommodation and attention between real-world scene and virtual graphics. Tested under a text-based visual search, context switching had a negative impact on performance when information was presented at the far distance (6 m), but not when it was presented at either the near (0.7 m) and medium (2 m) distances. Moreover, context switching between AR and real-world visual information resulted in significantly higher levels of eye fatigue at all distances. Furthermore, the effect of context switching was also assessed by changing the focal length distance. When colour semantics are not important, white text on a blue billboard has been recommended for both optical and video HMD. Background illuminance strongly affects optical HMD but hardly affects video HMD. Both types of HMD support similar performance at higher lighting levels, about 1000 lux. In the brightest lighting condition, about 4000 lux, user performance with optical HMD dropped significantly, and hence, video HMD might be more suitable for higher lighting setting. The main disadvantage for the optical HMD concerns brightness and contrast limitations, whereas the major drawback of the video HMD is the real-time perception of the real world.

In regard to perceiving virtual and real information simultaneously, one study examined the negative effect of context switching (i.e. real to AR view) on user perception and eye strain wearing a monocular optical HMD (Gabbard et al., 2019). Since an object of interest in the real world can occur at a range of distances, an optical HMD with a fixed focal length requires users to continuously switch both accommodation and attention between the real-world scene and virtual graphics. Their study found a significant effect of focal distance switching on task performance. Participants completed more sub-tasks and were more accurate when focal distance switching was not required. They further concluded that context switching, and focal distance switching are important AR user interface design issues.

Using monocular HMD, the impact of information display position (i.e. peripheral vs. central) for preventive car maintenance was assessed (Zheng et al., 2015). Their user-based study found that the display position of task instructions had a significant effect on task performance; with a peripheral display yielding a longer completion time than a central display. In their view, a peripheral display is not natural for human eyes to capture information, which leads to inefficient information fixation. One interesting finding in their study related to the poor choice of AR interface, is that a peripheral display yielded significantly longer completion times than any other conditions including the traditional approach for the searching task (Locate action).

In a different study, the level of information requirement was also studied on a Computer Numerical Control (CNC) machine repair task using a HMD (Zhu et al., 2013). They found that the appropriate level of detail of AR instructions could help the expert users to complete the task more quickly than the unregulated detail of AR information. This finding has a practical design implication in making AR systems context-aware toward its users' background for developing a more effective AR solution. The level of AR effectiveness with respect to task difficulty has also been studied in a remote maintenance application where users were asked to perform repair tasks with a variation in object and procedural awareness (Fernández del Amo et al., 2020). Object awareness (OA) refers to the ability of messages to identify the real-world objects being referred to. Meanwhile, procedure awareness (PA) refers to the ability of messages to clearly define the task that is being referred to. The study found that tasks with a lower difficulty of OA and PA might not show significant benefits of using an AR system over the traditional method, whereas the increasing complexity of OA and PA would result in greater performance gain of an AR solution over a traditional approach. This finding has a design implication for the appropriate level of AR information detail provision, which should be regulated to account for the varying skill levels of technicians and task difficulties in order for the AR solution to be optimal.

In relation to the limitation of HMD compared to other AR devices (e.g. projector, handheld device), HMD has a drawback with its small Field-of-View (FoV). This leads to difficulty in directing user's attention to the objects that are not in the field of view. One study conducted a user study to investigate the effect of AR-based viewpoint guidance towards off-screen targets on mental work and visual attention (Markov-Vetter et al., 2020). They tested three types of viewpoint guidance along the exocentric (Map, Map+) – egocentric (Arrow) dimension on a cued localisation task. The localisation task required the detection of off-screen targets, which were presented one after another, while the participant's time and accuracy were recorded. Their evaluation revealed a significant effect of type of guidance on the speed of off-screen target detection and accuracy. Arrow type guidance led to a faster target detection (even with a secondary task), lower workload, and higher accuracy compared to other conditions. The finding suggests that egocentric cues (e.g. Arrow) are an efficient form of viewpoint guidance for the user towards an off-screen target.

In the recent study, the effect of different instruction modes (google glass based-animation, based-video, and a printed manual) were examined on task performance, visual fatigue, subjective workload and usability between gender (Wang et al., 2019). The study found that in comparison with traditional manual, AR animation was helpful in minimising the errors while video visualisation complicated the task due to the need to switch attention. In terms of visual fatigue, female participants were reported to experience higher visual fatigue, higher workload, and lower usability perception when using AR compared to male participants. It is apparent that system usability and visual fatigue for AR HMD needs to be further improved. This finding suggests that google-glass based animation presents a better alternative to traditional manual for the disassembly task that is prone to error. Notwithstanding the fact that the dynamic visual modality of video or animation has been agreed to be engaging and efficient for information communication, the difference between these two in terms of how users interact with different types of dynamic information is still unclear. Some research suggest that animation might be more engaging and emphasizing the components of the process to learn (Rieber et al., 2004), other research suggest that visualisations might distract from the actual process being learnt (Lowe, 1999, 2004) suggesting that videos might be a more effective media for learning. Furthermore, visual modalities have the advantage of simplifying complex information, easing comprehension of the process and illustrating how different aspects of the process relate to each other (Smaldino, 2012). Although a number of the studies indicate that the combination of modalities (i.e. video with text annotation) delivers the best learning performance and engagement (Houts et al., 2006), others suggest that dynamic visualizations improve learning (Höffler and Leutner, 2007) in particular from concrete to abstract learning (Nguyen et al., 2012; Wu et al., 2013).

Table 1 summarises the existing research in AR interface design evaluation for HMD in manufacturing and maintenance applications. Although there are some other usability studies that investigated AR modalities, those studies were not included in this review because they consider the comparison of AR interactions under different AR displays (Marques et al., 2020), have no clear contexts in the manufacturing and maintenance application (Ahn et al., 2019), and comes from other domains (Moosburner et al., 2019). This section summarised the current research into providing AR HMD design practice for manufacturing and industrial maintenance, and covered factors related to HMD attributes, AR interfaces, and task characteristics. Nonetheless, those factors have not been investigated under the construct of human cognitive theories as well as the way different user interactions with AR information is captured and their impact on task performance, system usability, and user safety. This paper sought to demonstrate how MRM and JCS principles are helpful to inform the prediction of the performance outcomes when selecting AR modalities such that higher number of resources taxed while performing the task using ARs is associated to higher cognitive efforts and suboptimal task performance, and therefore, affects user capabilities acting together with technology to successfully carry out the required task. In this study, we assess the different impacts of common information modes (IM) (i.e. video vs. 3D animation) used in AR for manual assembly/disassembly in maintenance, and the method of input interface (II) of the AR headset (i.e. hand-gesture vs. voice command) on task performance, workload, and usability. Finally, this

Table 1
AR HMD design evaluation in manufacturing and maintenance application.

Authors and Task	Attributes	Performance measures	Usability and safety-related issues
Kim et al. (2019) - Pick part - Assemble	HMD types - Monocular - Binocular	CT: NS E: NS	- The use of HMD for performing cognitive task while walking may have workplace safety implication such as slips, trips, and falls-related injuries
	Information mode - Text-based (T) - Graphic-based (G)	**CT: S** T (H) G (L) **E: S** T (H) G (L)	- Using HMD with graphic-based information led to improved job performance, reduced perceived workload, and increased discomfort compared to paper-based baseline condition
	Availability mode - Always-on (Ao) - on-demand (oD)	CT: NS **E: S** Ao(L) oD (H)	- On-demand presentation mode led to increased perceived workload and reduced mean usability at 11% compared to always-on
Zheng et al. (2015) - Assembly and disassembly	Display positions	CT: S C(L) P (H)	- Peripheral display position causes user to move their eyes to process the information, which is unnatural.
	- Central (C) - Peripheral (P)		- Peripheral look is normally used for quick glance and not for fixation. This is probably why central display produced a shorter processing time.
	Type of display	CT: NS	- Participants adapted well with the non-eyewear assistance by holding with one hand and work with the other hand.
	- eyewear - non eyewear		- The semi-transparent screen made participants difficult to focus to the extent that they had to cover with one of their hands to provide a dark background which contribute to the extended completion time
	Task difficulty	**CT: S**	- Insufficient information can make one task more difficult than the other. It was observed that when AR instruction was not sufficient, it took participant longer to complete the task than when using non-wearable assistance, possibly because eyewear technology decreased participant's ability to think themselves (over-reliance)
	Action - Read - Locate - Manipulate	**CT: S** R (L) L (H) M (H)	- Task that requires searching information i.e., Locate induced a longer completion time

(*continued on next page*)

Table 1 (*continued*)

Authors and Task	Attributes	Performance measures	Usability and safety-related issues
	- Compare		for peripheral display compared to other types of displays. This is because human vision is less efficient at fixating eccentric targets compared to eye-centred targets.
Zhu et al. (2013) -Assembly and disassembly	LOD	CT: (S N/A) L (H) H (L)	- With the same expertise level, a group of participants that was presented with sufficient level of detail of AR information completed task faster compared to the other group with more detail of AR information.
	- Low (L) - High (H)		- Participants rated the adjustable level of AR information detail for completing the task more satisfying, intuitive, and easy to use compared to the non-adjustable detail of AR information
Gattullo et al. (2015) -Reading the graphical instruction	Text styles	**RT = S** B (H) E = NS	- There was a statistically significant effect of background and text style on user's response time but did not affect error rates
	- plain text (P) - outline (O) - billboard (B)		- The use of billboard improved legibility on real industrial background and supported mandatory industry colour.
	Backgrounds	**RT = S** E (H) E = NS	- The effect of background on the legibility could be attributable to different luminance profile of the background
	- testbed frame (TF) - tool workbench (TW) - engine (E)		- The background that has neutral luminance improved the legibility
	Billboard style	RT = (S N/A) E = (S N/A)	- White text on the mandatory billboard colour (e.g. red) produced shorter response time
	- white text - black text - colored text - no billboard		- It is important for the billboard to have a contrast between colour and text
	Outline thickness	RT = (S N/A) E = (S N/A)	- Adding outline improved the readability of the text for video-see through display whereas increasing the width of the outline did not improve the performance
	- no outline		- The use of plain black text in video see-through contrasts well with most backgrounds without the need for any outline
	- variation of outline		- Legibility does not improve for optical see-through even with the largest outlines
	HMD types	RT = (S N/A) E = (S N/A)	- When colour semantics are not important, it is recommended to use white text on blue billboard for both optical and video HMD
	- Optical		- Background illuminance strongly affects optical display but hardly affects video display
	- Video		- Both types of HMD support similar performance at higher lighting levels, about 1000 lux. In the brightest lighting condition, about 4000 lux, user performance with optical display dropped significantly. - The main disadvantage for the optical display concerns brightness and contrast limitations, whereas the major drawback of the video display is the real-time perception of the real world
Gabbard et al. (2019) Text-based visual search	Context switching	TC: S RR(H) RA (L) **E: S** RR(H) RA (L)	- Context switching had a negative impact on performance when information was presented at the far distance, but not at the near and medium distances.
	- Real to Real (RR) - Real to AR (RA)		- Context switching between AR and real-world visual information resulted in significantly higher levels of eye fatigue at all distances
	Focal distance switching	**TC: S** NR (H) R (L) **E: S** NR (H) R (L)	- Participants completed more subtasks, and were more accurate, when focal distance switching was not required.
	- Not required (NR) - Required (R)		- This will minimize changing accommodation and convergence, and therefore, put the least amount of strain on the eye's oculomotor mechanism
	Matched (M) vs. mismatched (mM)	**TC: S** M (L) mM(H) **E: S** M (L) mM(H)	- When the distance to real world text, distance to virtual text, and focal depth matched, participants performed better than when the distances did not match.
Markov-Vetter et al. (2020) Detecting	Type of guidance - Map (M) - Map+ (M+)	**CT: S** A(H) M+ (L) M (H) **E: S**	- Arrow type guidance led to a faster target detection (even with a secondary task), lower

(*continued on next page*)

Table 1 (*continued*)

Authors and Task	Attributes	Performance measures	Usability and safety-related issues
off-screen target	- Arrow (A)	A (L) M+ (H) M (H)	workload with higher accuracy compared to other conditions.
Wang et al. (2019) Disassembly task	Instruction modes	**CT: S** M(L) A(L) V (H)	- AR method (i.e. animation) had the best task performance in reducing the errors
	- Manual (M) - 3D animation (A) - Video (V)	**E: S** M (H) A (L) V (H)	- Female participants suffered higher level of visual fatigue, subjective mental workload, and perceived lower system usability when using google glasses-based AR than males

CT: Completion Time, RT: Reaction Time, E: Error, H: High, L: Low, S: Significant at $p < 0.05$, NS: Not significant.

paper attempts to test four key research questions: 1) how is 'task performance gain' using AR affected by different types of information mode (i.e. video vs. 3D animation) and interaction mode (i.e. hand gesture vs. voice command), 2) do AR instructions lead to a lower workload than a paper-based manual, 3) can eye gaze behaviour uncover different performance gains between different AR modalities 4) what are the participant attitudes towards AR technology which offers a new way of accessing and interacting with information to enhance their task performance.

4. Methods

4.1. Participants

A total of sixty-three volunteers recruited through email and an announcement on the University's intranet homepage participated in a paper-based manual and AR-based maintenance task. Nine of them were excluded due to incomplete data caused by errors in the technical systems, and three of them were excluded due to age-related performance outliers (>50 years). The technical errors were related to the limitation of AR tracking library that failed to recognise the object target which result in AR content not or properly overlaid in the real environment. The recognition errors could be due the physical object that being used consists of moving parts, surface with poor texture, and a limited number of contrast-based features (PTC, 2019). As a result, some participants had trouble to see AR contents to perform the task. The age-related performance factors include vision distance related problems and slow hand-eye coordination. The background of the volunteers was a combination of students and staff with varying levels of maintenance experience and exposure to AR technologies. Their mean age, gender, maintenance and AR experience were balanced across the groups and are presented in Table 3. Maintenance experience was calculated based on how many maintenance related activities they had done in the past five years (e.g. repair computer, assemble/disassemble home appliances, change the oil of the machine). AR exposure was calculated based on how many times approximately they had used AR system in the past five years. All volunteers reported being healthy and not having any musculoskeletal injuries in the past 12 months.

4.2. Experiment Apparatus

The equipment used for the maintenance task was a gearbox machine, which is commonly used for demonstrating condition-based monitoring for teaching purposes at Cranfield University. This machine was selected because it comes with a paper-based manual that is used for troubleshooting and repair. A set of repair toolkits were

Table 2

The distribution of resources allocation according to MRM in different maintenance conditions.

Conditions	Input modality	Processing stage	Processing code	Responses
CG	Visual	Searching (visual/cognitive, responding)	Verbal	Manual
	Visual	Processing (visual/cognitive)	Verbal	–
	Visual	Locating (visual/cognitive)	Spatial	–
	Visual	Working (responding)	Spatial	Manual
	Visual	Transitioning (visual/cognitive, responding)	Spatial	Manual
VH	Visual	Processing (visual/cognitive)	Verbal, Spatial	–
	Visual	Locating (visual/cognitive)	Spatial	–
	Visual	Working (responding)	Spatial	Manual
	Visual	Transitioning (responding)	Spatial	Manual
AH	Visual	Processing (visual/cognitive)	Verbal, Spatial	–
	Visual	Working (responding)	Spatial	Manual
	Visual	Transitioning (responding)	Spatial	Manual
VV	Visual	Processing (visual/cognitive)	Verbal, Spatial	–
	Visual	Locating (visual/cognitive)	Spatial	–
	Visual	Working (responding)	Spatial	Manual
	–	Transitioning (responding)	Verbal	Vocal
AV	Visual	Processing (visual/cognitive)	Verbal, Spatial	–
	Visual	Working (responding)	Spatial	Manual
	–	Transitioning (responding)	Verbal	Vocal

provided next to the gearbox for the participant to use during the maintenance task. For the AR-based system, a Microsoft Hololens 2 (Microsoft, 2019a) was used as the HMD that consists of a holographic display and user input sensors including hand-detection, a microphone, and eye-tracking to capture user inputs. Besides environment sensors, a Microsoft Hololens 2 is also equipped with a web camera that can be used for vision-based tracking in the AR system. Vuforia version 8.3.8 object tracking (PTC, 2019) was used for the AR tracking library and registration. The virtual content management, user interface, and data collections were programmed in Unity version 2019.1.10f (Unity, 2019) and interfaced with a local web and database server. To assess the impact of different maintenance conditions on the productivity, the following measures were employed:

- Performance indexes were used to measure how effective the task completed in each condition. The quality of performance was assessed in terms of completion time and the number of errors. The completion time was measured (using a stopwatch) from when the participant started to when they finished the maintenance task. Errors were recorded when the participant misinterpreted the given instructions, which were coded in five categories for ease of data collection including: 0-no error, 1-wrong tool, 2-wrong tool movement, 3-wrong tool orientation, 4-wrong part orientation, 5-other error.
- NASA-TLX (Hart and Staveland, 1988), which consists of six workload components (Mental, Physical, Temporal, Performance, Effort, and Frustration), was used to measure the subjective perception of workload. The self-rating score ranges from 0 to 100 and a raw score was computed. This measure is crucial to identify usability issues in AR systems as an increase in workload where demands exceed workload resources available could lead to a decrease in the task

Table 3
Demographic data and correlation to task performance.

Condition	N	Gender	Mean (standard deviation)			Spearman's coefficient					
			Age (yrs)	M exp.(number/5 yrs)	AR exp (number/5 yrs)	CT - Age	CT – Mexp.	CT - ARexp	Error - Age	Error - Mexp.	Error - ARexp
CG	10	M = 7; F = 3	26.6 (4.55)	219.5 (437.05)	3.6 (9.33)	−0.48	0.01	−0.18	−0.22	0.09	0.49
VH	11	M = 9; F = 2	26 (3.77)	11.45 (16)	4.72 (15.02)	0.37	−0.07	−0.42	.602*	.725*	−0.41
AH	10	M = 9; F = 1	25.2 (3.93)	11.6 (16.43)	3.9 (5.95)	0.60	−0.01	−0.08	−0.36	−0.41	−0.06
VV	10	M = 9; F = 1	27.5 (4.69)	32.5 (50.62)	2.2 (6.28)	−0.46	−0.36	0.07	−0.36	−0.11	0.49
AV	10	M = 9; F = 1	26 (4.47)	28.3 (39.08)	0.7 (1.15)	-.739*	-.779**	0.35	–	–	–

M exp: maintenance experience, AR exp: exposure to AR, CT: completion time, *$p < 0.05$, **$p < 0.01$.

performance (Jeffri and Awang Rambli, 2021), and hence, AR can lose its effectiveness when supporting knowledge-intensive work.

- Empatica E4 wristband (Garbarino et al., 2014) was used to capture continuous physiological measures of electrodermal activity (EDA) and interbeat interval (IBI) to measure levels of autonomic arousal. Physiological measures are informative for evaluating user experience and allow conclusions to be drawn regarding technology development and user adoption. A study by Stephenson et al. (2020) illustrates how the skin conductance level can inform the development of notifications for autonomous vehicle journeys in older drivers and at the same time encourage greater uptake of this technology. Furthermore, a recent systematic review by Zaki and Islam (2021) suggests that physiological measures and particular ECG derived measures are widely used tools to evaluate user experience and provide recommendations for the new systems and tools uptake. In particular root mean square of successive differences (RMSSD) is one of the most widely used heart rate variability (HRV) metrics (Aubert et al., 2003) and has been used for evaluating user technology experience in autonomous vehicles (Wintersberger et al., 2016). On the other hand, heart rate is a popular measure assessing users' in-game experiences (Drachen et al., 2010). The sampling rate of EDA was 4 Hz and it was separated to skin conductance level (SCL) and skin conductance response (SCR) using Ledalab (Benedek and Kaernbach, 2010a, 2010b) IBI data was used to calculate heart rate (HR) and heart rate variability measure represented by RMSSD via Kubios (Niskanen et al., 2004).
- Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1992) was used to measure perceived satisfaction on three sub-scales, namely system usefulness, information quality, and interface quality. Since the information quality component is usually intended for a website and software, the developed AR system was evaluated with respect to the system usefulness (6 questions) and interface quality (3 questions). Previous research found that perceived usefulness was an important factor in determining users' intention to use mixed reality technology (Yusoff et al., 2011). Further, poor design of how the interface is controlled may lead to an unnecessary secondary task (Hollnagel, 2005), and thus, impair task performance.
- Subjective assessment consisted of four questions administered to capture the participants' opinions concerning the usability of the AR system. The questions asked if participants like the use of the information and interaction mode, if they experienced difficulties during the task while being assisted by the AR system, and if the AR headset was comfortable or heavy. This assessment was
- intended to capture any external factors that may influence acceptance and ergonomics. Furthermore, for paper-based maintenance group, a following up question was asked to examine user experience while using the paper-based manual for the maintenance task in comparison to AR-based visualisation.
- An eye-tracker on the Microsoft Hololens 2 was used to record eye movements during user interaction with the AR contents or Area of Interests (AOIs). Eye movement measures were used to explore the different strategies utilised by the users to acquire the information required to perform the task and to evaluate design options. It collects data at a sampling rate of 30 Hz approximately, which results in 30 individual gaze points per second (Microsoft, 2019b). Fig. 1 shows the AOIs for the video-based and animation-based visualisations. Since the animation was overlayed on the gearbox (see Fig. 1b), and thus, virtual and real parts were overlapped alternately, the AOI for the animation was regarded as one with the gearbox. To capture the user's eye gaze behaviour, the collected raw data was converted into three common parameters, namely dwell time, number of fixations, and transition between AOIs or fixations within one AOI. While dwell time is typically used as an indicator for the level of visual engagement with the AOIs (Tullis and Albert, 2008), the number of fixations can be used to indicate the level of importance of an AOI (Poole et al., 2007) or inefficient information processing (Goldberg and Kotval, 1999) depending on the task and user context. Transition between AOIs or fixations conveys eye scanning behaviour for information searching (Goldberg and Kotval, 1999). Dwell time is registered as one visit (from entry to exit) on an AOI, which is a sum of all fixations and saccades in a specific AOI (Holmqvist et al., 2013). Fixation was computed as a minimum period of time during which the eye remains within a small area for 100 msec (Eyenal, 2007). The number of fixations is therefore calculated as the fixation count for a particular AOI. In addition, another metric, namely Fixation Rate (FR) was utilised to complement the analysis. FR is defined as the number of fixations on specific AOIs divided by the total number of fixations on the Area of Glance (AOG) and could be used as an indicator of information decoding complexity for comprehension tasks (Sharafi et al., 2015). A higher ratio may indicate that users show a greater interest in the AOI or that the AOI is difficult to interpret (Jacob and Karn, 2003). Using these metrics, it is possible to identify what and how much information is processed by the users to complete the task across different AR interfaces.

4.3. Procedure

Prior to their participation, participants read and signed a consent form that was approved by the Cranfield University Research Ethics System. Before proceeding to the experiment session, an online form was filled in by the participant to collect demographic data such as age, gender, maintenance experience, and the experience with AR. Upon arrival, the participant was equipped with a wristband sensor (Empatica E4) to record their physiological responses. The Empatica E4 wristband was placed on participants' non-dominant hand and fastened as tightly as possible without causing discomfort to the participant in order to

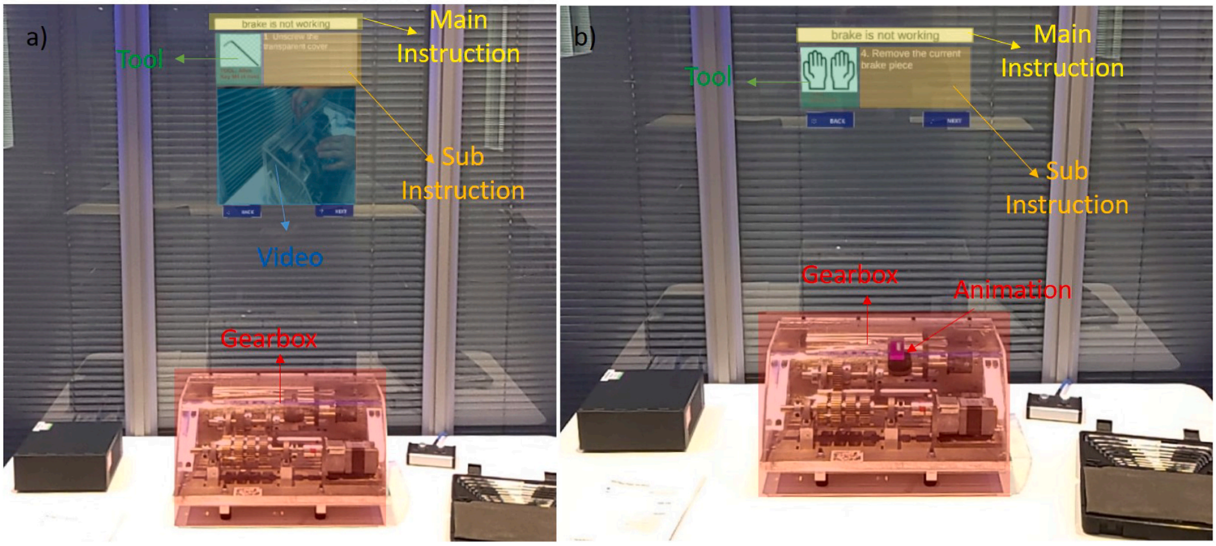

Fig. 1. AOIs in the AR-based maintenance view: a) video-based visualisation and b) animation-based visualisation.

reduce motion artifacts. The wristband also collected accelerometer data on a 3-axis accelerometer (sampling rate 32 Hz), which was later used to monitor the wrist movement of the participants. A 5-min rest position baseline was recorded before participants were asked to complete the maintenance task. After the baseline recording the participant was briefed about the experiment, asked to put the HMD on, and guided to perform eye-tracking calibration on the device. Following successful calibration, each participant was given a training session on how to use the AR device and AR interface that they would interact with during the experiment. Participants took time to familiarise with themselves with the AR system until they felt ready to proceed to the experiment. This session was given to allow the participants to familiarise themselves prior to starting the experiment, and therefore, minimising the novelty and learning effect in using the AR system.

The maintenance task was designed to represent a real-life situation in which the participant was presented with a gearbox machine that had a maintenance problem (i.e. the brake is not working, and they were asked to replace the brake). The tasks involved in the maintenance were as follows: 1) **searching** the maintenance instructions, 2) **processing** what they mean, 3) **locating** where it is applied, 4) **working** on the maintenance task (1. Unscrew the transparent cover, 2. Remove the transparent cover, 3. Unscrew the support that holds the brake wheel, 4. Remove the current brake piece, 5. Replace the brake with a new piece, 6 Tighten the support that holds the brake wheel, 7. Put back the transparent cover, 8. Tighten the transparent cover), and 5) **transitioning** from one to another instruction until the whole task is done. There were two main conditions in which this experiment was carried out. Firstly, one group of participants was asked to perform the maintenance task using a paper-based manual. The participant had to search for the relevant information within the manual (visual/cognitive and verbal as well as manual), understand the correct instructions (visual/cognitive and verbal), and locate/interpret the location of the right component on the gearbox (visual/cognitive and spatial), perform the maintenance task (responding and spatial), and move to the next instruction (visual/cognitive and spatial). This group served as a control group (CG).

Secondly, some other groups of participants were given the AR system that guided them step by step through the completion of the maintenance task. The AR system consisted of textual instructions that displayed the maintenance problem and the procedure to fix the problem. It also displayed the picture of the tool required for each step and additional textual information describing the name and the tool specification (visual/cognitive and verbal). In this case, searching task was eliminated in AR conditions.

In addition, there were two different forms of visualisation administered that assisted the participant to perceive the maintenance instruction in a more detailed and straightforward way: 1) processing video-based visualisation that showed footage of an expert using the tool to carry out the assembly/disassembly of the gearbox component step by step (visual/cognitive and spatial) where user still needed to transfer the knowledge from video to the gearbox (visual/cognitive and spatial), and 2) processing animation-based visualisation that overlaid virtual instructions and components on the top of the gearbox and the animated assembly and disassembly process right where it should be done on the gearbox (visual/cognitive and spatial). In this case, locating was eliminated in the animation-based visualisation. Both visualisations were set to always-on mode rather than on-request as suggested in previous research (Kim et al., 2019). Moreover, the AR system was also designed to have two distinct interaction modalities to navigate through the virtual instructions such as: 1) Hand gesture interaction (requires visual modality) which allows the user to use press buttons (responding and spatial) in the air, and 2) Voice command which allows the user to say some keywords such as “Back” or “Next” to control the instructions (responding and verbal). In this fashion, AR makes the maintenance task easier for the user because task-relevant information is presented directly to the user’s viewpoint right when and where it is needed, and thus, visual, and cognitive activities that involves working memory are reduced. Besides, it also facilitates a natural interface that merges cyberspace and the real world to channel direct interaction with digital information e.g. a “touchable button”. This type of interface is considered less effortful as it does not add distraction to the user in completing the primary task. Table 2 shows how the attentional resources are mapped to MRM for each maintenance condition. In all conditions, participants needed to perform a maintenance task with multiple task components. However, none of these tasks interfere, and can be completed sequentially and they tap on multiple resources. For the AR-based maintenance system, there were four groups with a combination of information mode and input modality such as *video-hand gesture (VH), video-voice command (VV), animation-hand gesture (AH), and animation-voice command (AV)*. Fig. 1 shows the user’s viewpoint and user interface while using the AR-based maintenance system. Based on the number of resource allocation, participants will perform the best in the AV condition, and the least performance improvement over paper-based manual will be in the VH condition.

After the familiarisation, the participant was asked to carry out the maintenance task in one of the conditions and their performance data was collected. The apparatus used in collecting participant data was presented in sub-section 4.2 Experiment Apparatus. The maintenance task involved a disassembly and assembly process using the combination of bare hands and some Allen keys. During the experiment, the participant was not allowed to interact with the researcher unless they encountered a technical problem.

Upon completing the task, the participant was asked to fill out the questionnaires to assess their subjective perception toward the task and the method in accomplishing the task in terms of perceived workload and usability. For the CG, after completing the maintenance task, each participant was asked to view and run through the maintenance tasks with the AR system without actually doing the task. After that, they were asked about their experience to capture their perception in terms of perceived workload and usability compared to a paper-based manual.

4.4. Data analysis

Statistical computation was performed on the data collected to determine whether there was a statistically significant impact of independent variables. In analysing the data, the comparison was first studied for all AR conditions against the control group for task performance and subjective workload. One-way ANOVA was run on the normally distributed data, and a significant effect was followed by a post hoc pairwise comparison (Tukey's HSD test). In this regard, outliers and non-homogeneity of variances should not be present.

For the non-normally distributed data indicated by Shapiro-Wilk ($p < 0.05$) and the presence of outliers, logarithmic transformation was applied. Otherwise, the Kruskal-Wallis test was applied for non-normally distributed data as an alternative to one-way ANOVA if data transformation failed to achieve normally distributed data. Bonferroni correction was applied to *p* value for multiple tests. Secondly, a two-way ANOVA was run to examine the interaction between information and interaction mode. A Mann-Whitney *U* test was run to assess the main effect of the independent variables for the non-normality distributed data. Thirdly, the analysis of baseline and during task measurement for within-subject factor of physiological data (e.g. SCL, SCR, HR, RMSSD) was run using a repeated measure ANOVA and between-subject measure for different maintenance conditions. To standardise data for the between subject design, z-scores were calculated for each of the variables. As SCL, SCR, RMSSD variables were normally distributed (Shapiro-Wilk $p > 0.05$) parametric test were used. The heart rate measure was not normally distributed (Shapiro-Wilk $p < 0.001$) and therefore natural logarithmic transformation was applied to the raw scores and later z-score normalization was conducted. This resulted in normally distributed data appropriate for the parametric tests (Shapiro-Wilk $p = 0.064$). Accelerometer data was used to calculate Euclidian distance between x, y and z coordinates and compared the movement between the participant groups. However, as there was no significant difference in the Euclidian distance neither during baseline nor during task times of interest (Kruskal-Wallis $H = 4.24$, $p = 0.375$, and $H = 4.06$, $p = 0.398$, respectively) the accelerometer data was not included as a covariate in further analysis. Fourthly, a Spearman's correlation was employed to investigate the association between variables. Fifthly, for within-subject factors analysis on eye tracking data e.g. total dwell time, total number of fixations, and so on, a repeated measures ANOVA was carried out. Where the assumption of sphericity was violated, a Greenhouse-Geisser value was considered. Some additional tests will be specifically mentioned in the results to complement the analysis. SPSS v26 (IBM, 2019) was used to perform statistical analysis with statistical significance determined when $p < 0.05$, when not changed by the Bonferroni correction.

5. Results

5.1. Task performance

Performance measures were assessed in different AR modalities against a control group, namely a paper-based manual. The task completion time was affected by maintenance conditions ($F = 10.263$, $p = 0.001$). A post hoc test revealed that the completion time in all AR conditions was significantly lower than the control group as indicated in Fig. 2. However, there was no completion time difference observed between the different AR conditions. The percentage of time reduction and improvement in time variability when compared to the control group is shown in.

Table 4. In conjunction with Table 2, it is apparent that the task performance in terms of time completion was higher as the task processing demand (e.g. involving working memory) decreased following the use of AR and different modality assignment. With regards to task errors, there was a statistically significant difference in the distribution of errors across maintenance conditions ($\chi^2 = 20.201$, $p = 0.0001$). Nevertheless, not all AR conditions resulted in significant reduction of task errors over the control group as indicated by the mean rank. Fig. 3 shows that only AH and AV led to statistically error reduction compared to the paper-based manual condition ($p = 0.001$ and $p = 0.003$, respectively). Further analysis on the impact of each modality was conducted. The result shows that no statistical one-way interaction was observed between information and interaction mode ($F = 0.160$, $p = 0.692$) for task completion time. When the main effect was examined, there was a statistically significant main effect of information mode ($F = 8.410$, $p = 0.006$) but not on the interaction mode ($F = 0.440$, $p = 0.512$) toward task completion time. The average completion time for 3D animation users was 501.1 s (SD = 83.60 s), and 583.1 s (SD = 93.73 s) for video users. In the same way, main effect of information mode ($U = 129$, $p = 0.005$) was also found for number of errors as indicated by mean rank (16.95 and 25.64 for 3D animation and video, respectively) but did not show the statistically significant main effect for interaction mode ($U = 191$, $p = 0.509$).

It was noted that despite the varying backgrounds of the participants with respect to their age, maintenance experience, and AR exposure, this study did not find consistent evidence that indicated the effect of these variables on the task performance as shown by Spearman's correlation in Table 3. Since the participants were recruited from the same environment and were randomly balanced, it remains unclear as to why correlations in VH and AV were observed between performance (i.e. CT and errors) and age as well as maintenance experience.

5.2. Subjective workload

The statistical analysis of workload measures revealed that there was a statistically significant difference in subjective workload distribution between AR and control group conditions as assessed by NASA-TLX measure in Mental ($\chi 2 = 9.96$, $p = 0.041$) and Temporal workload ($\chi 2 = 10.40$, $p = 0.034$), but not in Physical ($\chi 2 = 7.11$, $p = 0.13$), Effort ($\chi 2 = 5.75$, $p = 0.218$), Performance ($\chi 2 = 1.97$, $p = 0.741$), Frustration ($\chi 2 = 7.49$, $p = 0.112$) and total workload component. ($\chi 2 = 8.16$, $p = 0.086$). Following the statistically significant results, post hoc comparison did not reveal any statistically significant difference across maintenance conditions for Mental and Temporal workload. Furthermore, the workload measures were analysed to examine the effect of different information and interaction mode. Fig. 4 shows workload distribution of NASA-TLX. The statistical analysis did not reveal a statistically significant difference of workload changes for information mode in Mental ($U = 232.5$, $p = 0.752$), Physical ($U = 242$, $p = 0.577$), Temporal ($U = 177$, $p = 0.273$), Performance ($U = 255$, $p = 0.369$), Effort ($U = 195$, $p = 0.523$), Frustration ($U = 224.5$, $p = 0.906$), and Total ($U = 236$, $p = 0.687$) or interaction mode in Mental ($U = 272.5$, $p = 0.188$), Physical ($U = 290$, $p = 0.079$), Temporal ($U = 281.5$, $p = 0.121$), Performance (U

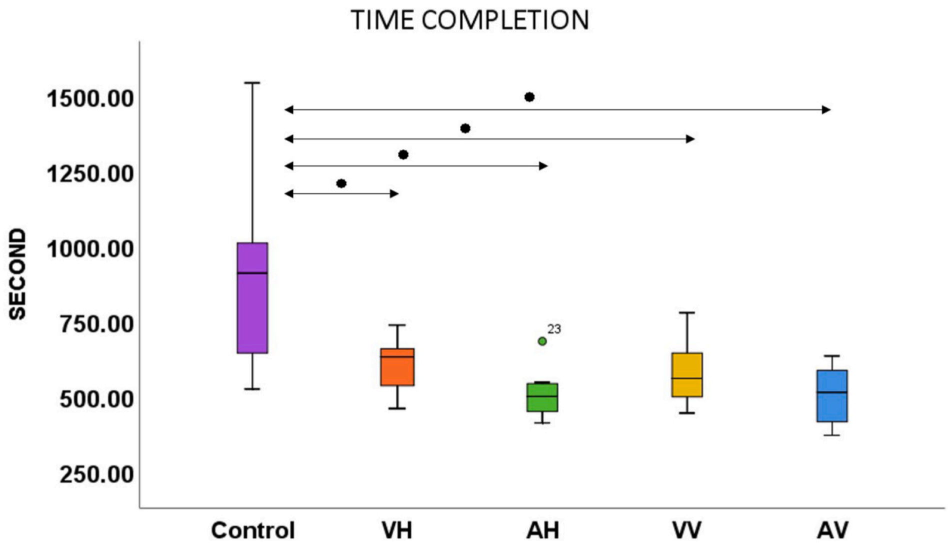

Fig. 2. Task completion time for different maintenance conditions. For an outlier in the data (23), log transformation was applied. *$p < 0.05$.

Table 4
Percentage improvement over paper-based maintenance.

	Improvement over control group	Improvement Variability
VH	31.8	71.8
AH	42.7	75.1
VV	35.1	63.9
AV	43.1	68.3

= 217, p = 0.929), Effort (U = 267, p = 0.235), Frustration (U = 231, p = 0.782), and Total (U = 277.5, p = 0.151).

A paired-sample T-test was performed on the workload data for CG after the completion of a paper-based manual and run-through with AR-based maintenance system. The results showed that workload components from paper-based were significantly lower than AR-based for Mental (t = 3.272, p = 0.011), Physical (t = 2.828, p = 0.022), Effort (t = 4.316, p = 0.003) and Total (t = 4.035, p = 0.004), but not for Temporal (t = 1.335, p = 0.219), Performance (t = 1.892, p = 0.095), and Frustration (t = 1.348, p = 0.214). Moreover, correlation analysis between task performance and each workload component in NASA-TLX was carried out. For each group of maintenance condition, there was no association established between components of task performance and perceived workload components of NASA-TLX.

5.3. Physiological data

Physiological arousal data consisted of 46 participants, as due to technical issues, 5 participants did not have the complete physiological measures data sets. Each experimental group had 9 participants each, and control group had 10. To assess if different maintenance conditions elicited different autonomic arousal, a mixed ANOVA was run for within-subject factor (baseline vs. task) and between-subject factor (maintenance conditions) on each dependent variable of SCL, SCR, RMSSD and HR. Statistical results revealed that there was no one-way interaction between within- and between-subject factor for SCL, SCR, RMSSD, and HR (F = 1.530, p = 0.210, F = 0.44, p = 0.778, F = 0.14, p = 0.246, F = 2.45, p = 0.061, respectively). The main effect of within-subject factor was significant for SCL, SCR, and HR where task related physiological response was significantly higher than baseline (F = 100.77, p < 0.001, F = 169.85, p < 0.001, F = 131.56, p < 0.001 respectively), but not for RMSSD (F = 0.420, p = 0.520). The main effect of between-subject was not statistically significant for SCL, SCR, RMSSD,

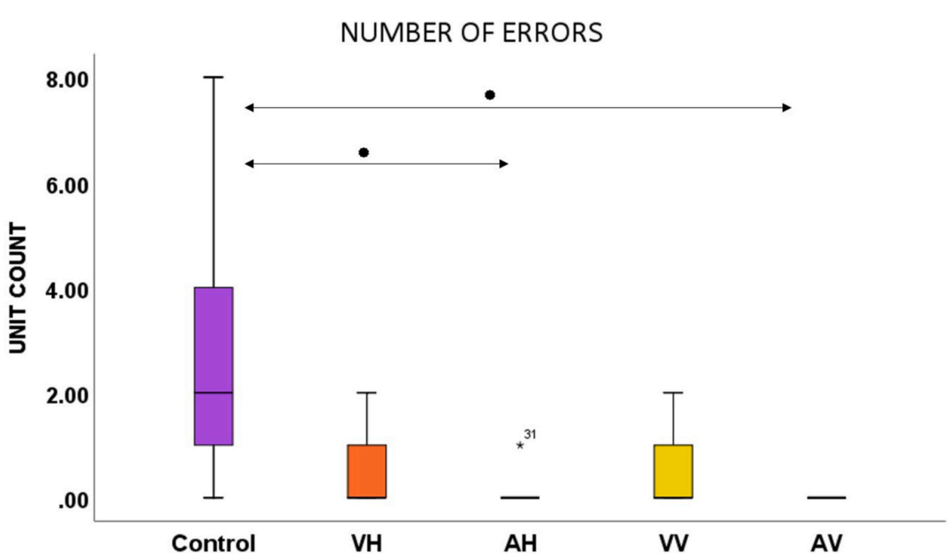

Fig. 3. Number of errors for different maintenance conditions. *$p < 0.05$.

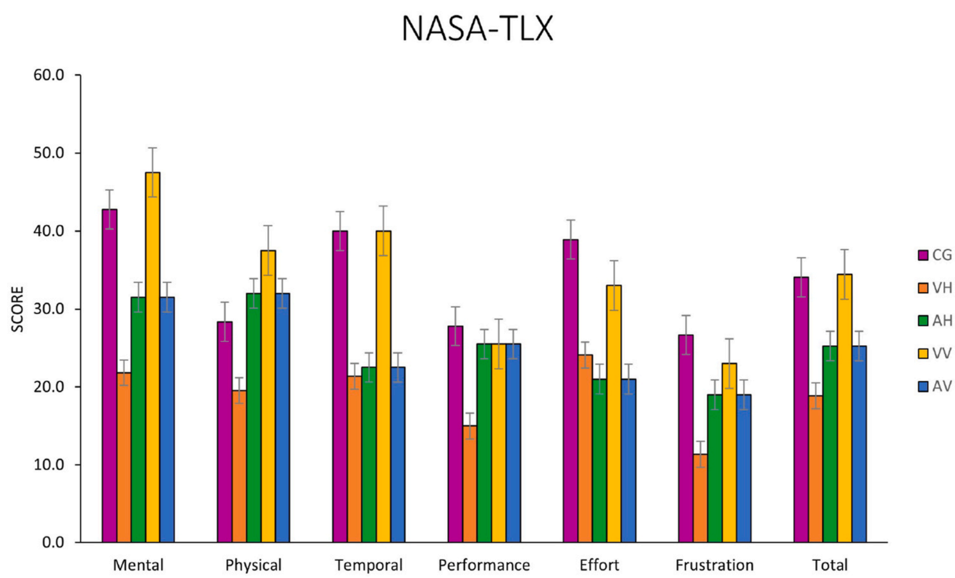

Fig. 4. Distribution of perceived workload across conditions.

and HR ($F = 1.09$, $p = 0.373$, $F = 1.360$, $p = 0.264$, $F = 2.330$, $p = 0.073$, $F = 1.870$, $p = 0.135$).

To further analyse if there was a difference in the amount of autonomic arousal increment across maintenance conditions, the measure of physiological response difference [task response – baseline] as shown in Table 5 was analysed with a one-way MANOVA for SCL, SCR, RMSSD and HR. The analysis showed a significant result only for the difference in HR ($F = 2.65$, $p = 0.047$), but not for SCL, SCR, RMSSD ($F = 1.6$, $p = 0.185$, $F = 0.18$, $p = 0.948$, $F = 1.42$, $p = 0.246$). Post-hoc test revealed only a significant higher of HR for AH compared to CG ($p = 0.048$).

5.4. Gaze behaviour

Data from eye movement was obtained for each AR condition. However, due to unforeseen technical issues, the server failed to log eye movement data from some participants. This is probably due to a failure in the calibration process which makes the Hololens 2 unable to detect the eye movements (Microsoft, 2019b).

Table 6 shows the total sample of participants with regard to data availability for statistical analysis. Since the number of samples of VV minus the number of groups is equal to zero, it results in insufficient residual for degree of freedom for conducting an analysis of variance. Furthermore, samples in different groups of information mode variables were compared to examine if there is any meaningful comparison to be gained from the analysis. Fig. 5 and Fig. 6 show similar distributions between interaction mode across information mode samples for total dwell time, number of fixations, and transition between AOIs.

Considering the similarities, more meaningful information of eye gaze behaviours would be gained by analysing differences between distinct information mode while keeping interaction mode constant (VH vs. AH).

For the VH condition, the analysis revealed a statistically significant difference for total dwell time ($F = 34.05$, $p = 0.001$). Post hoc analysis showed that there was a statistically significant difference ($p = 0.002$) between video (V) and main-instruction (M), video and sub-instruction (Sb) ($p = 0.002$), video and tool (To) ($p = 0.003$) as well as gearbox (Gb) and main instruction ($p = 0.015$), gearbox and sub-instruction ($p =$

Table 6
Eye-tracking data availability in each AR condition.

AR conditions	Total samples	Data availability	Analysis of variance
VH	11	6	YES
VV	10	4	Insufficient residual for degree of freedom
AH	10	9	YES
AV	10	10	YES

Table 5
Physiological response across conditions.

Conditions	*SCL (Skin Conductance Level)*			*SCR (Skin Conductance Response)*			*RMSSD (root mean square of successive differences)*			HR (Heart Rate)		
	T	B	D	T	B	D	T	B	D	T	B	D
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
AH	0.57 (0.32)	−0.92 (0.38)	1.49 (0.62)	0.46 (0.20)	−0.60 (0.20)	1.06 (0.34)	0.33 (0.49)	0.22 (0.66)	0.11 (0.92)	0.80 (0.34)	−0.93 (0.32)	1.73 (0.55)
AV	0.46 (0.40)	−0.75 (0.61)	1.22 (0.94)	0.35 (0.38)	−0.48 (0.38)	0.83 (0.59)	0.49 (0.69)	0.46 (0.56)	0.16 (0.54)	0.63 (0.45)	−0.94 (0.29)	1.57 (0.70)
VH	0.39 (0.46)	−0.32 (0.66)	0.71 (1.03)	0.43 (0.29)	−0.57 (0.29)	1.01 (0.45)	0.21 (0.68)	0.15 (0.20)	0.06 (0.70)	0.65 (0.29)	−0.59 (0.40)	1.24 (0.64)
VV	0.65 (0.36)	−0.89 (0.65)	1.54 (0.84)	0.38 (0.84)	−0.44 (0.18)	0.82 (0.41)	0.46 (0.66)	0.48 (0.30)	−0.18 (0.42)	0.51 (0.54)	−0.72 (0.60)	1.23 (1.05)
CG	0.36 (0.22)	−0.69 (0.45)	1.08 (0.53)	0.29 (0.53)	−0.60 (0.27)	0.92 (0.55)	0.67 (0.58)	1.02 (0.87)	−0.49 (0.69)	0.15 (0.29)	−0.54 (0.54)	0.70 (0.77)

T: Task B: Baseline D: Difference.

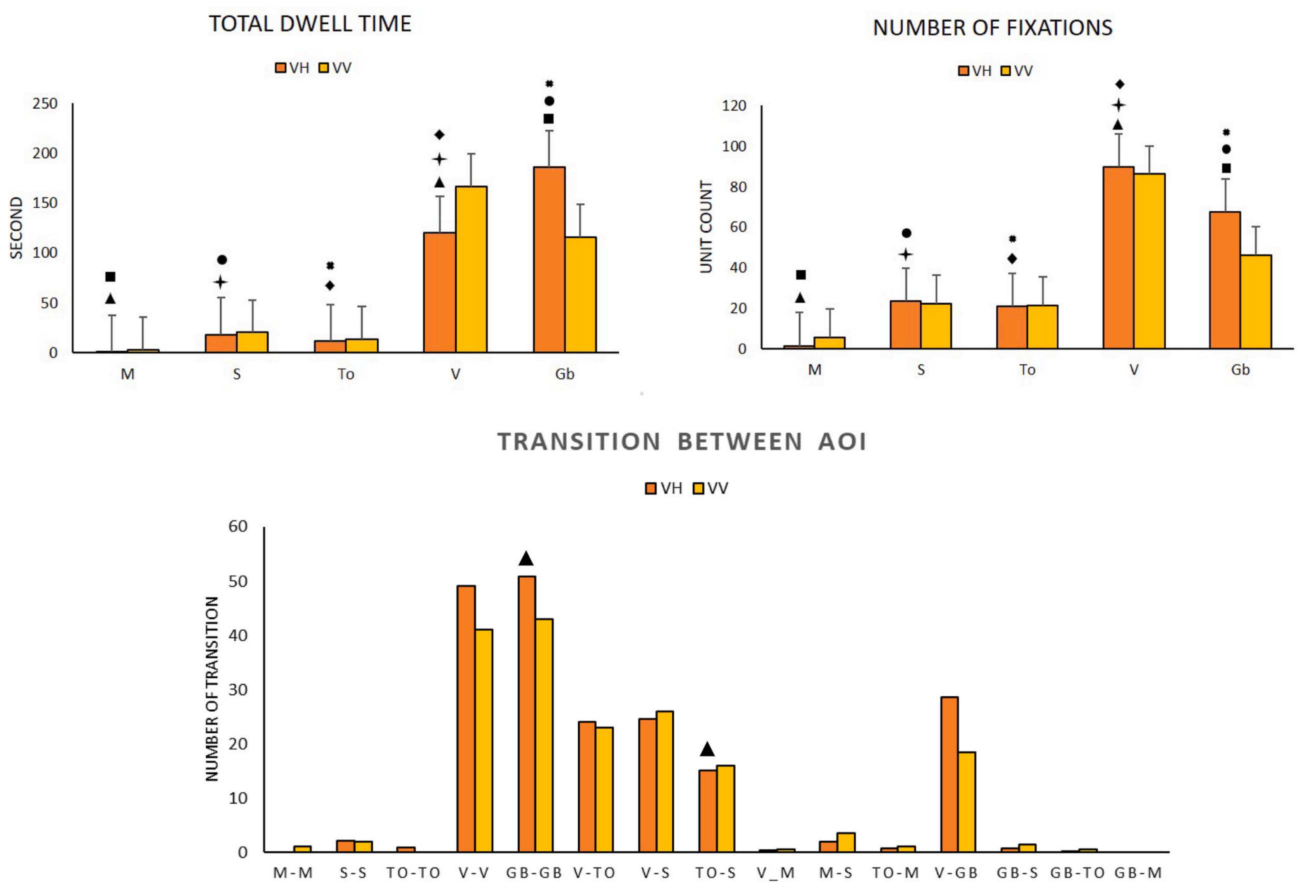

Fig. 5. Eye gaze behaviours of Video-Hand gesture (VH) and Video-Voice command (VV): a) Dwell time, b) Number of Fixations, and c) transition between AOIs or fixations. The symbols indicate the statistically significant differences.

0.022), gearbox and tool (p = 0.021), but not between video and gearbox (p = 0.775). The analysis also elicited a statistically significant difference for the total number of fixations (F = 35.44, p = 0.001). Post hoc test showed that there was a statistically significant difference between video and main instruction (p = 0.005), video and sub-instruction (p = 0.016), video and tool (p = 0.009) as well as gearbox and main instruction (p = 0.006), gearbox and sub-instruction (p = 0.035), gearbox and tool (p = 0.012), but not between video and gearbox (p = 1.0). These results show that user engagement during the AR-based maintenance with video-based visualisation was mainly placed on video (35.68%) and gearbox (55.18%) rather than textual instructions (5.64%) and 2D image (3.5%). Users also glanced more on video (44.17%) and gearbox (33.25%) in gaining information for the completing the task in comparison with textual information (12.32%) and 2D image (10.26%). Analysis into user scanning behaviour revealed information search between fixations or AOIs was focused on Gearbox-Gearbox (Gb-Gb) followed by Video-Video (V–V), Video-Gearbox (V-Gb), Video-Sub-instruction (V–S), Video-Tool (V-To), and Tool-Sub-instruction (To-S). Statistical analysis showed there was a significant difference in the number of transitions for highly transitioned area or fixations (F = 6.34, p = 0.001). A post hoc test showed the transition counts between Gb-Gb and To-S were statistically significant different (p = 0.024) whereas no statistically significant difference was observed for other areas of eye movement.

In the same manner, there was a statistically significant difference for total dwell time (F = 93.21, p = 0.001) in the AH condition. The post hoc test showed the total dwell time for Gb was statistically significantly higher than M (p = 0.009), S (p = 0.013), and To (p = 0.012). Similarly, there was a statistically significant difference for the number of fixations for different AOIs (F = 225.53, p = 0.001). Post hoc analysis eliciting the number of fixations of Gb was statistically significantly higher than M (p = 0.0001), Gb and S (p = 0.0001), Gb and To (p = 0.0001). This implies that users in the AH condition were more concentrated on the information available on the gearbox (i.e. animation, accounts for 85.3%) compared to textual instructions (11.08%) and 2D images (3.61%) to complete the task while being assisted by the AR system. There was also evidence from the fixation count that users paid more attention to information displayed on the gearbox (58.07%) than textual instruction (28.22%) and 2D image (13.7%). The analysis of eye scanning behaviours demonstrated the predominant information search was centred on transition between Gb-Gb fixations, followed by To-S, S–S, Gb-S, and Gb-To. Statistical analysis showed a significant difference among transitions (F = 74.68, p = 0.0001). The post hoc test revealed transitions count for Gb-Gb was statistically significantly higher than To-S (p = 0.003), S–S (p = 0.001), Gb-S (p = 0.0001), and Gb-To (p = 0.0001). Further, To-S and Gb-S was statistically significantly higher than Gb-To (p = 0.042 and 0.027, respectively).

Since both information modes showed users' main reliance on graphical information (e.g., animation and video) rather than textual cues during maintenance task, further analysis was carried out to investigate different perceptual operations between VH and AH conditions. To examine the comparison for both conditions, statistical analysis considered the primary AOIs, namely video and gearbox for VH and gearbox for AH upon which users fixated their eye most often during the maintenance task. In terms of total dwell time, there was no statistically significant difference (F = 0.01, p = 0.925) between video + gearbox (90.86%) in VH and gearbox (85.3%) in AH. On the other hand, the analysis showed there was a statistically significant difference with respect to the number of fixations (F = 24.57, p = 0.001) between video + gearbox (77.42%) in VH and gearbox (58.07%) in AH. Furthermore,

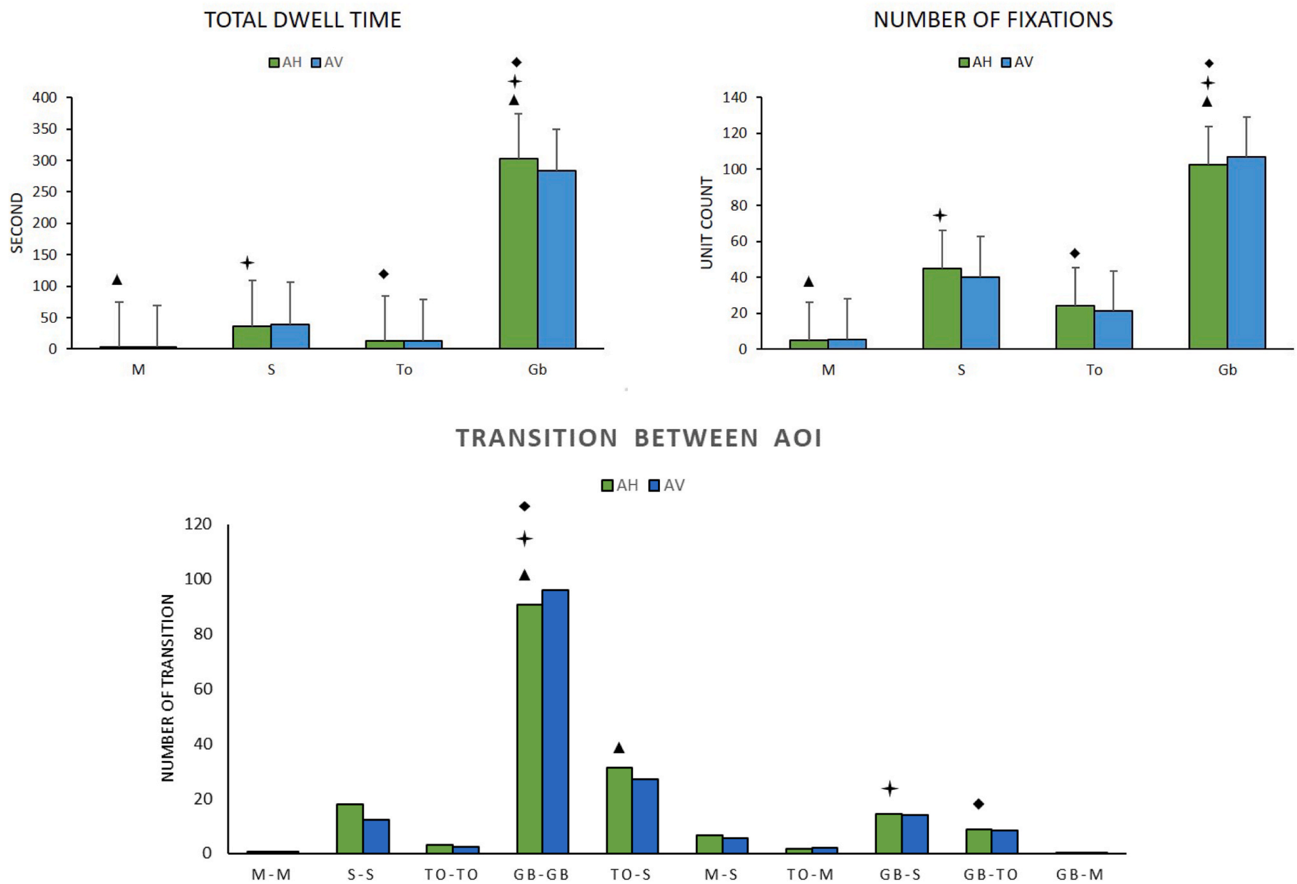

Fig. 6. Eye gaze behaviours of Video-Hand gesture (VH) and Video-Voice command (VV): a) Dwell time, b) Number of Fixations, and c) transition between AOIs or fixations. The symbols indicate the statistically significant differences.

the analysis of FR elicited a statistically significantly high ratio of AOIs for video + gearbox in VH compared to AOI for gearbox in AH condition ($F = 38.65$, $p = 0.0001$).

5.5. Correlation between Eye movement and performance

In order to find the association between participants' perceptual strategies in different AR-based visualisations and task performance, correlation analysis was conducted on the eye movement metrics (e.g. total dwell time, total number of fixations, FR) and performance metrics (e.g. completion time and errors). The total number of fixations from the sum of video and gearbox fixation counts elicited by participants in the VH condition was statistically positively correlated with the time taken to complete the task ($r_s = 0.841$, $p = 0.036$). Similarly, the total number of fixations on the gearbox elicited by participants in the AH condition positively correlated with task completion time ($r_s = 0.717$, $p = 0.03$). This implies that the more AR modality could facilitate users to integrate required information for task completion within the vicinity (higher number of fixations in the specific AOI, 77.42% (video + gearbox) in VH and 58.07% (gearbox) in AH), the better the task performance would be achieved in terms of task completion time.

5.6. Perceived usability

In regard to the usability, the use of PSSUQ was not suitable for the control group due to the absence of an interface. Therefore, the analysis was made for AR conditions with respect to system usefulness and interface quality. The results showed that there was no one-way interaction between information mode and input modality on the measure of system usefulness ($F = 2.292$, $p = 0.139$) and interface quality ($F = 0.244$, $p = 0.625$). Further, both system usefulness and interface quality were not influenced by main effect of information mode ($F = 2.814$, $p = 0.102$ and $F = 1.638$, $p = 0.209$) and input modality ($F = 1.866$, $p = 0.178$ and $F = 0.066$, $p = 0.799$). Furthermore, a two-way MANOVA was run to examine the impact of independent variables on combined variables of PSSUQ subscales. The result did not show any statistical effect of AR modalities on a combined metric of system usefulness and interface quality as shown in Table 7.

5.7. Subjective assessment

Subjective perceptions were collected for all AR conditions (VH, AH, VV, AV). Concerning the video-based visualisation, almost all (20 out of 21) participants said that the video instruction was simple, easy to follow, intuitive, and effective as if having someone giving practical examples. Only one participant said: *"no, because you can't stop or pause the video"* whereas one participant who liked the video instructions said: "*Yes, because they were visual and easy to follow (copy). Also, I appreciated the fact that they were played in loop*". Some other caveat of video-based visualisation included the limitation of the single perspective. One participant said: "*The video was useful, but another view is required from the*

Table 7
Two-way MANOVA analysis.

	IM	II	IM X II
TWO-WAY MANOVA	F = 1.454 (p = 0.247)	F = 1.652 (p = 0.206)	F = 1.201 (p = 0.313)
(System usefulness & Interface Quality)	Wilks' Λ = .925	Wilks' Λ = .916	Wilks' Λ = .937

IM: Information Mode II: Input Interface.

front perspective because one instruction took more time to identify the location of the bolts". Another limitation involved the need to switch attention between the video and the mechanical component. One participant said: "*it could be better and more direct if instead of video, the system could highlight the action directly on the hardware. Using video adds time because I need to re-oriented the view*". Another said: "*the screen is a little small so the application and video is not complete in my sight, I have to look up and down to see the whole screen"*. Furthermore, there was a safety implication from using the video during the maintenance task. Researchers observed that participants often looked at the video instead of the machine while doing the assembly and disassembly process. As one participant noted: "*I did not face any difficulty during the test, however I realize a situation when I tried to remove the wheel part, I was focused on the video too much that I almost got my finger slightly stuck in between the parts, which I suppose is a health and safety concern if working with high risk working environment*".

In the same fashion, all participants in the animation group (20 out of 20) found that the animation-based visualisation is easy to understand, helps to quickly spot the right part and identify the correct orientation, and shows directly what is needed to do. Only three participants noted the misalignment of the animation overlaid on the machine, yet still they admitted that the instructions were useful and understandable. One participant said: "*Yes - it made finding the required parts easier. However, the animation was not in the correct position/orientation during the 'removing the brake'' phase. The system showed me the shape of the part however, so I could make an educated guess as to where it actually is*". Some other drawbacks found in the current design of the animation-based visualisation were related to occlusion and brightness. One participant said: "*Somewhat, it makes it far easier to find a certain part, however there should be an option to turn it off after the part is found to avoid distraction*" and another said: *"When re-connecting the brake, I wanted to look closely to find the holes for the screws, however the instructions were showing a large purple block which was bright and therefore made it difficult for me to see the holes beneath"*. Regarding the brightness, one said: *"The animation showing the step was a bit of lack of the brightness or it was a bit too transparent on the first task, however the pictures and animation for steps after were clear and bright*". The other one said: "*The lightning was slightly low when looking at the gear box to see the task required to operate"*. This highlights that occlusion, improper brightness of the animation (i.e. poor visibility), and inaccurate overlay could lead to confusion and place a cognitive demand on understanding the required task.

For the type of interaction mode, all participants (21 out of 21) liked the hand gesture to navigate through the instructions. Hand gesture was perceived as easy to use, intuitive, and novel. One participant mentioned:" *The hand gesture was well detected and allowed me to easily use the system's functions, such as the menu*". Nonetheless, the limitation of this type of interaction mode concerns the correct position of the user's fingertip on the button, which requires users to readjust their positions and do multiple clicks. As one participant noted:" *Yes, the hand gestures were easy to do and you couldn't do them by mistakes; however, to tap the button it seems I thought the image was a bit further than it was, so I often tap just behind the button instead of on the button"*. This type of interaction would be especially difficult for those who have depth issue in their visual system. One participant added: "*the hand gesture is good, but I had some problems with the depth when trying to touch the buttons*". The other downside of hand gestures concerns the limitation of navigating the menu when both hands are occupied as one participant said: "*when my hand is occupied, it is faster to use voice command*". Although, the limitations do not seem to pose adverse impact on occupational safety, the use of an air press button might be restricted to a wide workspace area in which no hindrance of physical objects are present in the user's viewpoint. The workaround for hand gesture might be the use of an air tap that makes use of the user's gaze point as a cursor and air tap like a mouse click. Another issue with hand gesture is related to ergonomics. For a task that is physically demanding, hand gestures could be tiring because users need to raise their hand to interact with the AR contents. For this reason, some users preferred voice command over hand gesture, as one participant stated: "*I like the voice command. It reduces the work of lifting and lowering the hand to select obvious options while performing the task*".

All participants in the voice group (20 out of 20) said they liked voice command as an interaction modality. Generally, they felt it was easy to use and handy especially when both hands are busy and could help them finish the task quickly, as one mentioned: "*Yes. Using my voice to control the steps helps me to concentrate on the main task*. The main hindrance of using voice command could be ambient noise, which may reduce the effectiveness of this interaction mode. One participant said:" *Voice command is useful and it's great that is not necessary to calibrate my voice, however some maintenance tasks are made in very loud environments, I don't know how it will react*". The workaround would be to make both hand and voice interaction modes available to deal with the surrounding noise.

Finally, regarding the AR headset, almost all AR participants (38 out of 41) said that the Microsoft Hololens 2 was comfortable and not too heavy. But, since they were using it only for 7–8 min on average, most of them questioned if it would be comfortable for prolonged use. Some other aspects that need further improvement according to the participants are Field of View, heated surface after long usage, and rendering quality.

6. Discussion

The purpose of this study was to investigate the impact of different AR modalities concerning commonly used types of information mode (i.e. video vs. 3D animation) and interaction modes (i.e. hand gesture and voice command) for AR HMD systems in manufacturing and maintenance applications.

6.1. How is AR task performance affected by different types of information mode (i.e. video vs. 3D animation) and interaction mode (i.e. hand gesture vs. voice command)?

The findings show that although all AR conditions significantly improved task performance in terms of completion time over a traditional paper-based manual, the 3D animation-based instruction show a significantly faster completion time (14%) compared to a video-based visualisation. This result differed from the findings observed by Morillo et al. (2020) and Loch et al. (2016) who found no time saving difference between video and 3D animation for assembly tasks. However, they used different AR devices such as a smartphone and a screen in their study, which might contribute to the different finding. The difference in the performance gain might be explained by the minimum amount of attention switching from the instruction to the gearbox in the animation-based instructions using HMD. Since animation was overlaid on top of the gearbox, participants could directly see the location of the right component and the procedures to perform the task as shown in Fig. 1b. On the contrary, although video-type instruction was displayed in close proximity to the gearbox, they would still be unable to see the video and the gearbox at the same time and have to switch their attention to match what they were looking at with what they were having to do, which was analogous to using smartphone or looking at the screen. The effect of attention switching to task completion time was evident when comparing an AR tablet and AR HMD in the complex assembly task (Hoover et al., 2020; Wang et al., 2019). The impact of attention switching could also be observed in the study that compared video-annotated animation with traditional video tutorial that found no significant differences in completion time and errors (Yamaguchi et al., 2020). Although, the users found the annotated animation helpful in identifying the corresponding real part, the task performance remained similar which is likely due to both visualisations are screen-based and require attention switch between instruction and the real part while performing the task.

In regard to the interaction mode, the study did not observe any significant performance gain between hand gesture (i.e. air press button) and voice command in navigating the instructions for the given task. It is likely that a difference would be apparent when the task involves a highly physically demanding task, which requires both hands in the operation as noted in the subjective feedback. Furthermore, this study also found that the number of task errors was significantly reduced when using AR compared to paper-based manuals but only for the 3D animation-based visualisation. It was also found that participants who used video-based instructions made significantly more mistakes than those who experienced 3D animation-based instructions. This finding is consistent with the previous study (Loch et al., 2016) who found the same deficiency of video-based instructions compared to animation using screen display for AR systems. Similar to task completion, no effect of distinct interaction modes was found for the number of errors. Taken all this, this study found that the interaction mode did not affect the performance gain whereas type of information mode caused variations in the task performance (i.e. time completion and accuracy). In the light of MRM (see Table 2), although transitioning with hand gesture demanded more resource allocations (visual modality, spatial processing code, and manual responses) than voice command (verbal processing code, and vocal responses), this did not sufficiently manifest in performance benefits. However, concerning the visualisation mode, AR eliminates the number of task components from using manual (searching task) and from modality assignment (locating task reduced attention switch in animation), which contribute to increased task performance. With regard to considering the interaction between the AR HMD and the participant as a JCS, these findings indicate that the 3D animation overlaid on the real component better supported the participant with the information provision aspect of the time management principle.

6.2. Do AR instructions lead to lower workload than a paper-based manual?

Whilst performance increment and variation were apparent between both the AR and Non-AR approaches as well as the information mode, the measure of subjective workload did not reveal any statistical evidence on the account of workload difference. Nevertheless, the comparison of within-subject analyses for the paper-based manual showed a statistically significant difference for Mental, Physical, Effort, and Total workload after they experienced the AR assistance systems. The difference in the perceived workload could be attributed to the extra efforts when using a paper-based solution compared to AR. The aspect of searching for information in the manual and interpreting how the instruction applied to the required task could place a significant demand in terms of visual/cognitive resources, and thus, more working memory allocation (see Table 2). In contrast, AR eliminates the visual search task and eases the instruction comprehension by presenting rich and relevant information for the task which, therefore, reduces the mental workload (Rehman and Cao, 2020). Furthermore, the feedback collected from the paper-based participants confirmed that the AR instructions would improve the task significantly. One participant said: "*The AR experience was amazing. It provided clear instructions and steps. Clear visualisation of the task was achieved on AR rather than a two-step method of paper-based system, which were reading and then interpreting. Overall, I would prefer AR-based systems over paper-based systems*".

The analysis of physiological arousal (SCL, SCR, and HR) indicated an increase of workload from baseline to task for all maintenance conditions. In particular, all AR conditions elicited a higher level of HR compared to a paper-based manual (Table 5) with the significant result observed only between the paper-based manual group and the Animation-Hand (AH) group. Looking at the performance data and perceived usability, they may suggest that the increased workload while using AR was neither attributable to the task-related workload since AR helps to complete the task faster and with fewer errors nor with usability issues as participants perceived that the visualisations and interactions in AR systems were easy to use and helpful. The possible reason underlying the workload increment in AR may be associated with the increased engagement to the task when using AR, which lead to higher physiological activation. This is also shown in the other study that AR HMD can provide engaging experience due to the novelty effect associated with AR that made it more effective and entertaining for solving assembly tasks (Westerfield et al., 2015). Taken all this together, this study found that the use of AR can reduce task-related workload (higher resource allocation for searching and interpreting according to MRM in Table 2), which is evident from NASA-TLX, subjective perception, and performance measures. However, AR can also induce non-task related workload as indicated by their physiological responses, which is associated with the increased engagement in solving manual tasks.

6.3. Can eye gaze behaviour uncover different performance gains between different AR modalities?

Further, it is important to examine the parts of AR visualisation where visual attention was focused, the amount of processing effort, and the information search from one location to another for different visualisation modalities. The analysis of eye movement revealed that participants spent more time interpreting and relating dynamic graphical instructions (e.g. animation and video) than textual instructions or static graphical instructions, which was indicated by the significant difference in total dwell time. This is likely due to the fact that dynamic graphical instructions are easier to understand, more detailed and straightforward. Moreover, the significant difference in the total number of fixations signifies that participants processed a larger amount of information from the dynamic graphical information. This is apparent since dynamic instructions consist of more information components than the static instructions. Nevertheless, the information search between fixations or AOIs in the video-based instruction showed that participants possibly still relied on other information such as textual instructions, pictures, and the machine to form a complete understanding of the task guidance. This is indicated by no significant difference between Video-Video and Video-tool, Video-Video and Video-SubInstruction, Video-Video and Gearbox-Video. In contrast, there was a significant difference between Gearbox-Gearbox and Gearbox-SubInstruction, Gearbox-Tool, and Gearbox-Tool in animation-based instruction, which may suggest that overlaying 3D animation directly over the gearbox comprised enough information to perform the required task. In line with MRM, users who used video still allocate resources to encode both verbal (textual instruction) and spatial information (video) to get an integrated information for accomplishing manual task. While, animation users relied only spatial information (animation on top of machine) to perform manual task.

To further understand the impact of different visualisations on visual effort, a comparison was made between 3D animation and video to unravel users' strategies in acquiring information and their impact on the task performance. The results show that the amount of time users spent looking at the video and gearbox was equivalent to the total time spent looking at the animation and gearbox. Nonetheless, the total number of fixations was significantly greater in the sum of the video and gearbox than in animation and gearbox. A possible explanation could be due to attention switching to relate information displayed by the video with the gearbox, which is associated to a higher frequency of visual scanning in the video-based visualisation. Unlike video that displayed a scene of an expert technician performing the task, 3D animation only overlaid relevant components on the gearbox, and therefore possibly, associated to less effortful visual scanning. Besides, the Fixation Rate indicated a significant higher ratio of video than animation, which may imply higher complexity in information decoding. This is probably why participants searched over other AOIs to gain more information in understanding the required task in the video-based instruction. Moreover, the significant correlation between the total number of fixations and the completion time in both conditions revealed that fixations frequency could be a good predictor of task completion time, namely those who gazed more often at the dynamic graphical instructions tend to spend more time on completing the task. This finding aligns with a previous study in assessing task assembly performance and fixation count with assistive systems (Stork and Schubö, 2010), whereby attentional guidance with the least attention switching is correlated to reduced task completion time and lower fixation counts. In addition, this study presented more evidence on the increased fixation frequency which was not only caused by attention switching due to difficulty in task memorisation (Stork and Schubö, 2010), but it can also stem from interpretational uncertainty of lead information (Poole et al., 2007) which is shown by other information search efforts (Wickens, 2008; Wickens et al., 1983). Moreover, according to the resource compatibility concept, performance benefits are better realised when the visual/manual modalities are associated with spatial encoding (animation) than when the spatial is paired with verbal encoding (text + video). Although the number of errors were not marked by any association with the eye tracking metrics used in this study, there are some evidences that eye tracking metrics (e.g. number of fixations) could discriminate different user strategies to interact with different information modes, and their impacts on the task performance.

6.4. Discussion of different implications of AR visualisation and interaction modalities

Perceived usability was not statistically significantly different across AR modalities. This could imply that participants did not have a strong preference of one modality over the others for the type of information and interaction mode. This is also evident from their feedback regarding the system's usefulness in which they found both information modes useful, easy to use, and helpful for quicker task completion. Nonetheless, both information modes have their own drawbacks. In the video-based instruction, the presentation of the information is limited to one perspective. Although, showing a recorded video of how the instruction is done is intuitive, in some cases, users might require looking at the how the task is done from a different angle to better understand the guidance. Therefore, providing the choice to the users to switch the video to a different perspective may improve the usability of this type of information mode. The other limitation was related to the attention shift from video to the task at hand, which requires head movement. In a complex and long assembly task, the increase in switching attention does not only place additional workload on the users but would also result in frequent head movements that could lead to physical ergonomic issues such as fatigue or neck pain. Further, there is a safety-related issue with the use of video as the information mode. During the maintenance task, participants tended to focus their attention on the video instead of on the component while performing the assembly and disassembly process. The reason is probably that they would like to make sure that they did not miss any important information when they were looking at the equipment while the video was running or else, they would have to re-watch the video from the beginning. This tendency could lead to the risk of occupational hazards as they might get physical injuries from the mechanical component.

On the other hand, the presentation of 3D animation right on the equipment could introduce confusion when the overlay is misaligned with the physical component. Although, most participants found no issue in comprehending the animation-based instructions, in some cases users may find it difficult to relate the guidance with the required task. In addition, some other problems found in this study were the brightness and occlusion. It was reported that occlusion of the real object by the 3D animation blocks the user view in searching for the relevant part. In other cases, improper occlusion between the virtual and the real part (e.g. due to poor brightness) may cause a misunderstanding of spatial properties, incorrect operation of the task, and increased eyes-strain (Shah et al., 2012).

In terms of interface quality, participants found both interaction modes effective and easy to use. Nonetheless, participants found that voice commands could help them concentrate on the task better and complete the task quicker if both their hands are busy. The use of hand-gesture may be tiring due to the need to raise and to lower the hand when interacting with interface. With regards to the hardware, participants found that the Hololens 2 was comfortable and not too heavy for the task duration tested. However, further research is needed to test for a prolonged usage (Wang et al., 2019). Considering all AR conditions, the findings address the fourth question, namely participants liked to use AR as a novel way of interacting with information required to perform an assembly task.

6.5. Discussion on designing AR and the user as a joint cognitive system

The results indicate the need to address AR usability issues from a Joint Cognitive System (JCS) perspective (Hollnagel, 2005). JCS is concerned with the co-agency between human and the artefact (i.e. AR system) in achieving goals and objectives. This achievement is assessed through three main principles: supporting the coping strategies employed by individuals to effectively do their job, supporting the users' time management through the provision of clear information and easy to use interaction mechanism, and by providing adaptive systems that enhance users anticipation in coping strategies. This research has looked to address the first two principles, by showing that an AR system can be designed to improve human capabilities to meet the task demand quicker and more accurately. However, the current system implementation was not equipped with the decision support capabilities necessary to provide adaptive system support. A concern with providing this level of support is that AR may lead human users to misguided information and even overreliance on the system capability. The inability of the system to recognise the information and action selection of the human user and to provide adaptive feedback accordingly, could hinder a user's understanding of the whole process necessary to take the right action. To achieve the full realisation of the JCS in AR, a system comprehension of human cognitive and perceptual operations during the task would be required and would involve visual scanning behaviour and allocation of attentional resources. MRM could help in framing the resource allocations and predict the performance outcomes with respect to the modalities assignment, type of processing code, and responses. Based on these insights, the AR system can then be designed to present stimuli that enhance a user's ability to anticipate and understand the outcome of their selected actions to meet the predictability principle of JCS in AR.

6.6. Limitations

Finally, there are some limitations that need to be considered when interpreting the findings in this study. Firstly, the maintenance tasks tested in this study were relatively simple and limited, consisting of the use of an Allen key to (un)screw bolts and simple hand movements to perform the assembly and disassembly process. It is therefore questionable whether this result could be generalised for a more complex assembly task such as an aircraft cable assembly task. In the current study, a slight misalignment of 3D animation was still comprehensible for the AR users, but the question arises whether this would still be the case with assembly tasks that require more precision in terms of tracking and registration of virtual objects to real components such as in spacecraft cable assembly. Further research should investigate the effect of imprecise overlay of AR on complex assembly tasks and the optimal type of information mode. Secondly, the task duration and environment considered in the present study may not be sufficient to reflect a real industrial setting that can be noisy, with poor lighting as well as exposure to time pressure. These factors may have impact on the efficacy of information and interaction mode.

7. Conclusion and further research

This paper offers new design implications in the choice of information mode for AR HMD systems for assembly and disassembly tasks in manufacturing and maintenance applications. The findings show that 3D animation could help to achieve faster task completion times and fewer errors in guiding task assembly compared to video-based instructions. The advantage of using 3D animation is more favourable to users' perceptual system since it allows direct visualisation on the equipment and minimises the need to shift attention when interpreting and relating to the required task. Nevertheless, the use of 3D animation could cause confusion to the users and possibly, provide misguided information when the registration is misaligned, and the brightness and occlusion are not appropriately designed. Regarding the video-based instruction, although participants indicated a high acceptance of video usage, the eye movement analysis showed inefficient scanning behaviour when processing the information, which is likely, contribute to task performance. Furthermore, video usage could introduce occupational hazards if the user is not warned to focus their attention on the equipment while doing the required task. Concerning the interaction mode, voice command is perceived as useful when the task at hand requires the use of both hands though it might not be effective in a noisy environment. Besides, it could alleviate physical strain when the maintenance task is physically demanding. Therefore, multi-modal input that include eye tracking, hand gesture, and voice recognition could help users to accomplish specific tasks better based on their preference (Dünser et al., 2007).

Finally, this paper has summarised existing user-based studies and has contributed to new design practices for AR interfaces on HMD systems in maintenance and manufacturing applications. Unlike the 2D desktop environment, there is currently no established design guideline or interaction metaphor for designing effective 3D interfaces like AR, particularly for industrial usage. As noted, user-based studies are critical in driving design activities as they help to form a collection of tried-and-true guidelines which will eventually become standards for effective AR design (Gabbard and Swan, 2008). This incremental domain-specific approach is critical to gain a clear understanding of how design aspects affect the capabilities of the user and provide users with support in coping with complex environments and enhance their capabilities prior to generalisation to wider use (Hollnagel, 2005). MRM could help in framing the human cognitive resource allocations and predict the performance outcomes with respect to the modalities assignment, type of processing code, and responses (Wickens, 2002, 2008). Future research should investigate some other usability issues that maximise efficient visual scanning behaviour, address over-reliance on AR information, and other ergonomic issues that might put the users at risk of occupational hazards. It also important to address the use of AR for a wide range of users including those who experience vision distance related problems and slow hand-eye coordination to ensure inclusive implementation of AR in the long term. From the technological side, the AR tracking and registration were still not without flaws even in the laboratory setting. Therefore, the robustness of AR tracking and registration is also critical and should be improved, especially for the implementation in the industrial contexts where the environment is relatively more complex. It is also questionable to what extent the design practices are transferable across different applications (e.g. medical). Further investigations in this area could help to clarify and accelerate the formation of design standards for AR systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the EPSRC as part of the "Digital Toolkit for optimisation of operators and technology in manufacturing partnerships" project (DigiTOP; https://digitop.ac.uk; EP/R032718/1). Data underlying this study can be accessed through the Cranfield University repository at https://10.17862/cranfield.rd.15195594.

References

Ahn, J. gil, Ahn, E., Min, S., Choi, H., Kim, H., Kim, G.J., 2019. Size perception of augmented objects by different AR displays. Commun. Comput. Inf. Sci. 1033, 337–344. https://doi.org/10.1007/978-3-030-23528-4_46.

Akçayir, M., Akçayir, G., Pektaş, H.M., Ocak, M.A., 2016. Augmented reality in science laboratories: the effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories. Comput. Hum. Behav. 57, 334–342. https://doi.org/10.1016/j.chb.2015.12.054.

Angeli, C., 2013. Examining the effects of field dependence-independence on learners' problem-solving performance and interaction with a computer modeling tool: implications for the design of joint cognitive systems. Comput. Educ. 62, 221–230. https://doi.org/10.1016/j.compedu.2012.11.002.

Aubert, A.E., Seps, B., Beckers, F., 2003. Heart Rate Variability in Athletes, Sports Medicine. https://doi.org/10.2165/00007256-200333120-00003.

Benedek, M., Kaernbach, C., 2010a. Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47, 647–658. https://doi.org/10.1111/j.1469-8986.2009.00972.x.

Benedek, M., Kaernbach, C., 2010b. A continuous measure of phasic electrodermal activity. J. Neurosci. Methods 190, 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028.

Bottani, E., Vignali, G., 2019. Augmented reality technology in the manufacturing industry: a review of the last decade. IISE Trans. 51, 284–310. https://doi.org/10.1080/24725854.2018.1493244.

Braly, A.M., Nuernberger, B., Kim, S.Y., 2019. Augmented reality improves procedural work on an international space station science instrument. Hum. Factors 61, 866–878. https://doi.org/10.1177/0018720818824464.

Dey, A., Billinghurst, M., Lindeman, R.W., Swan, J.E., 2018. A systematic review of 10 Years of Augmented Reality usability studies: 2005 to 2014. Front. Robot. AI 5. https://doi.org/10.3389/frobt.2018.00037.

Drachen, A., Yannakakis, G., Nacke, L.E., Pedersen, A.L., 2010. Correlation between Heart Rate, Electrodermal Activity and Player Experience in First-Person Shooter Games, pp. 49–54. https://doi.org/10.1145/1836135.1836143. Proc. - Sandbox 2010 5th ACM SIGGRAPH Symp. Video Games.

Dünser, A., Grasset, R., Seichter, H., Billinghurst, M., 2007. Applying HCI principles to AR systems design. Proc. 2nd Int. Work. Mix. Real. User Interfaces Specif. Authoring, Adapt. (MRUI '07) 37–42.

Endsley, T.C., Sprehn, K.A., Brill, R.M., Ryan, K.J., Vincent, E.C., Martin, J.M., 2017. Augmented reality design heuristics: designing for dynamic interactions. Proc. Hum. Factors Ergon. Soc. 2017-Octob. https://doi.org/10.1177/1541931213602007, 2100–2104.

Eyenal, 2007. Eye Tracker System Manual ASL EyeTrac 6 EyeNal Analysis Software. Data Process.

Fernández del Amo, I., Erkoyuncu, J., Vrabič, R., Frayssinet, R., Vazquez Reynel, C., Roy, R., 2020. Structured authoring for AR-based communication to enhance efficiency in remote diagnosis for complex equipment. Adv. Eng. Inf. 45, 101096. https://doi.org/10.1016/j.aei.2020.101096.

Funk, M., Bachler, A., Bachler, L., Kosch, T., Heidenreich, T., Schmidt, A., 2017. Working with augmented reality? A long-term analysis of in-situ instructions at the assembly workplace. ACM Int. Conf. Proc. Ser. Part F1285, 222–229. https://doi.org/10.1145/3056540.3056548.

Gabbard, J.L., Mehra, D.G., Swan, J.E., 2019. Effects of ar display context switching and focal distance switching on human performance. IEEE Trans. Visual. Comput. Graph. 25, 2228–2241. https://doi.org/10.1109/TVCG.2018.2832633.

Gabbard, J.L., Swan, J.E., 2008. Usability engineering for augmented reality: employing user-based studies to inform design. IEEE Trans. Visual. Comput. Graph. 14, 513–525. https://doi.org/10.1109/TVCG.2008.24.

Garbarino, M., Lai, M., Tognetti, S., Picard, R., Bender, D., 2014. Empatica E3 - A Wearable Wireless Multi-Sensor Device for Real-Time Computerized Biofeedback and Data Acquisition, pp. 39–42. https://doi.org/10.4108/icst.mobihealth.2014.257418.

Gattullo, M., Uva, A.E., Fiorentino, M., Gabbard, J.L., 2015. Legibility in industrial AR: text style, color coding, and illuminance. IEEE Comput. Graph. Appl. 35, 52–61. https://doi.org/10.1109/MCG.2015.36.

Goldberg, J.H., Kotval, X.P., 1999. Computer interface evaluation using eye movements: methods and constructs. Int. J. Ind. Ergon. 24, 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. 52, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Höffler, T.N., Leutner, D., 2007. Instructional animation versus static pictures: a meta-analysis. Learn. InStruct. 17, 722–738. https://doi.org/10.1016/j.learninstruc.2007.09.013.

Hollnagel, E., 2005. Designing for joint cognitive systems. IET Semin. Dig. 47–51. https://doi.org/10.1049/ic:20050450, 2005.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Weijer, J. van de, 2013. Eye tracking A comprehensive guide to methods and measures. J. Chem. Inf. Model. 53, 1689–1699.
Hoover, M., Miller, J., Gilbert, S., Winer, E., 2020. Measuring the performance impact of using the microsoft HoloLens 1 to provide guided assembly work instructions. J. Comput. Inf. Sci. Eng. 20, 1–7. https://doi.org/10.1115/1.4046006.
IBM, 2019. SPSS [WWW Document].
Jacob, R.J.K., Karn, K.S., 2003. Eye tracking in human-computer interaction and usability research. Mind's Eye 573–605. https://doi.org/10.1016/b978-044451020-4/50031-1.
Jeffri, N.F.S., Awang Rambli, D.R., 2021. A review of augmented reality systems and their effects on mental workload and task performance. Heliyon 7, e06277. https://doi.org/10.1016/j.heliyon.2021.e06277.
Jones, A.T., Romero, D., Wuest, T., 2018. Modeling agents as joint cognitive systems in smart manufacturing systems. Manuf. Lett. 17, 6–8. https://doi.org/10.1016/j.mfglet.2018.06.002.
Kim, S., Nussbaum, M.A., Gabbard, J.L., 2019. Influences of augmented reality head-worn display type and user interface design on performance and usability in simulated warehouse order picking. Appl. Ergon. 74, 186–193. https://doi.org/10.1016/j.apergo.2018.08.026.
Lewis, J.R., 1992. Proceedings of the 36th annual meeting of the human factors society. Proc. Hum. Factors Soc. 1, 1259–1263.
Livingston, M.A., Rosenblum, L., Macedonia, M., 2005. Projects in VR: evaluating human factors in augmented reality systems. IEEE Comput. Graph. Appl. 25, 6–9. https://doi.org/10.1109/MCG.2005.130.
Loch, F., Quint, F., Brishtel, I., 2016. Comparing video and augmented reality assistance in manual assembly. Proc. - 12th Int. Conf. Intell. Environ. IE 147–150. https://doi.org/10.1109/IE.2016.31.
Lowe, R., 2004. Interrogation of a dynamic visualization during learning. Learn. InStruct. 14, 257–274. https://doi.org/10.1016/j.learninstruc.2004.06.003.
Lowe, R.K., 1999. Extracting information from an animation during complex visual learning. Eur. J. Psychol. Educ. 14, 225–244. https://doi.org/10.1007/BF03172967.
Markov-Vetter, D., Luboschik, M., Islam, A.T., Gauger, P., Staadt, O., 2020. The Effect of Spatial Reference on Visual Attention and Workload during Viewpoint Guidance in Augmented Reality. https://doi.org/10.1145/3385959.3418449. Proc. - SUI 2020 ACM Symp. Spat. User Interact.
Marques, B., Alves, J., Neves, M., Justo, I., Santos, A., Rainho, R., Maio, R., Costa, D., Ferreira, C., DIas, P., Santos, B.S., 2020. Interaction with virtual content using augmented reality: a user study in assembly procedures. Proc. ACM Human-Computer Interact. 4 https://doi.org/10.1145/3427324.
Microsoft, 2019a. Microsoft Hololens 2. WWW Document]. URL. https://www.microsoft.com/en-us/hololens/hardware.
Microsoft, 2019b. Eye tracking on Hololens 2 [WWW Document]. URL. https://docs.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking. accessed 2.9.21.
Moosburner, S., Remde, C., Tang, P., Queisner, M., Haep, N., Pratschke, J., Sauer, I.M., 2019. Real world usability analysis of two augmented reality headsets in visceral surgery. Artif. Organs 43, 694–698. https://doi.org/10.1111/aor.13396.
Nee, A.Y.C., Ong, S.K., Chryssolouris, G., Mourtzis, D., 2012. Augmented reality applications in design and manufacturing. CIRP Ann. - Manuf. Technol. 61, 657–679. https://doi.org/10.1016/j.cirp.2012.05.010.
Nguyen, N., Nelson, A.J., Wilson, T.D., 2012. Computer visualizations: factors that influence spatial anatomy comprehension. Anat. Sci. Educ. 5, 98–108. https://doi.org/10.1002/ase.1258.
Nichols, S., Patel, H., 2002. Health and safety implications of virtual reality: A review of empirical evidence. Appl. Ergon. 33, 251–271. https://doi.org/10.1016/S0003-6870(02)00020-0.
Niskanen, J.P., Tarvainen, M.P., Ranta-Aho, P.O., Karjalainen, P.A., 2004. Software for advanced HRV analysis. Comput. Methods Progr. Biomed. 76, 73–81. https://doi.org/10.1016/j.cmpb.2004.03.004.
Norros, L., 2014. Developing human factors/ergonomics as a design discipline. Appl. Ergon. 45, 61–71. https://doi.org/10.1016/j.apergo.2013.04.024.
Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H., 2018. A systematic review of augmented reality applications in maintenance. Robot. Comput. Integrated Manuf. 49, 215–228. https://doi.org/10.1016/j.rcim.2017.06.002.
Poole, A., Ball, L.J., Phillips, P., 2007. In search of salience: a response-time and eye-movement analysis of bookmark recognition. People Comput. XVIII — Des. Life 363–378. https://doi.org/10.1007/1-84628-062-1_23.
Poushneh, A., 2018. Augmented reality in retail: a trade-off between user's control of access to personal information and augmentation quality. J. Retailing Consum. Serv. 41, 169–176. https://doi.org/10.1016/j.jretconser.2017.12.010.
PTC, 2019. Vuforia [WWW Document]. URL. https://library.vuforia.com/features/objects/object-reco.html.
Rehman, U., Cao, S., 2020. Comparative evaluation of augmented reality-based assistance for procedural tasks: a simulated control room study. Behav. Inf. Technol. 39, 1225–1245. https://doi.org/10.1080/0144929X.2019.1660805.
Rieber, L.P., Tzeng, S.C., Tribble, K., 2004. Discovery learning, representation, and explanation within a computer-based simulation: finding the right mix. Learn. InStruct. 14, 307–323. https://doi.org/10.1016/j.learninstruc.2004.06.008.
Rizzo, a. a., Kim, G.J., Yeh, S.C., Thiebaux, M., Hwang, J., Buckwalter, J.G., 2005. Development of a benchmarking scenario for testing 3D user interface devices and interaction methods. Proc. 11th Int. Conf. Hum. Comput. Interact.
Schwerdtfeger, B., Reif, R., Günthner, W.A., Klinker, G., Hamacher, D., Schega, L., Böckelmann, I., Doil, F., Tümler, J., 2009. Pick-by-Vision: a first stress test. Sci. Technol. Proc. - IEEE 2009 Int. Symp. Mix. Augment. Reality, ISMAR 2009 115–124. https://doi.org/10.1109/ISMAR.2009.5336484.
Shah, M.M., Arshad, H., Sulaiman, R., 2012. Occlusion in augmented reality. Proc. - ICIDT 2012, 8th Int. Conf. Inf. Sci. Digit. Content Technol. 2, 372–378.
Sharafi, Z., Soh, Z., Guéhéneuc, Y.G., 2015. A systematic literature review on the usage of eye-tracking in software engineering. Inf. Software Technol. 67, 79–107. https://doi.org/10.1016/j.infsof.2015.06.008.
Smaldino, 2012. Instructional media and technology for learning. Int. J. Distrib. Parallel Syst. 3, 8.
Stork, S., Schubö, A., 2010. Human cognition in manual assembly: theories and applications. Adv. Eng. Inf. 24, 320–328. https://doi.org/10.1016/j.aei.2010.05.010.
Tullis, T., Albert, B., 2008. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann. https://doi.org/10.1088/1751-8113/44/8/085201.
Unity, 2019. Unity [WWW Document].
Wang, C.H., Tsai, N.H., Lu, J.M., Wang, M.J.J., 2019. Usability evaluation of an instructional application based on Google Glass for mobile phone disassembly tasks. Appl. Ergon. 77, 58–69. https://doi.org/10.1016/j.apergo.2019.01.007.
Westerfield, G., Mitrovic, A., Billinghurst, M., 2015. Intelligent augmented reality training for motherboard assembly. Int. J. Artif. Intell. Educ. 25, 157–172. https://doi.org/10.1007/s40593-014-0032-x.
Wickens, C.D., 2008. Multiple resources and mental workload. Hum. Factors 50, 449–455. https://doi.org/10.1518/001872008X288394.
Wickens, C.D., 2002. Multiple resources and performance prediction. Theor. Issues Ergon. Sci. 3, 159–177. https://doi.org/10.1080/14639220210123806.
Wickens, C.D., Sandry, D.L., Vidulich, M., 1983. Compatibility and resource competition between modalities of input, central processing, and output. Hum. Factors 25, 227–248. https://doi.org/10.1177/001872088302500209.
Wintersberger, P., Riener, A., Frison, A.K., 2016. Automated driving system, male, or female driver: who'd you prefer? Comparative analysis of passengers' mental conditions, emotional states & qualitative feedback. AutomotiveUI 2016 - 8th Int. Conf. Automot. User Interfaces Interact. Veh. Appl. Proc. 51–58. https://doi.org/10.1145/3003715.3005410.
Wu, H.K., Lin, Y.F., Hsu, Y.S., 2013. Effects of representation sequences and spatial ability on students' scientific understandings about the mechanism of breathing. Instr. Sci. 41, 555–573. https://doi.org/10.1007/s11251-012-9244-3.
Yamaguchi, M., Mori, S., Mohr, P., Tatzgern, M., Stanescu, A., Saito, H., Kalkofen, D., 2020. Video-annotated augmented reality assembly tutorials. UIST 2020 - Proc. 33rd Annu. ACM Symp. User Interface Softw. Technol. 1010–1022. https://doi.org/10.1145/3379337.3415819.
Yusoff, R.C.M., Ibrahim, R., Zaman, H.B., Ahmad, A., 2011. Evaluation of user acceptance of mixed reality technology. Australas. J. Educ. Technol. 27, 1369–1387. https://doi.org/10.14742/ajet.899.
Zheng, X.S., Foucault, C., Da Silva, P.M., Dasari, S., Yang, T., Goose, S., 2015. Eye-wearable technology for machine maintenance: effects of display position and hands-free operation, 2015-April Conf. Hum. Factors Comput. Syst. - Proc.. https://doi.org/10.1145/2702123.2702305, 2125–2134.
Zhu, J., Ong, S.K., Nee, A.Y.C., 2013. An authorable context-aware augmented reality system to assist the maintenance technicians. Int. J. Adv. Manuf. Technol. 66, 1699–1714. https://doi.org/10.1007/s00170-012-4451-2.