



HAL
open science

Augmented reality on industrial assembly line: Impact on effectiveness and mental workload

Mathilde Drouot, Nathalie Le Bigot, Emmanuel Bricard, Jean-Louis De Bougrenet, Vincent Nourrit

► To cite this version:

Mathilde Drouot, Nathalie Le Bigot, Emmanuel Bricard, Jean-Louis De Bougrenet, Vincent Nourrit. Augmented reality on industrial assembly line: Impact on effectiveness and mental workload. Applied Ergonomics, 2022, 103, pp.103793. 10.1016/j.apergo.2022.103793 . hal-03667191

HAL Id: hal-03667191

<https://hal.science/hal-03667191>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Augmented reality on industrial assembly line: impact on effectiveness and mental workload

Mathilde Drouot^{1,2,3*}, Nathalie Le Bigot⁴, Emmanuel Bricard³, Jean-Louis de Bougrenet^{1,2},
Vincent Nourrit^{1,2}.

¹*Optics Department, IMT Atlantique, 655 avenue du Technopôle, Brest, France*

²*LaTIM, INSERM UMR1101, 22 avenue Camille Desmoulins, 29200, Brest, France*

³*elm.leblanc (Bosch Group), Drancy, France*

⁴*Univ Brest, Lab-STICC, CNRS, UMR 6285, Brest, France*

*Corresponding author: Mathilde Drouot. Email-address: mathilde.drouot@imt-atlantique.org

Abstract

Studies examining the potential of augmented reality (AR) to improve assembly tasks are often unrepresentative of real assembly line conditions and assess mental workload only through subjective measurements and leads to conflicting results. We proposed a study directly carried out in industrial settings, to compare the impact of AR-based instructions to computerized instructions, on assembly effectiveness (completion time and errors) and mental workload using objective (eye tracking), subjective (NASA-TLX) and behavioral measurements (dual task paradigm). According to our results, AR did not improve effectiveness (increased assembly times and no decrease in assembly errors). Two out of three measurements indicated that AR led to more mental workload for simple assembly workstation, but equated computer instructions for complex workstation. Our data also suggest that, AR users were less able to detect external events (danger, alert), which may play an important role in the occurrence of work accidents.

Keywords: augmented reality, industrial assembly, mental workload.

1. Introduction

Augmented reality (AR) displays enable users to see additional digital information (usually a virtual image) over the real environment. Unlike virtual reality, the user still experiences the real world, but enhanced with digital content. As a result, AR is increasingly used for industrial processes, and companies are looking to embrace AR in production lines, for operator training of industrial assembly tasks. The integration and use of this type of device in professional environment, however, raises many questions from a technological, industrial and user's point of view. The use of AR to display work instructions was first proposed by Caudell et Mizell (1992), who suggested that overlaying information on real-world objects using a Head-Up-Display device would reduce costs and improve the effectiveness of human operations.

In their review, Anastassova et al., (2007) listed several hypotheses that have been proposed by different authors to explain the benefits of AR for training and procedural learning. AR allows to associate an abstract concept with a physical object, improving the understanding of some complex concepts (Stedmon and Stone, 2001). AR also provides contextualized information in real time, which would reduce the risk of errors during training sessions (Neumann and Majoros, 1998). Furthermore, AR could increase the user's motivation due to the novelty of its interaction mode (Zhong et al., 2003). Finally, AR leaves the user's hands free and allows both autonomous learning and communication with an expert on the task if the user needs help. Based on the potential of AR to improve operator's training of assembly tasks, the use of AR instructions for assembly tasks has been the subject of numerous studies, particularly focused on its effectiveness on user performance (see Drouot et al, (2021c) unpublished results, and Werrlich et al., (2017) for a recent review). The evaluation of its effectiveness is based on three indicators: assembly time, assembly errors and mental workload.

Assembly time is an indicator frequently used in the literature, for which there are contradictory observations. Indeed, some studies observe a decrease in time needed to perform assembly tasks using instructions displayed with projective AR, compared to classical paper instructions (Hou et al., 2015, 2013; Hou & Wang, 2013; Yang et al., 2019). However, many studies also showed that the use of AR does not result in time savings, and might even result in longer completion times when compared to paper instructions (Haniff & Baber, 2003; Min et al., 2020), but also to other types of instructions, such as videos (Loch et al., 2016; Min et al., 2020a), virtual reality (Gavish et al., 2015), and tablets (Funk et

al., 2016). One explanation for these divergent results is that the benefit of AR instructions depends on the complexity of the task, with a benefit only when the task is complex (Wiedenmaier et al., 2003).

Reduction of assembly errors when using AR is the most common benefit observed in studies. Indeed, users performed assembly tasks with fewer errors using instructions displayed in projective AR compared to conventional paper instructions (Bode, 2019; Hou et al., 2015, 2013; Hou and Wang, 2013; Yang et al., 2019). According to Tatić et Tešić (2017), one explanation would be that using AR prevents users from skipping steps or completing them in the wrong order.

The third indicator used in the literature is mental workload. Fewer studies have evaluated this parameter and, similarly to what is observed for assembly time, data are not congruent. Indeed, studies showed that the use of AR causes, either a decrease in mental workload (Hou & Wang, 2013; Yang et al., 2019), an increase (Blattgerste et al., 2017; Funk et al., 2016) or no change (Funk et al., 2017). AR allows to spatialize information directly in the assembly area, Tang et al. (2003) suggested that this reduces head movements, thus decreasing the time spent searching for information, which could decrease the user's mental load. However, the low level of user experience and familiarity with AR devices, could also generate additional effort to learn how to use them and thus could explain the increased mental workload compared to other instructional media.

In previous studies, mental workload has been evaluated using a single questionnaire, i.e., the NASA TLX (Hart and Staveland, 1988), that includes different workload dimensions (physical, mental, temporal, performance, effort, frustration). The different dimension of mental workload could provide various information on the source of workload (e.g., distinguish the workload due to the task from the one due to wearing the HMD). However, in these studies, only the overall score was considered, as mental workload was only investigated as a quantitative indicator, such as assembly time and errors. However, mental workload is a complex process representing the physical, cognitive and motivational resources of a user, which can be further evaluated using different measurements (subjective, objective, and behavioral).

Thus, in addition to the subjective measurements, in which participants are asked to provide a judgement of mental workload required by the task after they performed it (e.g., NASA TLX), mental workload can be evaluated by objective and behavioral measurements, see Boyer (2017), Miller (2001) and Cain (2007) for reviews. Physiological reaction (e.g., heart rate, blink rate, pupil size) can be recorded to obtain objective measurement of mental workload directly during the task (Charles & Nixon, 2019; Tao et al., 2019), but they require the use of specific equipment such as sensors or cameras. Behavioral measurements are a direct assessment of mental workload, as they are based on the effects of mental workload during the task, i.e. the level of performance achieved by the individual. They can be carried out by assessing performance on the main task (de Waard, 1996; Hicks and Wierwille, 1979) or on an additional task performed simultaneously (de Waard, 1996; Kellogg, 1990; Piolat et al., 1999).

Those different measurements of mental workload (i.e. subjective, physiological and behavioral) can be distinguished from each other by their technical requirements and by the quantity and the quality of data obtained (see Cegarra & Chevalier (2008) for a review). In this review, Cegarra and Chevalier (2008) show the advantages of combining physiological, behavioral and subjective measures of mental workload to compensate the different limitations of each measurement. Due to the multi-dimensionality of mental workload, many authors have chosen to measure it with a set of variables rather than with a single variable (Bednarik et al., 2018; Jafari et al., 2020; Luque-Casado et al., 2016; Wanyan et al., 2014). While having important advantages for the study of mental workload, using the NASA TLX in combination with other measures would provide more evidence on the impact of AR. To the best of our knowledge, no study on mental workload generated by AR, combining several types of measurement (subjective, objective and behavioral) has been carried out, either in a laboratory environment or in an industrial setting.

Considering advantages and limitations of each of the measurement, as well as their feasibility in an industrial setting, we selected one measure from each category: the ocular measures, the dual task paradigm and the NASA TLX. Although requiring a controlled environment, the ocular measures (eye blinks and pupil size) allow a dynamic measurement of mental workload. The eye trackers used to make these measurements are becoming less expensive and intrusive, and are beginning to be integrated directly into AR or VR devices, making them, a relevant tool for physiological measurement of mental workload in AR. The dual-task paradigm also allows for dynamic measurement, and requires very few resources, making it easy to implement in industrial settings. Finally, the NASA TLX has the fewest technical requirements is non-intrusive, and easily administered, and measures several dimensions of mental workload.

The present study aimed at evaluating whether the use of AR for assembly instructions presentation presents a benefit compared to existing solutions. Most of these studies on assembly were conducted in controlled laboratory environments, and they often used an artificial assembly task, e.g., with LEGO blocks (Drouot et al., 2021a; Funk et al., 2016; Rodriguez et al., 2021). Although laboratory studies are necessary, more representative studies that consider real assembly line conditions are also needed. In order to address the gaps identified in the literature, we chose to study the impact of AR directly in an industrial environment with collaborators from the factory, i.e., on assembly workstations in a factory. To this end, we have compared the effects of AR-based instructions and computerized instructions, on assembly performances and mental workload. We also compared a simple and a complex assembly workstation to determine whether the advantage of AR is only observed for complex tasks. If contextualization and spatialization of AR information reduces head movement and eliminates ambiguity about the location of the component to use, it could reduce overall assembly time and errors. Finally, if AR allows for optimal integration of information in the area of interest and therefore avoids the need for the operator to hold this information in memory, it could reduce mental workload.

2. Methodology

2.1 Participants

Twenty-seven employees (5 females, 22 males) from elm.leblanc factory (company manufacturing gas heating and hot water production devices) from a wide range of education levels (from BTEC First Diploma to PhD) and positions (e.g., apprentices, managers, directors) were recruited for this study. Participants included were aged between 24 and 63 ($M=35.88$; $SD=12.67$) and used their usual optical correction (glasses, contacts). Among participants, three employees reported having a manual assembly experience, but none of them had any experience or knowledge of the assembly workstations used in this study. Seventeen participants reported having used virtual reality at least once and 15 for AR. Prior to participation, all volunteers were fully informed about the procedure and signed an informed consent form. The study was conducted in accordance with the principles of the Declaration of Helsinki.

2.2 Experimental design

In this study, two factors were manipulated: the instruction medium (AR, Computer) and the complexity of the assembly workstation (Complex, Simple). In the *AR* condition, the instructions were presented using an AR Head Mounted display (HMD). In the *Computer* condition, the same instructions were presented using PowerPoint slides on a computer. To ensure that the weight of the HMD did not affect the data, the HMD was worn by participants in both conditions.

In the *Complex* condition, the assembly workstation (workstation 101.1) required 33 instructions of moderate complexity, such as screwing and rotations. In the *Simple* condition, the assembly workstation (workstation 101.2) required 13 instructions of low complexity such as positioning.

Due to industrial constraints, only one workstation of each complexity was available, thus, each participant performed the tasks under only two of the four experimental conditions (i.e., Complex AR and Simple Computer or Simple AR and Complex Computer), avoiding working twice on the same workstation and thus avoiding training effects. In addition, to avoid order effects, the order of presentation was counterbalanced, i.e. half of the participants worked first on the simple workstation and then the complex workstation while the other half worked on the complex workstation first.

2.3 Materials

2.3.1 Software and devices

A Microsoft HoloLens 2 was used to display instruction in AR condition, and Unity 3D and MRTK were chosen to implement the instruction creation and visualization tool (see Lavric et al., (2021) for a full description of the development methodology). A Fujitsu 15.6" computer was used to display the instructions in the computer condition. PsychoPy software was used to generate sound stimuli randomly every 15-25 seconds. WavePad software was used to record and analyze participants' oral responses. The Pupil Labs eye tracker (integrated directly onto the HoloLens 2 using custom 3D printed parts) and the PupilCapture software were used to record the participants' eye data, and the PupilPlayer software was used to analyze and extract the information (Kassner et al., 2014).

2.3.2 Assembly workstations

The chosen assembly workstations are the first two of Line 1 in elm.leblanc, which form the basic structure of a boiler. In these two workstations, the assembly involves installing components with or

without help of tools, on a defined assembly structure (see Figure 1). The components to use are very diverse in size, shape and weight: very small (e.g., screws, nuts), medium (e.g., brackets, plates) and large (e.g., uprights, boiler shells).

The assembly tasks to be performed can be divided into four categories:

- Picking: e.g. picking up the component (screw, nut) at the specified location;
- Positioning: e.g., positioning the component at the specified location;
- Tool uses: e.g., using the screwdriver, or locking the clamps;
- Handling: e.g., turning, moving, joining two components together.

Assembly workstation 101.1 consists of 33 steps of medium complexity: 12 x Picking, 8 x Positioning, 6 x Tool uses, 7 x Handling. Assembly workstation 101.2 consists of 13 steps of low complexity: 4 x Picking, 5 x Positioning, 3 x Tool uses, 1 x Handling.

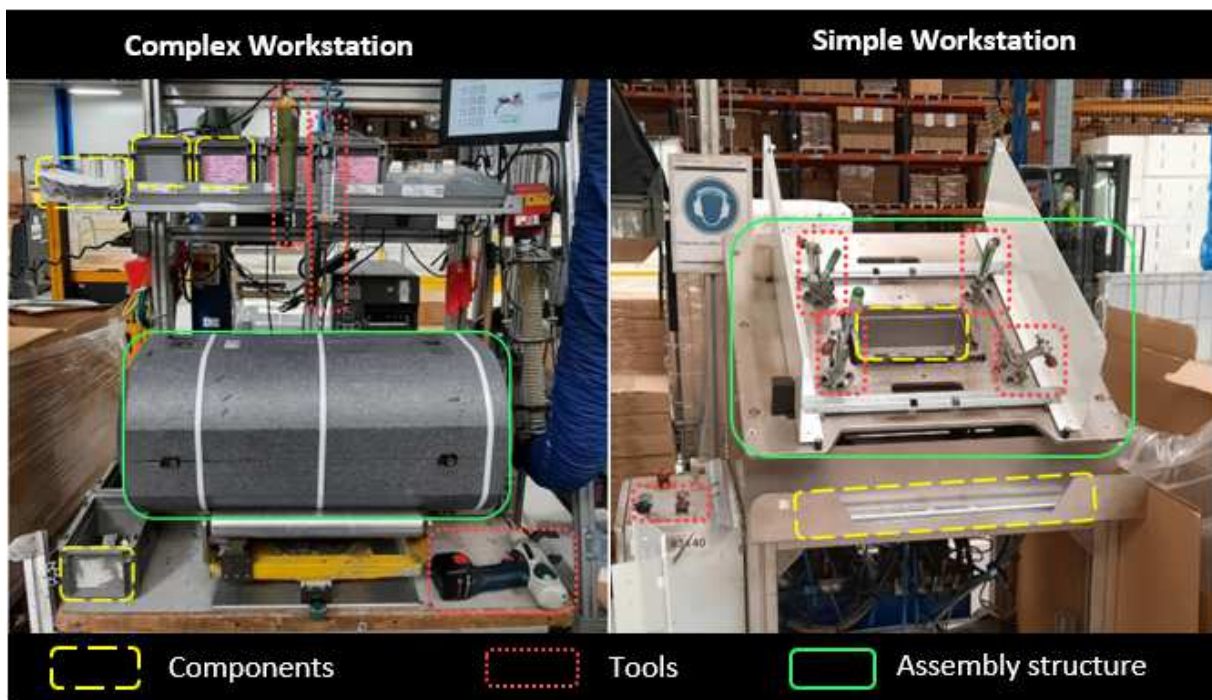


Figure 1 - Pictures of elm.leblanc workstations 101.1 and 101.2

2.3.3 Assembly instructions

The existing paper workstation instructions proposed by elm.leblanc were not suitable for novices, as they do not allow the assembly steps to be carried out without help of a team leader. We therefore developed our own assembly instructions, and designed it to ensure that instructions in AR are as equivalent as possible to those presented in the computer. Given the complexity of some of the operations to be performed, for the instructions that cannot be easily explained with pictures, we have included videos in addition to the textual information. Instructions (e.g., text, symbols and videos) were presented both in AR and in the computer, allowing hands-free operation, which is essential for manual assembly.

Resources used to present assembly information, either in the AR HMD or in the computer, include: text, signs and symbols, photos and videos. Each assembly instruction is visually composed of three elements providing three different pieces of information:

- *What*: A textual information describing the assembly task to perform;
- *Where*: An arrow pointing the physical location of the instruction to execute;
- *How*: Depending on the complexity of the task, an image of the component to be used or a video indicating the sequence of movements to perform.

The instructions presented in the AR HMD or in the computer were identical, with the same number of steps and the same information (i.e. where, what, how), see Figure 2. The difference between the two media is that AR instructions are spatialized, whereas those in the computer are fixed to a specific

location. The other difference relates to the possibilities of interaction between the user and the instruction medium. For instructions on the computer, the user interacts with the two keyboard arrows, the top one to return to the previous step and the bottom one to move on to the next. For the instructions in the AR HMD, participants can perform different actions by *touching* the "next", "previous" and "hide" buttons with their index fingertip (of their preferred hand). Participants can also perform the same actions using voice commands, by saying the words "next", "previous" and "hide" respectively. After hiding the instructions (e.g., if they obstruct the field of view), participants can make them reappear using the voice command, "help me". To use the AR instructions, a short calibration phase of the HoloLens 2 is automatically started if a new user is detected, in order to allow a personalized display (using the interpupillary distance). Then, to be able to interact with the developed interface, users must scan with the HoloLens 2 a QR code which allows workstation identification. After selecting the designed application in the interface, users can access the instruction steps.

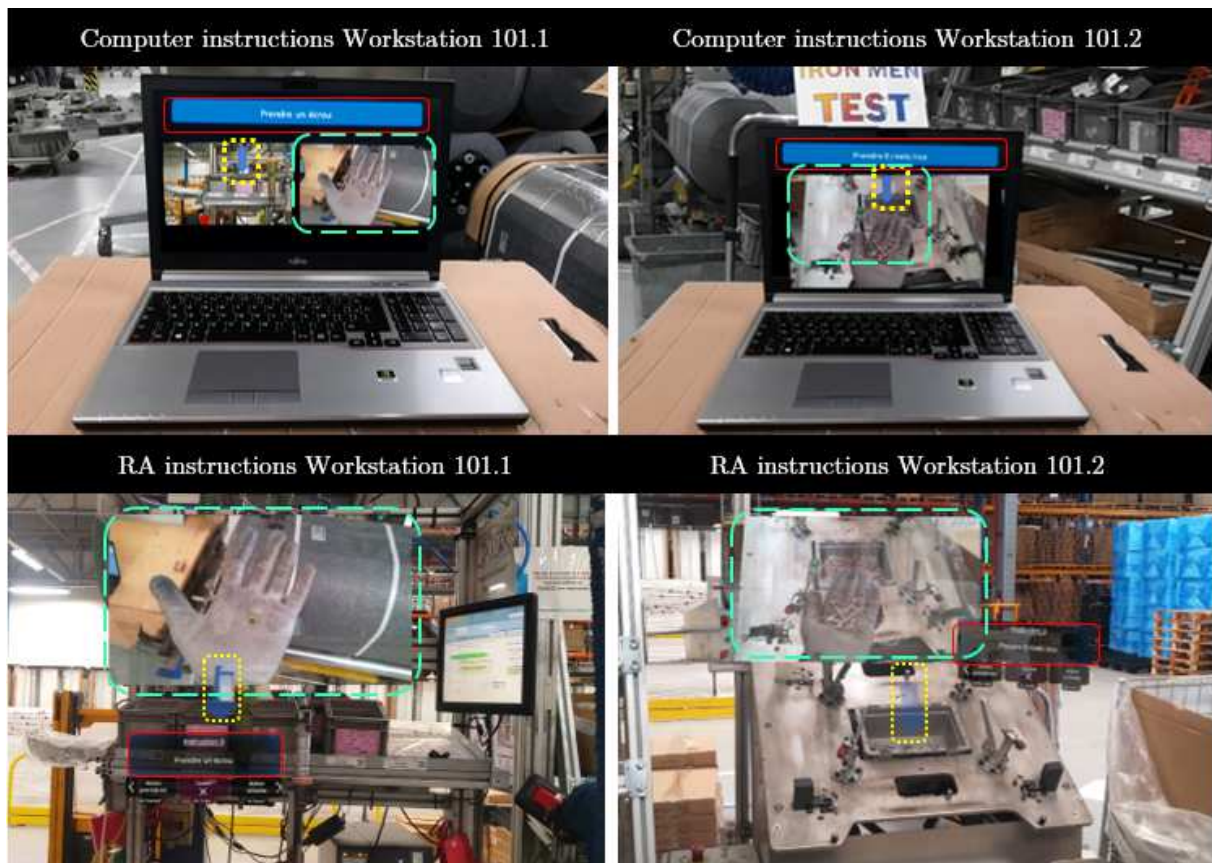


Figure 2 -Pictures of assembly for workstation 101.1 (top) and 101.2 (bottom), for computer condition (left) and AR condition (right).

2.4 Procedure

The two tasks to be performed were explained to participants, and were first asked to perform a main task corresponding to the assembly of all components for each workstation using the corresponding instructions (i.e. in the AR HDM or in the computer). Participants were asked to follow the instructions carefully, trying to make as few mistakes as possible.

It was then explained to them that, in addition to the main assembly task, a second task had to be performed at the same time. This second task was an auditory stimulus detection task. Participants had to respond orally as quickly as possible by pronouncing the syllable "ta" following the auditory stimulus. These stimuli were "beeps" (100ms) produced randomly every 15 to 25s, and were emitted from a mobile speaker positioned near participants. Participants were told that the assembly task was the primary task and took priority over the beep detection task.

Depending on the instruction medium condition, participants were told how to interact with the corresponding interface. When using the AR instructions, the HoloLens 2 was placed on the participant's head, ensuring that it was correctly positioned. After the HoloLens 2 calibration phase, the eye tracker

cameras were repositioned so that they were in front of participants' pupils, ensuring that the software detected them.

After participants had confirmed that they had understood all the instructions, the recordings were started (audio recording of the oral responses, recording of the eye tracker cameras, recording of the assembly times for each step). Once participants were ready, PsychoPy program started to generate the beeps, and the participants could begin the first assembly step on the corresponding instructional support.

While participants were performing the assembly, the experimenter checked the assembly steps and if an error was made by participants and not resolved, they were notified. Once the assembly workstation was completed, participants removed the HMD and had to fill in the NASA TLX. After a 10-minute break, they assembled the second workstation in the other complexity and instruction medium condition. At the end of the second assembly, participants also had to fill in the NASA TLX and a form about their visual health, their previous experience and habit with technologies (i.e., video games, virtual reality and AR), their manual assembly experience, and also about their preference for the instructional medium.

2.5 Data collection and analyses

The fact that our participants performed tasks under only two of the four experimental conditions (e.g., complex AR and simple computer), which were counterbalanced using a balanced Latin square, only allows us to analyse the main effects. Thus, a 2x2 ANOVA, with instruction medium (AR, Computer) and complexity (complex, simple) was calculated to assess the main effect on each of the measured parameters.

Our objective is to determine if the impact of AR is different between a simple assembly workstation and a complex one. For this reason, participants' data was gathered in two groups: a first group "G1" for those who used augmented reality in the "simple" condition and a "G2" group for the others who used the AR medium in the "complex" condition. Thus, a 2x2 mixed ANOVA, with complexity (complex, simple) as a within factor and group as a between factor (G1, G2) was calculated for each measured parameter.

Statistical analyses were performed with JASP (JASP Team, 2019). The analyses were conducted after carefully verifying that the underlying assumptions of the parametric test were satisfied. Shapiro-Wilk test was performed to ensure the samples were normally distributed. When normality assumption was violated, a logarithmic transformation of the data was used. Levene's test of sphericity was conducted to ensure that error variance in groups of samples are equivalent. Due to missing data, two participants were removed from the analyses.

Blink rate, blink duration and pupil size were measured using the PupilLab eye tracker for all participants, however, due to problems with data acquisition and recording (e.g., camera failure), the analysis of eye data was only performed for the 17 participants (with the full set of eye data recorded). Only data with an associated pupil detection confidence of 80% were analyzed. Data associated blinks with a duration of less than 50ms and more than 500ms were removed from the analysis (Caffier et al., 2003), representing 18% of data.

An audio recording of the entire experimental session was made using a loudspeaker microphone placed near the subject and Wavepad software. Reaction times associated with the participant's oral response were extracted in the Wavepad sound editor from this recording including auditory stimuli and the participant's oral responses. No specific processing was performed on the audio signal.

For behavioral measurement, a lower limit at 100ms is usually considered as a valid human reaction time (Whelan, 2008), and a cutoff was selected as a function of the proportion of responses eliminated (between 5% and 10%, Ratcliff, 1993). Thus, reaction times to beep detection task below 100ms and above 1500ms, representing 6% of the data, were removed from the analysis.

3. Results

Main effect analyses

3.1 Effectiveness

3.1.1 Assembly time

A main effect of instruction medium was observed on assembly time (see Figure 3), $F(1, 46)=21.67$, $p<0.001$, $\eta_p^2=0.161$. Participants performed assembly (in seconds) faster in the computer condition ($M=540.50$, $SD=165.35$) than in the AR condition ($M=851.58$, $SD=267.76$). A main effect of complexity was also found on assembly time $F(1, 46)=66.42$, $p<0.001$, $\eta_p^2=0.494$. Participants performed the assembly faster in simple condition ($M=423.71$, $SD=135.28$) than in complex condition ($M=968.36$, $SD=297.83$).

3.1.2 Assembly errors

We found no main effect of instruction medium and complexity on unsolved assembly errors ($ps>0.450$).

3.2 Mental workload

3.2.1 Physiological measurement: eye tracking

A main effect of instruction medium was observed on blink rate, $F(1, 30)=3.30$, $p=0.008$, $\eta_p^2=0.099$. Blink rate is lower in AR condition ($M=12.40$, $SD=5.77$) than in computer condition ($M=16.40$, $SD=6.80$). No main effect of complexity was found, $F<1$.

No main effect of complexity and instruction medium was found on blink duration, ($ps>0.394$).

A main effect of instruction medium was observed on pupil size, $F(1, 46)=4.84$, $p=0.033$, $\eta_p^2=0.097$. Pupil size was smaller in AR condition ($M=24.84$, $SD=4.29$) than in computer condition ($M=28.08$, $SD=5.75$). No main effect of complexity was observed ($p>0.502$).

3.2.2 Behavioral measurement: secondary task

A main effect of instruction medium was found on detection task reaction time, $F(1, 46)=9.46$, $p=0.004$, $\eta_p^2=0.171$. Reaction time were faster in computer condition ($M=692.31$, $SD=128.66$) than in AR condition ($M=811.19$, $SD=136.78$). No main effect of complexity was found, $F<1$.

We observed a main effect of instruction medium on omission numbers for the detection task $F(1, 46)=5.42$, $p=0.024$, $\eta_p^2=0.105$. The number of missed beeps is greater in AR condition ($M=2.06$, $SD=3.32$) than in computer condition ($M=0.18$, $SD=1.86$). No main effect of complexity was observed, $F(1, 46)=2.62$, $p=0.112$, $\eta_p^2=0.05$.

3.3.3 Subjective measurement: NASA-TLX

We found a main effect of complexity only on the physical dimension of the mental workload, $F(1, 46)=4.43$, $p=0.041$, $\eta_p^2=0.088$. Participants reported more physical workload (here due to the weight of some components and the force required for some actions), for complex workstation ($M=31.20$, $SD=24.10$) than for simple workstation ($M=18.44$, $SD=16.81$). No significant difference was observed for the other workload dimension ($ps>0.125$).

3.3.4 Post-experimental questionnaire

At the post-experimental questionnaire, significantly more participants reported AR as their favorite instruction medium (20 participants), compared to computer (6 participants), $X^2(1,26)=7.54$, $p<0.001$.

Interaction effect analyses

We observed an interaction effect between complexity and group on assembly time, $F(1, 23)=38.36$, $p<0.001$, $\eta_p^2=0.625$. Simple effect analysis (see Figure 3), showed a significant difference between the simple ($F(1, 23)=19.71$, $p<0.001$) and complex ($F(1, 23)=8.77$, $p=0.007$) condition groups. Thus, the increase in assembly time with AR is greater in the complex condition (AR ($M=558.43$, $SD=186.17$); computer ($M=289.00$, $SD=6.46$)), than in the simple condition (AR ($M=1144.73$, $SD=349.36$); computer ($M=792.00$, $SD=246.30$)).

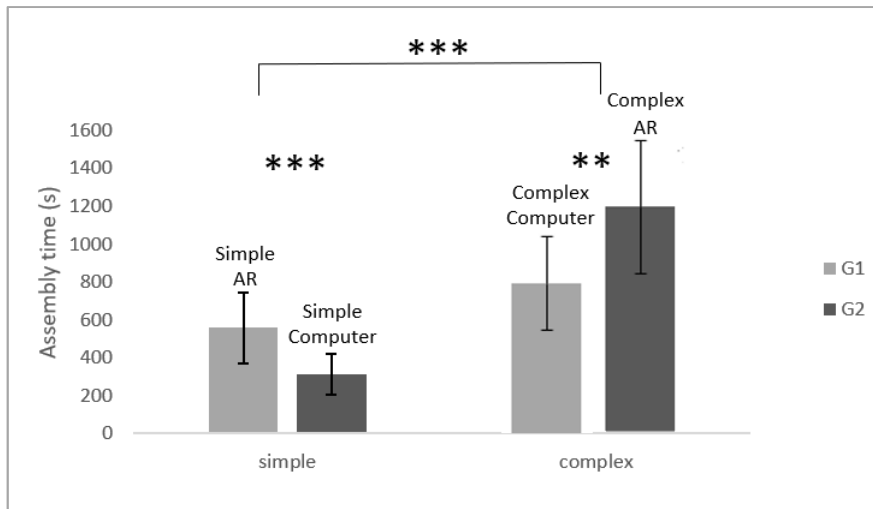


Figure 1 - Mean assembly time depending on complexity and groups. Error bars represent standard deviation. ** Significant at 0.01, *** Significant at 0.001.

An interaction effect between complexity and group was found on blink rate, $F(1, 11)=6.340$, $p=0.029$, $\eta_p^2=0.366$. Simple effect analysis (see Figure 4) showed a significant difference between the groups in the simple condition $F(1, 11)=7.48$, $p=0.019$, but not for the complex condition $F<1$. Thus, in the simple condition, blink rate is lower with AR ($M=9.90$, $SD=3.88$) than with computer ($M=18.42$, $SD=7.49$), whereas in the complex condition, there is no difference between AR ($M=13.95$, $SD=7.35$) and computer ($M=14.78$, $SD=7.38$).

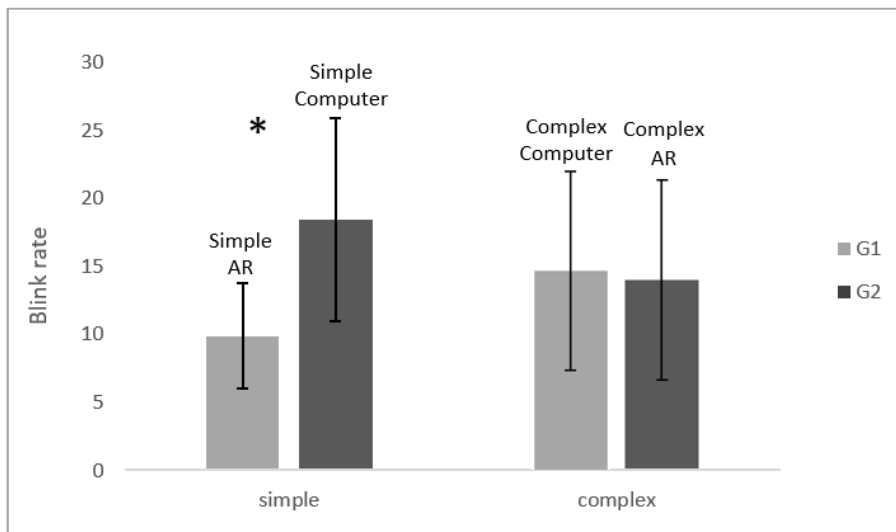


Figure 4 - Mean blink rate, according to complexity and groups. Error bars represent standard deviation. * Significant at 0.05.

We found an interaction effect between complexity and group on pupil size, $F(1, 22)=13.96$, $p=0.001$, $\eta_p^2=0.388$. Simple effect analysis showed a significant difference between groups in the simple condition $F(1, 22)=5.91$, $p=0.024$, but not for the complex condition $F(1, 22)=1.21$, $p=0.301$. Thus, in the simple condition, pupil size is smaller with AR ($M=23.84$, $SD=4.45$) than with computer ($M=28.79$, $SD=5.49$), whereas in the complex condition, there is no difference between AR ($M=25.82$, $SD=4.13$) and computer ($M=28.07$, $SD=5.77$).

We found an interaction effect between complexity and group on detection task reaction time, $F(1, 23)=59.39$, $p<0.001$, $\eta_p^2=0.721$. Simple effect analysis (see Figure 5) showed a significant difference between groups in the simple condition $F(1, 23)=14.29$, $p<0.001$, but not for the complex condition $F<1$. Thus, in the simple condition, reaction time were longer with AR ($M=862.26$, $SD=158.78$) than with

computer ($M=642.83$, $SD=122.32$), whereas in the complex condition, there is no difference between AR ($M=760.11$, $SD=114.89$) and computer ($M=741.80$, $SD=135.01$).

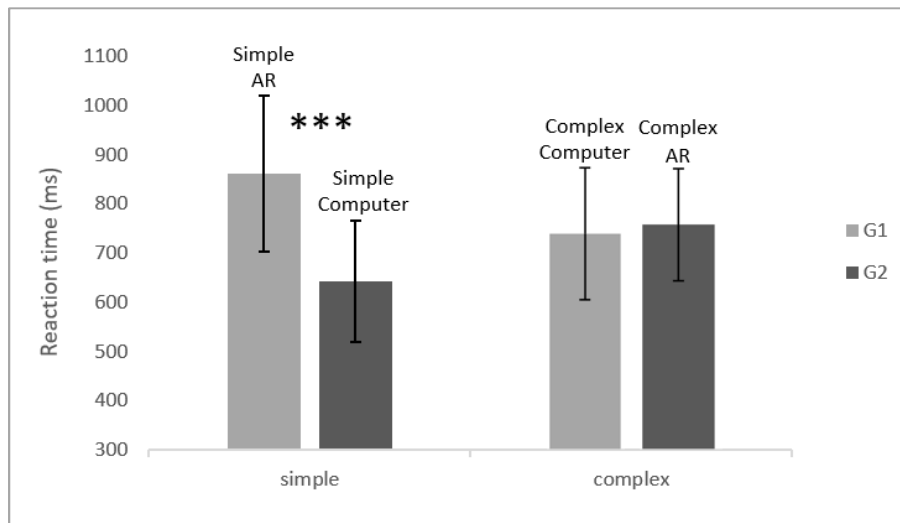


Figure 5 - Mean reaction time to beep detection task, according to instruction medium and complexity. Error bars represent standard deviation. *** Significant at 0.001.

An interaction effect was also observed between complexity and group on the number omissions for the detection task, $F(1, 23)=11.53$, $p=0.003$, $\eta_p^2=0.334$. Simple effect analysis showed a significant difference between groups in the simple condition $F(1, 23)=4.50$, $p=0.045$, but not for the complex condition $F(1, 23)=1.45$, $p=0.24$. Thus, in the simple condition, the number of omissions is higher with AR ($M=2.78$, $SD=3.53$) than with computer ($M=0.45$, $SD=0.94$), whereas in the complex condition, there is no difference between AR ($M=3.64$, $SD=3.11$) and computer ($M=2.21$, $SD=2.78$).

No interaction effect between complexity and group was found on assembly errors and blink duration, nor on the different workload dimension ($ps > 0.30$)

4. Discussion

This study aims at evaluating the effects of AR, on assembly performance and mental workload of users, directly in industrial assembly lines. Our study shows that the use of AR instructions leads to longer assembly times than the same instructions displayed on a computer. This result is coherent with numerous studies that have also found longer assembly times with AR compared to paper (Haniff & Baber, 2003; Min et al., 2020), video (Loch et al., 2016) or tablet (Funk et al., 2016) instructions. These longer assembly times suggest that the use of AR is associated with a cost for the user. A possible explanation for this cost could be related to spatialization of information in AR (not present in the computer). The fact that the information is spatialized (i.e. directly where the action must be performed) means that the user does not know where the information is, and additional time is needed to locate it. On the other hand, with the computer, the information is fixed, so the participant always knows where to look. Thus, the increase in AR assembly time, which is greater for a complex workstation than for a simple one, could be explained by the fact that there are more assembly steps to be carried out and by the incompressible cost associated with the search for information at each step.

Nevertheless, in contrast to several studies showing a benefit of AR in reducing the number of assembly errors (Bode, 2019; Hou et al., 2015, 2013; Hou and Wang, 2013; Yang et al., 2019), we did not find any decrease in errors with AR in our study. This could be explained by the interdependence of the assembly steps. Indeed, in our study, contrary to some studies (e.g. with Lego blocks), if a mistake was made, participants often noticed it themselves, as they could not perform the next step directly. Thus, as only errors unsolved by participants themselves were reported, there were few mistakes overall, regardless of the type of instructional medium used.

Although AR does not allow faster and accurate assembly performances, our post-experimental questionnaire shows that most participants reported AR as their favorite instruction medium, compared to computer. Based on participants' comments, this better acceptability could be explained by the

benefits of voice commands (leaving the user's hands free), the spatialization of AR instructions and the innovative dimension associated with AR. For other participants, their preference is explained by a better knowledge and a greater familiarity with computer. As observed in some AR studies (Drouot et al., 2021b), the level of familiarity with new technologies may have an impact on user performance and thus explain the increased assembly times associated with AR.

As many studies have highlighted the importance of using a combination of physiological, subjective and behavioral measures to assess mental workload, we have chosen a combination of dual-task paradigm, ocular measures and NASA TLX in our study. Our comparison of the three assessment methods used in our study, showed that ocular and secondary task measurements were sensitive to mental workload and showed the same pattern of results. This supports the relevance of using multiple measures to increase methodological confidence in the data.

Results of behavioral measurement (i.e., secondary task), indicate that participants detected auditory stimuli less often when using AR instructions. This suggests that mental workload is greater when using AR. This is supported by participants' reaction times auditory stimuli, which are also longer when using AR. The use of AR would result in a higher mental workload and thus prevent users from responding as effectively and quickly to an external stimulus as when using a computer. These results can be related to cognitive distraction (Kim and Gabbard, 2019), in which AR impacts the allocation of users' visual attention. In our study, virtual objects in AR seem to attract user's attention at the expense of detecting non-virtual stimuli.

Data from the behavioral measurement suggested that the effect of AR differs depending on the complexity of the assembly to perform. Indeed, AR leads to longer reaction times and more omissions compared to the computer, only when the assembly workstation is simple. These data are consistent with some studies suggesting that AR is only effective and beneficial when the task to perform has a higher level of complexity (Wiedenmaier et al., 2003). Although there is no gain here from the use of AR when performing a complex task, the negative cost associated with AR disappears and the level of mental workload would be equivalent to that of computer instructions.

The same pattern of results observed for the behavioral data, is also found for the physiological data. Indeed, blink rate is lower when using AR than when using computer instructions, suggesting a greater mental workload (Ahlstrom and Friedman-Berg, 2006; Cain, 2007). Furthermore, the interaction effect observed between complexity and instruction medium on blink rate and pupil size are also shows that mental workload generated by AR for a simple assembly becomes equivalent to that of the computer when the assembly is complex. The analysis of pupil size shows that the pupil is smaller in AR than with computer. Although this may indicate a greater mental workload in AR (Peinkhofer et al., 2019), and the fact that pupil size can be used in non-controlled environment (Bækgaard et al., 2019), the significant difference in brightness between the computer and AR conditions in our study does not allow us to conclude on mental workload variation based on this measure.

Both behavioral (reaction time and omission) and physiological (blink rate and pupil size) data strongly suggested that mental workload increases in AR and that this increase is observed in the case of simple assembly. We suggest that these data could be explained by the fact that the presentation of assembly instructions in AR would present both a cost and a benefit, the ratio of which would vary according to the complexity. Thus, when the assembly is simple (low mental and physical demand), the potential benefit of AR is reduced to its minimum, leaving only the cost associated with its use. However, when the assembly is more complex, benefit and cost offset each other. Future studies using different levels of physical and mental complexity could address this hypothesis and identify the level of complexity required for the benefit of AR to exceed the cost of using it.

Although the behavioral and physiological measures indicate results in the same direction, the subjective measurement of NASA TLX does not indicate a greater perceived mental workload in AR. The fact that physiological and behavioral data show different directions has already been observed (Luque-Casado et al., 2016; Miyake, 2001). The study of the different NASA TLX dimensions indicates that participants perceive and report a difference only on the physical dimension between simple and complex workstation. As the HMD was worn regardless of the condition, the physical load is probably not related to it, but more to the fact that the complex workstation requires more strength due to the use of screwdrivers and the weight of the ball (approximately 15kg).

5. Conclusion

This study aimed at evaluating effectiveness of assembly performance and mental workload when using instruction displayed in AR. An AR application developed in the Microsoft HoloLens 2 (Lavric et al., 2021) was tested directly in industrial assembly lines. To our knowledge, this study is the first to study mental workload directly in an industrial setting, using a combination of three different types of measurement (objective, subjective and behavioral).

Our results show that AR does not provide any benefit in terms of assembly performance effectiveness, as it does not reduce errors and it can increase assembly time. Regarding mental workload, four measurements (two physiological and two behavioral) indicate that AR results in a greater mental workload when the assembly workstation is simple but that, when it is complex, AR seems equivalent to computer instructions. Behavioral data indicates that when participants use AR, they detect external stimuli less often and less quickly, suggesting a cognitive distraction in AR. This cognitive distraction underlines the fact that the attentional resources of an operator using AR could be less available for detection of external events (danger, alert) and thus play an important role in the occurrence of work accidents. Although AR does not necessarily reduce attention and awareness for the real-world environment (Aromaa et al., 2020), AR should be used with caution in industrial use. We recommend, that any integration of AR instructions in an industrial setting should consequently be associated with an evaluation protocol in order to quantify the potential impact on performance and to ensure that the proposed AR solution brings more benefits than the existing solution. It could therefore be interesting to work on the ergonomics of the AR interface and more specifically how to manage spatialization to reduce the associated cost.

Our study presents some limitations, further studies are thus needed in order to better assess the conditions under which AR could be used in industrial settings. Indeed, it might be interesting to evaluate how mental workload in AR varies according to levels of complexity while controlling the physical difficulty associated, which was not controlled for in this study. A better control of environmental conditions (e.g., brightness levels) would increase the validity of the data from the various physiological measurements (e.g. we could not use here the pupils' diameter data). Finally, our study only focused on the use of AR for assembly guidance and not for learning. However, the study of learning in AR plays a major role in industrial training. Thus, it would be interesting to evaluate more precisely how the parameters measured in this study may vary after several cycles of use and how learning and increasing familiarity influence the mental workload with time of use. Further studies could then analyze the use of AR for longer and more regular assembly cycles, as well as cycles without AR instruction support.

Funding

This work was supported by elm.leblanc S.A.S, as part of the IRON-MEN project, and the French National Association of Research and Technology (ANRT) with the Grant CIFRE 2018/0994. The sponsors did not have any specific role in the production of the paper.

Bibliography

- Ahlstrom, U., Friedman-Berg, F.J., 2006. Using eye movement activity as a correlate of cognitive workload. *Int. J. Ind. Ergon.* 36, 623–636. <https://doi.org/10.1016/j.ergon.2006.04.002>
- Aromaa, S., Väättänen, A., Aaltonen, I., Goriachev, V., Helin, K., Karjalainen, J., 2020. Awareness of the real-world environment when using augmented reality head-mounted display. *Appl. Ergon.* 88. <https://doi.org/10.1016/j.apergo.2020.103145>
- Bækgaard, P., Jalaliniya, S., Hansen, J.P., 2019. Pupillary measurement during an assembly task. *Appl. Ergon.* 75, 99–107. <https://doi.org/10.1016/j.apergo.2018.09.004>
- Bednarik, R., Bartczak, P., Vrzakova, H., Koskinen, J., Elomaa, A.P., Huotarinen, A., De Gómez Pérez, D.G., Von und zu Fraunberg, M., 2018. Pupil size as an indicator of visual-motor workload and expertise in microsurgical training tasks. *Eye Track. Res. Appl. Symp.* <https://doi.org/10.1145/3204493.3204577>
- Blattgerste, J., Strenge, B., Renner, P., Pfeiffer, T., Essig, K., 2017. Comparing conventional and augmented reality instructions for manual assembly tasks, in: *ACM International Conference Proceeding Series. Association for Computing Machinery*, pp. 75–82. <https://doi.org/10.1145/3056540.3056547>
- Bode, M., 2019. Evaluation of an augmented reality assisted manufacturing system for assembly

guidance.

- Boyer, S., 2017. Contribution de l'analyse du signal vocal à la détection de l'état de somnolence et du niveau de charge mentale. Université Paul Sabatier - Toulouse III.
- Cain, B., 2007. A Review of the Mental Workload Literature. Def. Res. Dev. Toronto 4-1-4–34.
- Caudell, T.P., Mizell, D.W., 1992. Augmented reality: an application of heads-up display technology to manual manufacturing processes. Institute of Electrical and Electronics Engineers (IEEE), pp. 659–669 vol.2. <https://doi.org/10.1109/hicss.1992.183317>
- Cegarra, J., Chevalier, A., 2008. The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. Behav. Res. Methods 40, 988–1000. <https://doi.org/10.3758/BRM.40.4.988>
- Charles, R.L., Nixon, J., 2019. Measuring mental workload using physiological measures: A systematic review. Appl. Ergon. <https://doi.org/10.1016/j.apergo.2018.08.028>
- de Waard, D., 1996. The Measurement of Drivers' Mental Workload. Groningen.
- Drouot, M., Le Bigot, N., Bolloc, J., Bricard, E., De Bougrenet, J., Nourrit, V., 2021a. The visual impact of augmented reality during an assembly task ☆. Displays 66, 101987. <https://doi.org/10.1016/j.displa.2021.101987>
- Drouot, M., Le Bigot, N., De Bougrenet, J., Nourrit, V., 2021b. Effect of Context and Distance Switching on Visual Performances in Augmented Reality. 2021 IEEE Conf. Virtual Real. 3D User Interfaces Abstr. Work. 476–477. <https://doi.org/10.1109/VRW52623.2021.00120>
- 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, in: ACM International Conference Proceeding Series. Association for Computing Machinery, pp. 222–229. <https://doi.org/10.1145/3056540.3056548>
- Funk, M., Kosch, T., Schmidt, A., 2016. Interactive worker assistance: Comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions, in: UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Association for Computing Machinery, Inc, pp. 934–939. <https://doi.org/10.1145/2971648.2971706>
- Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., Tecchia, F., 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. Interact. Learn. Environ. 23, 778–798. <https://doi.org/10.1080/10494820.2013.815221>
- Haniff, D.J., Baber, C., 2003. User evaluation of augmented reality systems, in: Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003. pp. 505–511. <https://doi.org/10.1109/IV.2003.1218032>
- 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](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hicks, T.G., Wierwille, W.W., 1979. Comparison of Five Mental Workload Assessment Procedures in a Moving-Base Driving Simulator. Hum. Factors J. Hum. Factors Ergon. Soc. 21, 129–143. <https://doi.org/10.1177/001872087902100201>
- Hou, L., Wang, X., 2013. A study on the benefits of augmented reality in retaining working memory in assembly tasks: A focus on differences in gender. Autom. Constr. 32, 38–45. <https://doi.org/10.1016/j.autcon.2012.12.007>
- Hou, L., Wang, X., Bernold, L., Love, P.E.D., 2013. Using animated augmented reality to cognitively guide assembly, in: Journal of Computing in Civil Engineering. pp. 439–451. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000184](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000184)
- Hou, L., Wang, X., Truijens, M., 2015. Using augmented reality to facilitate piping assembly: An experiment-based evaluation. J. Comput. Civ. Eng. 29. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000344](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000344)
- Jafari, M.J., Zaeri, F., Jafari, A.H., Payandeh Najafabadi, A.T., Al-Qaisi, S., Hassanzadeh-Rangi, N., 2020. Assessment and monitoring of mental workload in subway train operations using physiological, subjective, and performance measures. Hum. Factors Ergon. Manuf. 30, 165–175.

<https://doi.org/10.1002/hfm.20831>

- Kassner, M., Patera, W., Bulling, A., 2014. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction, in: *UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, Inc, pp. 1151–1160. <https://doi.org/10.1145/2638728.2641695>
- Kellogg, R.T., 1990. Effectiveness of Prewriting Strategies as a Function of Task Demands. *Am. J. Psychol.* 103, 327. <https://doi.org/10.2307/1423213>
- Kim, H., Gabbard, J.L., 2019. Assessing Distraction Potential of Augmented Reality Head-Up Displays for Vehicle Drivers. *Hum. Factors* 18720819844845. <https://doi.org/10.1177/0018720819844845>
- Lavric, T., Bricard, E., Preda, M., Zaharia, T., 2021. Exploring Low-Cost Visual Assets for Conveying Assembly Instructions in AR, in: *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. Kocaeli, Turkey.
- Loch, F., Quint, F., Brishtel, I., 2016. Comparing video and augmented reality assistance in manual assembly, in: *Proceedings - 12th International Conference on Intelligent Environments, IE 2016*. Institute of Electrical and Electronics Engineers Inc., pp. 147–150. <https://doi.org/10.1109/IE.2016.31>
- Luque-Casado, A., Perales, J.C., Cárdenas, D., Sanabria, D., 2016. Heart rate variability and cognitive processing: The autonomic response to task demands. *Biol. Psychol.* 113, 83–90. <https://doi.org/10.1016/j.biopsycho.2015.11.013>
- Miller, S., 2001. Literature Review Workload measures.
- Min, J.S., Kwak, G., Hwang, W., 2020a. Comparison among types of assembly manuals based on paper, video and augmented reality. *ICIC Express Lett.* 14, 303–310. <https://doi.org/10.24507/icicel.14.03.303>
- Min, J.S., Kwak, G., Hwang, W., 2020b. Comparison among types of assembly manuals based on paper, video and augmented reality. *ICIC Express Lett.* 14, 303–310. <https://doi.org/10.24507/icicel.14.03.303>
- Miyake, S., 2001. Multivariate workload evaluation combining physiological and subjective measures. *Int. J. Psychophysiol.* 40, 233–238. [https://doi.org/10.1016/S0167-8760\(00\)00191-4](https://doi.org/10.1016/S0167-8760(00)00191-4)
- Neumann, U., Majoros, A., 1998. Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance, in: *Proceedings - Virtual Reality Annual International Symposium*. pp. 4–11. <https://doi.org/10.1109/vrais.1998.658416>
- Peinkhofer, C., Knudsen, G.M., Moretti, R., Kondziella, D., 2019. Cortical modulation of pupillary function: Systematic review. *PeerJ Comput. Sci.* 2019. <https://doi.org/10.7717/peerj.6882>
- Piolat, A., Olive, T., Roussey, J.Y., Thunin, O., Ziegler, J.C., 1999. SCRIPTKELL: A tool for measuring cognitive effort and time processing in writing and other complex cognitive activities. *Behav. Res. Methods, Instruments, Comput.* 31, 113–121. <https://doi.org/10.3758/BF03207701>
- Ratcliff, R., 1993. Methods for Dealing With Reaction Time Outliers. *Psychol. Bull.* 114, 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- Rodriguez, F.S., Saleem, K., Spilski, J., Lachmann, T., 2021. Performance differences between instructions on paper vs digital glasses for a simple assembly task. *Appl. Ergon.* 94. <https://doi.org/10.1016/J.APERGO.2021.103423>
- Stedmon, A.W., Stone, R.J., 2001. Re-viewing reality: Human factors of synthetic training environments. *Int. J. Hum. Comput. Stud.* 55, 675–698. <https://doi.org/10.1006/ijhc.2001.0498>
- Tang, A., Owen, C., Biocca, F., Mou, W., 2003. Comparative effectiveness of augmented reality in object assembly, in: *Proceedings of the Conference on Human Factors in Computing Systems - CHI '03*. ACM Press, New York, New York, USA, p. 73. <https://doi.org/10.1145/642611.642626>
- Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., Zhang, T., 2019. A systematic review of physiological measures of mental workload. *Int. J. Environ. Res. Public Health.* <https://doi.org/10.3390/ijerph16152716>
- Tatić, D., Tešić, B., 2017. The application of augmented reality technologies for the improvement of

- occupational safety in an industrial environment. *Comput. Ind.* 85, 1–10.
<https://doi.org/10.1016/j.compind.2016.11.004>
- Wanyan, X., Zhuang, D., Zhang, H., 2014. Improving pilot mental workload evaluation with combined measures, in: *Bio-Medical Materials and Engineering*. pp. 2283–2290.
<https://doi.org/10.3233/BME-141041>
- Werrlich, S., Eichstetter, E., Nitsche, K., Notni, G., 2017. An Overview of Evaluations Using Augmented Reality for Assembly Training Tasks. *Int. J. Comput. Inf. Eng.* 11, 1129–1135.
- Whelan, R., 2008. Effective analysis of reaction time data. *Psychol. Rec.* 58, 475–482.
<https://doi.org/10.1007/BF03395630>
- Wiedenmaier, S., Oehme, O., Schmidt, L., Luczak, H., 2003. Augmented Reality (AR) for Assembly Processes Design and Experimental Evaluation. *Int. J. Hum. Comput. Interact.* 16, 497–514.
https://doi.org/10.1207/S15327590IJHC1603_7
- Yang, Z., Shi, J., Jiang, W., Sui, Y., Wu, Y., Ma, S., Kang, C., Li, H., 2019. Influences of augmented reality assistance on performance and cognitive loads in different stages of assembly task. *Front. Psychol.* 10, 1–17. <https://doi.org/10.3389/fpsyg.2019.01703>
- Zhong, X., Liu, P., Georganas, N.D., Boulanger, P., 2003. Entwurf einer kollaborativen Augmented-Reality-Anwendung für industrielles Training (Designing a Vision-based Collaborative Augmented Reality Application for Industrial Training). *it - Inf. Technol.* 45.
<https://doi.org/10.1524/itit.45.1.7.19032>