



**HAL**  
open science

# Methodologies and tools for expert knowledge sharing in manual assembly industries by using augmented reality

Traian Lavric

## ► To cite this version:

Traian Lavric. Methodologies and tools for expert knowledge sharing in manual assembly industries by using augmented reality. Emerging Technologies [cs.ET]. Institut Polytechnique de Paris, 2022. English. NNT : 2022IPPAS015 . tel-03940858

**HAL Id: tel-03940858**

**<https://theses.hal.science/tel-03940858>**

Submitted on 16 Jan 2023

**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.

# Methodologies and tools for expert knowledge sharing in manual assembly industries by using augmented reality

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à Télécom SudParis

École doctorale n°626 Institut Polytechnique de Paris (ED IP Paris)  
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Évry, le 02/12/2022, par

**Traian LAVRIC**

Composition du Jury :

Laurence NIGAY <i>Professeure, Université Grenoble Alpes (EHCI)</i>	Présidente
Thierry DUVAL <i>Professeur, IMT Atlantique (Lab-STICC)</i>	Rapporteur
Jean-Yves DIDIER <i>Maître de conférences, Université d'Evry Val d'Essonne (IBISC)</i>	Rapporteur
Fabien DUMON <i>Lead of Holographic Academy, Airbus Defence &amp; Space</i>	Examineur
Titus ZAHARIA <i>Professeur, Télécom SudParis (SAMOVAR)</i>	Directeur de thèse
Marius PREDA <i>Maître de conférences, Télécom SudParis (SAMOVAR)</i>	Co-Encadrant
Emmanuel Bricard <i>Directeur des systèmes d'information, ELM LEBLANC (Bosch France)</i>	Co-Encadrant industriel



# TABLE OF CONTENTS

Table of contents .....	iii
List of tables.....	vii
List of figures .....	ix
List of acronyms and abbreviations.....	xii
Acknowledgements .....	xiv
Abstract .....	xvi
<b>Chapter 1 . Introduction.....</b>	<b>1</b>
1.1 Research problem and motivation .....	2
1.2 Thesis outline .....	4
1.3 Contributions .....	6
<b>Chapter 2 . Industrial case study .....</b>	<b>7</b>
2.1 Introduction and objectives .....	8
2.2 Assembly context exploration at elm.leblanc.....	9
2.2.1 Research problem validation and refining.....	9
2.2.2 Human workforce and industrial automation .....	10
2.2.3 Training in manufacturing: importance and limitations .....	11
2.2.4 Assembly organization, process, and resources .....	14
2.2.5 Industrial requirements and challenges .....	23
2.3 Prototyping and preliminary studies .....	25
2.3.1 Informal field studies .....	25
2.3.2 User study: AR training methodology for assembly operations .....	28
2.3.3 Prototyping and evaluations .....	29
2.4 Conclusion.....	37
<b>Chapter 3 . Related work .....</b>	<b>38</b>
3.1 Content authoring in AR .....	39
3.1.1 State of the art surveys and classification .....	39
3.1.2 Methods and systems .....	40
3.2 AR training: information conveyance and guidance.....	45
3.2.1 AR training versus traditional training .....	45
3.2.2 Use cases in manufacturing sectors.....	46
3.2.3 Non-industrial AR training research works .....	47

3.3	AR adoption in the industry .....	49
3.4	Conclusion.....	50
<b>Chapter 4 . An analysis on information representation and conveyance in AR and HCI considerations .....</b>		<b>52</b>
4.1	Introduction .....	53
4.2	Digital assets .....	54
4.2.1	General classification .....	54
4.2.2	Media guidance strength analysis and cognitive considerations .....	56
4.2.3	Transient information effect. Segmentation and playback control.....	63
4.2.4	Cueing in hand-manipulated operations .....	64
4.2.5	Verbalizers <i>versus</i> visualizers and the <i>multimedia effect</i> .....	64
4.2.6	User and environment-adapted guidance .....	65
4.2.7	Authoring cost evaluation: a study on AR HMD-based content creation.....	66
4.2.8	Conclusions .....	69
4.3	Visual information conveyance mediums in AR (hardware devices).....	70
4.3.1	Handheld display (HHD).....	70
4.3.2	Head mounted display (HMD) .....	70
4.3.3	Desktop/monitor-based.....	71
4.3.4	Spatial augmented reality (SAR) .....	71
4.3.5	Screen-based video see-through .....	72
4.4	Human-computer interaction (HCI) considerations .....	72
4.4.1	Usability and usefulness.....	73
4.4.2	User interface and interaction techniques .....	74
4.5	Conclusions and guidelines .....	75
<b>Chapter 5 . ATOFIS, the proposed AR content authoring system.....</b>		<b>78</b>
5.1	Challenges and requirements .....	79
5.1.1	Technological challenges and expert knowledge .....	79
5.1.2	User acceptance and shop floor reluctance .....	79
5.1.3	System performance and usability .....	80
5.1.4	Safety, psychological and physiological concerns.....	80
5.1.5	Efficiency and viability.....	80
5.1.6	KPI and costs .....	81

5.2	Methodology and system implementation .....	81
5.2.1	Concept and principles.....	81
5.2.2	Authoring procedure and guidelines .....	87
5.2.3	Authoring extensions .....	89
5.2.4	System implementation.....	91
5.2.5	Authoring an AR work instruction.....	91
5.2.6	Interaction techniques .....	94
5.2.7	Key functional and design aspects.....	95
<b>Chapter 6 .</b>	<b>The proposed AR training system .....</b>	<b>97</b>
6.1	Information representation of assembly operations in AR .....	98
6.2	Information conveyance: principles and techniques.....	98
6.2.1	What.....	99
6.2.2	Where.....	99
6.2.3	How .....	100
6.3	Addressing different assembly environments .....	100
6.3.1	Seated assembly environment.....	101
6.3.2	Dynamic and mobile workstation environment .....	101
6.4	System implementation.....	102
6.5	Interaction techniques.....	103
6.6	Key functional and design aspects.....	105
<b>Chapter 7 .</b>	<b>Evaluations .....</b>	<b>107</b>
7.1	Training experiments .....	108
7.1.1	Low-cost instructions set training (field experiment 1 - T-FE1).....	109
7.1.2	Comparative evaluation: low-cost <i>versus</i> CAD-complemented instruction set training (field experiment 2 - T-FE2) .....	115
7.1.3	Discussion.....	122
7.2	Authoring experiments .....	124
7.2.1	Comparative authoring evaluation: ATOFIS <i>versus</i> Guides (field study - AT-FS)	124
7.2.2	Full comparative evaluation between ATOFIS and Guides (laboratory experiment - AT-LE)	128
<b>Chapter 8 .</b>	<b>Conclusion and future work.....</b>	<b>151</b>
8.1	Conclusion.....	152

---

8.1.1	Case study: manual assembly and industrial context.....	152
8.1.2	State-of-the-art in AR authoring and training.....	152
8.1.3	Information representation guidelines and HCI considerations.....	152
8.1.4	System implementation and evaluation .....	153
8.1.5	Final remarks.....	155
8.2	Future work.....	156
8.2.1	Large-scale system evaluation .....	156
8.2.2	User interface optimization .....	156
8.2.3	Automatic assembly validation and personalized training experiences.....	157
8.2.4	Context-aware, adaptive, and automatic authoring.....	157
	References.....	158

## LIST OF TABLES

Table 2-1. Departments and people involved in the considered industrial case study at elm.leblanc.....	8
Table 2-2. Questionnaire for line managers et elm.leblanc, designed by us, to identify details regarding the traditional training procedures conducted in the factory. ....	12
Table 2-3. Manual assembly operations (i.e., verbs) and their occurrences identified in the existing paper instructions of the assembly line L2.....	19
Table 2-4. Manual assembly operations identified in an automotive use case. ....	20
Table 2-5. Requirements and challenges stated by experts from the factory at elm.leblanc....	24
Table 2-6. Proposed vocal commands and gestures used during the authoring .....	33
Table 2-7. ATOFIS-BETA vs. Guides: presentation times. ....	35
Table 2-8. ATOFIS-BETA vs. Guides: reported authoring times. ....	36
Table 2-9. ATOFIS-BETA vs. Guides: systems complexity. ....	36
Table 4-1. Media types and their guidance strength.....	56
Table 4-2. Visual assets suitability classification for information representation in AR.....	57
Table 4-3. Visual assets characteristics regarding optimal ways of information representation in AR. Red dot = not appropriate, green dot = appropriate, blank cells = depends on multiple factors. ....	57
Table 4-4. Reported time and difficulty estimations for capturing low-cost assets (i.e., text, FPV photo and FPV video) by using Hololens 2. ....	69
Table 5-1. ATOFIS: summary of the authoring interaction techniques, user input and results. ....	95
Table 5-2. ATOFIS: summary of the 2W1H-based authoring workflow.....	95
Table 6-1. A summary of the interaction techniques used during training. ....	105
Table 6-2. Training UI and information conveyance.....	105
Table 7-1. T-FE1: error rates reported during the AR training. ....	111
Table 7-2. T-FE1: assembly time measurements reported during the AR training. ....	112
Table 7-3. The System Usability Scale questionnaire. ....	114
Table 7-4. T-FE2: participant division by instruction set (low-cost vs. CAD) and assembly experience (novice vs. experienced).....	115
Table 7-5. T-FE2: reported evaluation results – i.e., error rate, assembly completion times and instruction reading times, for each cycle of the two considered groups. ....	116
Table 7-6. T-FE2: reported error rates per group and subgroup during the WEC.....	117
Table 7-7. T-FE2: average ATCs and IRTs reported per group, on the entire assembly set (ATx) and on the CAD-complemented instructions (AT2), during the WEC. ....	118
Table 7-8. T-FE2: reported mean (M) and standard deviation (SD) of the ACT and IRT values over each of the three assembly cycles. ....	119
Table 7-9. AT-FS: authoring data composition per each of the five considered workstations. ....	126
Table 7-10. AT-FS: average number of media assets used for authoring an AR work instruction, per each of the five considered workstations. ....	126



---

Table 7-11. AT-FS: a comparison between authoring times reported by ATOFIS and Guides, respectively, for each of the workstations. ....	127
Table 7-12. AT-FS: average AR work instruction authoring time reported by the two systems, per workstation. ....	127
Table 7-13. AT-LE: AR and assembly experience of participants to the experiment. ....	129
Table 7-14. AT-LE: experiment procedure workflow and description. ....	130
Table 7-15. AT-LE: instruction set used for the experiment, including the information associated to each corresponding AR work instruction. ....	132
Table 7-16. AT-LE: reported training times for both cycles (systems). ....	136
Table 7-17. AT-LE: reported authoring times for both assembly cycles (systems). ....	144
Table 7-18. AT-LE: results of a paired two sample for means on the reported authoring times of the two systems, for both assembly cycles. ....	144

## LIST OF FIGURES

Figure 2-1. An overview of the assembly line L2 at elm.leblanc where most of the experiments conducted during this thesis have been conducted. ....	9
Figure 2-2. The first workstation (i.e., W201) of the assembly line L2 at elm.leblanc. The tools and assembly components are highlighted and labeled. ....	14
Figure 2-3. Common workstation components and organization: A (red) – storage areas, B (yellow) – assembly tools and informatic systems, C (blue) – assembly structure and D (green) – assembly cart. ....	16
Figure 2-4. Relevant existing digital data at elm.leblanc. ....	23
Figure 2-5. Field experiment: assembly of a boiler in progress. ....	26
Figure 2-6. Assembly components: a) small (e.g., screws, clips) and b) flexible (e.g., cables)...	27
Figure 2-7. Assembly complexity: a boiler assembly state difficult to represent digitally in AR environments. ....	27
Figure 2-8. AR Boiler Showroom. An AR application for visualizing and inspecting in AR the corresponding 3D model of a boiler produced at elm.leblanc. ....	29
Figure 2-9. AR TrainingAssistant screenshots. a) Welcome screen, b) “Instructions” screen, c) “Statistics” screen, d) A 3D animation of an assembly operation. ....	30
Figure 2-10. ATOFIS-BETA authoring workflow illustration: a) Location anchor placement – point with index fingertip & use vocal command, b) Photo capture – vocal command, c) Photo element manipulation – hand gestures, d) Capture and manipulate additional photos/videos.....	32
Figure 2-11. ATOFIS-BETA training procedure illustration: a) Directional arrow indicating the picking location, b) Location anchor indicating the assembly location, c) The trainee picks the assembly component: the location anchor is hidden, and a new indication arrow is displayed, d) The trainee identifies the next anchor, visualizes the augmentation media, and performs the assembly operation.....	33
Figure 2-12. ATOFIS-BETA authoring: summary of authoring key elements.....	34
Figure 2-13. ATOFIS-BETA information conveyance: summary of training key elements.....	34
Figure 2-14. ATOFIS-BETA vs. Guides authoring workflow and stages: the crossed-out blocks are no longer required by ATOFIS-BETA. ....	35
Figure 4-1. Visual assets used in industrial augmented reality (Gattullo, Evangelista, et al., 2020). ....	55
Figure 4-2. Abstract versus concrete AR visualizations. ....	60
Figure 4-3. A 3D capture of the assembly line L2 at elm.leblanc, obtained with Hololens 2. ....	66
Figure 4-4. Microsoft Hololens 2, a state-of-the-art AR see-through HMD. ....	71
Figure 4-5. A prototype of a SAR system proposed in (Uva et al., 2018).....	72
Figure 5-1. Authoring workflows of Guides and ATOFIS. A) Guides, a 3-step authoring composed of media capture, PC authoring and HMD authoring; B) ATOFIS, a 1-step in-situ, HMD-only authoring.....	87
Figure 5-2. Low-cost visual assets used to capture and represent the assembly information via AR in our proposed approach. ....	89

Figure 5-3. ATOFIS authoring workflow overview for an AR instruction: (1) what: assembly description - text, (2) where: location indication - arrow and (3) how: assembly demonstration - photo or video.....	92
Figure 5-4. AR authoring example. a) Step 1. Insert text instruction by using the virtual keyboard or dictation; b) Step 1 validated, step 2 active; c) Step 2. Positioning of the location arrow by using far interaction technique; d) Step 2 validated, step 3 active; e) Step 3. Photo-capture view; f) Step 3. Photo taken, the author validates or removes it. ....	93
Figure 6-1. Informal field experiment conducted in a different manufacturing factory. The worker is required to seat during training (i.e., the execution of the assembly procedure). ..	101
Figure 6-2. Example of how the assembly information is conveyed via AR. Instruction 1: “Grab 2 uprights”. a) Text & indication arrow; b) Location arrow & FPV photo; Instruction 2: “Place the first upright”. c) Text & indication arrow; d) FPV video & CAD model. ....	102
Figure 7-1. T-FE1 & T-FE2: the final product assembly obtained by participants during the two training field experiments.....	108
Figure 7-2. T-FE1 & T-FE2: the most relevant operations of the assembly procedure used during the training field experiments. ....	109
Figure 7-3. T-FE1: a participant executing an assembly during the training field experiment. ....	110
Figure 7-4. F-FE1: average assembly time (AvgTotal) and reading times per instruction (AvgRead), for a) Cycle 1, b) Cycle 2 and c) Cycle 3. ....	113
Figure 7-5. T-FE1: box and whiskers chart illustrating the assembly times for each of the four cycles.....	113
Figure 7-6. T-FE1: box and whiskers chart illustrating the raw SUS scores reported by the participants. ....	114
Figure 7-7. T-FE2: Error-prone assembly example: participant places a symmetrical assembly component incorrectly during the training – i.e., upside-down.....	117
Figure 7-8. T-FE2: Reported ACT and IRT scores for CAD-complemented instructions (of type AT2) during the WEC. ....	119
Figure 7-9. T-FE2: reported ACT values per each assembly instruction, over the three cycles.....	120
Figure 7-10. T-FE2: reported NASA-TLX scores per dimension and subgroups. ....	121
Figure 7-11. T-FE2: box and whiskers chart illustrating the raw SUS training scores reported by the participants on each dimension. ....	122
Figure 7-12. AT-LE: the assembly workstation setup with its main components and a participant performing an assembly operation.....	129
Figure 7-13. AT-LE: assembly operations 2 and 3, considered error-prone because of the symmetrical nature of the assembly component. ....	135
Figure 7-14. AT-LE: reported training times per participant, per cycle (training system). ....	137
Figure 7-15. AT-LE: average NASA-TLX scores reported by participants on each dimension, for both training systems. ....	138
Figure 7-16. AT-LE: NASA-TLX scores reported by each participant, for both training systems. ....	138

---

Figure 7-17. AT-LE: normalized SUS scores reported by each participant, for both training systems. ....	139
Figure 7-18. AT-LE: average normalized SUS scores reported by all participants on each dimension, for both training systems. ....	139
Figure 7-19. AT-LE: authoring times reported by each participant, per cycle (authoring system). ....	145
Figure 7-20. AT-LE: average NASA-TLX scores reported by participants on each dimension, for both authoring systems. ....	145
Figure 7-21. AT-LE: NASA-TLX scores reported by each participant, for both authoring systems. ....	146
Figure 7-22. AT-LE: normalized SUS scores reported by each participant, for both authoring systems. ....	146
Figure 7-23. AT-LE: average normalized SUS scores reported by all participants on each dimension, for both training systems. ....	147

## LIST OF ACRONYMS AND ABBREVIATIONS

<b>2D</b>	Two-Dimensional
<b>2W1H</b>	What, Where and How Principle
<b>3D</b>	Three-Dimensional
<b>ACT</b>	Assembly Completion Time
<b>AR</b>	Augmented Reality
<b>API</b>	Application Programming Interface
<b>ARAF</b>	Augmented Reality Application Format
<b>ARF</b>	Augmented Reality Framework
<b>ARML</b>	Augmented Reality Markup Language 2.0 (ARML 2.0)
<b>AT</b>	Authoring Tool
<b>ATOFIS</b>	Authoring Tool On-the-Fly and In-Situ
<b>CAD</b>	Computer-Aided Design
<b>CI</b>	Contextual Inquiry
<b>CIO</b>	Chief Information Officer
<b>CV</b>	Computer Vision
<b>DB</b>	Database
<b>EHS</b>	Environment, health, and safety
<b>elm.leblanc</b>	e.l.m. leblanc SAS
<b>ETSI</b>	European Telecommunications Standards Institute
<b>FoV</b>	Field-of-View
<b>FPV</b>	First-Person View
<b>GUI/UI</b>	Graphical User Interface
<b>Guides</b>	Microsoft Dynamics 365 Guides
<b>HCD</b>	Human-Centered Design
<b>HCI</b>	Human-Computer Interaction
<b>HFA</b>	High-Functionality Application
<b>HHD</b>	Handheld Device
<b>HMD</b>	Head Mounted Display
<b>HS2</b>	Microsoft Hololens 2
<b>HUD</b>	Heads-Up Display
<b>IAR</b>	Industrial Augmented Reality
<b>IRT</b>	Instruction Reading Time
<b>ISO</b>	International Organization for Standardization
<b>IT</b>	Information Technology
<b>KPI</b>	Key Performance Indicator
<b>LCD</b>	Liquid Crystal Display
<b>MRTK</b>	Mixed Reality Toolkit
<b>MTM</b>	Methods-Time Measurement
<b>NASA-TLX</b>	NASA Task Load Index
<b>NUI</b>	Natural User Interface
<b>PC</b>	Personal Computer
<b>SAR</b>	Spatial Augmented Reality
<b>SF</b>	Screen-Fixed
<b>SUS</b>	System Usability Score

---

<b>TMU</b>	Time Measurement Unit
<b>TUI</b>	Tangible User Interface
<b>UX</b>	User Experience
<b>VR</b>	Virtual Reality
<b>Vuforia</b>	PTC Vuforia Expert Capture
<b>W3C</b>	World Wide Web Consortium
<b>WEC</b>	Workstation Exploration Cycle
<b>WF</b>	World-Fixed
<b>WYSIWYG</b>	What You See Is What You Get
<b>XML</b>	Extensible Markup Language

## ACKNOWLEDGEMENTS

This thesis would not have been possible without the help of numerous people. First and foremost, I would like to express my deepest appreciation to my mentor and advisor Marius PREDA, which helped me decide to pursue a Ph.D. in the first place and offered me the unique opportunity to conduct this research work under the best possible conditions. His ideas and judgement guided and inspired me throughout this exciting, arduous journey. Thanks to him, I became a better student and a better person. I am profoundly grateful for his advice and support during all these years.

Further, I would like to thank Emmanuel BRICARD, my advisor from elm.leblanc, Bosch, the industrial partner of my thesis. Thanks to his vast knowledge, incredible dedication, and ardent passion, I could get a comprehensive understanding of the manufacturing context in the early stages of my thesis, which allowed me to identify and adjust my research work on relevant industrial needs and challenges. His problem-solving strategies and methods helped me become more organized, pragmatic, and effective. Working with him was not only a pleasure, but also a fruitful and rewarding professional and personal experience.

My deepest and sincere gratitude goes to my doctoral advisor, Titus ZAHARIA, for his support and extremely insightful feedback, who always took his time to discuss my research ideas in detail and meticulously analyze my scientific proposals. Thanks to him, I learned to accept and reframe failure, learn from mistakes, and realize that only sustained hard work leads to success. His pragmatism, objectivism and encouragement inspired and motivated me to do better and provided me vital emotional support during difficult moments of my thesis.

I am very grateful for having had the opportunity to conduct my research work in an industrial environment. I would like to thank Philippe LAFORGE, CEO of elm.leblanc and an exemplary leader, for accepting my application and providing me the perfect assembly context for conducting my research, for his encouragements, friendliness, interest, and positive attitude towards my work. Sincere thanks to the IRON-MEN project fellows Abhaya-Dhathri ARIGE, Hugo DURCHON, Mathilde DROUOT and Pierre BEGOUT. It has been a pleasure and an honor brainstorming, designing, coding, traveling, sharing moments of conviviality, and learning from them. There are about 100 people at elm.leblanc that I interacted with during this journey, and I am extremely grateful for their support and feedback, especially to those which took their time to participate to my experiments and evaluate the tools that I developed during this thesis. I would like to send my special thanks to the BDO team, including Yannick GRALL, Diego Oliveira DA COSTA, Eric FARAUT, Nicolas JOURNET, Mohammad Nabeel KAUDEER, Eugene MAINENTI, Blandine DE ROCHEFORT and Manuel LOPES. I would as well like to send my greatest appreciation for their help, interest in my work and feedback to Lea CHASLE, Pascal RENON, Delphine DEROUSSSEN, Quentin THOMAS, Adao ALVARAES MEIRA, Frederic SANTHA, Seddik BOUANHAK, Evariste MOURETTE, Alain BIGOT, Thierry VANDANDAELE, Marion OTTAVI and Laurent TORTRAT. Thank you so much to all operators I interacted with, especially those of assembly line 2, for their friendliness and openness to help, by participating to the shop floor experiments and informal discussions, providing invaluable feedback for my research.

This thesis would not have been possible without the help of people from Telecom SudParis. I owe a deep sense of gratitude to my colleagues from ARTEMIS department for their support, friendliness, and good humor, including Chao CAO, Christian TULVAN, Guillaume GRANDINEAU, Kaouther OUENNICHE, Léa SAUNIER, Minderis KRIR, Nicolas HOFFMANN, Rania BENSIED, Tan-Khoa MAI and Zied LAHIANI. Special thanks to my ARTEMIS colleagues and to TSP students who accepted to spend more than half a day participating to my last experiment and provide vital statistical data that allowed me to conclude my research.

Further, I must express my deepest gratitude to Evelyne TARONI for her support and compassion, one of the persons who genuinely cared about my work as well as my wellbeing. She encouraged me to take a step further and extend the thesis to conduct a final evaluation of my proposal, one that I could not carry out within the planned timeframe of the thesis. This turned out to be the most comprehensive experiments of my thesis.

Finally, I want to thank my family. First and foremost, I would like to thank my wonderful parents for their unconditional support and continuous encouragement. I cannot express in words how grateful I am for being blessed with such parents and friends, who made everything they could to help me in all my endeavors. I would like to thank my beloved children, David, and Caroline, for their unconditional love, for delighting my days and bringing joy into my life, and for giving me strength and energy to pursue my goals. Finally, I would like to express my sincere and profound gratitude to my beloved life partner, Elena, for her care, understanding and support. It is thanks to her willingness and tremendous efforts that I could dedicate the time required to conduct this work and commit to my other duties and responsibilities. This accomplishment would not have been possible without her.



## ABSTRACT

To remain competitive in the ongoing industrial revolution, as known as Industry 4.0, manufacturing sectors must ensure high flexibility at the production level, a need best addressed by skilled human workforce. As traditional training becomes increasingly inefficient, finding a better way of training novice workers becomes a critical requirement. Literature suggests that augmented reality (AR), an emerging technology proposed by Industry 4.0, can potentially address this concern.

Despite the progress of AR in recent years, no significant industrial breakthrough can be noted, however. We found that most AR systems are elaborated and evaluated under controlled settings, without the implication of the eventual end users. Guided by literature recommendations, we conducted a long-term case study in a manual assembly factory, to identify needs and expectations that an AR training system should meet, to optimally address the considered industrial sector.

Further, we conducted an in-depth analysis on information representation and conveyance in AR, with respect to cognitive implications and authoring efforts. We explored as well human-computer interaction paradigms to identify principles and design guidelines for elaborating an AR tool dedicated to the shop floor context. We found that (1) the visual representation of the assembly expertise in AR can rely on spatially registered low-cost visual assets while (2) a human-centered design should be adopted during the elaboration of the proposed AR system, prioritizing usability, and usefulness rather than performance.

We defined a formalized visual representation (i.e., 2W1H principle) of assembly operations in AR, that considers authoring concerns and supports training performance. We proposed an HMD-only immersive authoring that allows one to capture his assembly expertise *in-situ*, during the assembly itself. The authoring is a one-step process, does not rely on existing data or external services and does not require AR or technical expertise, pre- or post-processing of data. During the training, the assembly information is conveyed in AR by following the same 2W1H principle, designed to guide a novice worker in a natural, non-intrusive manner, minimizing user input and UI clutter, and aiming to optimize comprehension and learning.

We evaluated our proposal by conducted several experiments. The first, conducted on a real-world assembly workstation, confirmed the hypothesis that spatially registered low-cost visual assets can effectively convey manual assembly expertise to novice workers via AR in an industrial setup. The findings of the second experiment supported the assumption that the worthiness of authoring CAD-based AR instructions in similar industrial context is questionable. A final experiment proved that the proposed AR system, including both authoring and training procedures, can be used effectively by novices in a matter of minutes. The overall reported feedback demonstrated the usability and efficiency of the proposed AR training approach, indicating that a similar system implementation could be successfully adopted in shop floor environments.

Future work should validate the reported experimental findings in large-scale industrial evaluations and propose reliable “intelligent” modules (e.g., assembly validation and feedback) to better assist novice workers during training and optimize the authoring procedure as well.

# Chapter 1. INTRODUCTION

---

## 1.1 Research problem and motivation

We are at the beginning of a revolution that is fundamentally changing the way we live, work, and relate to one another, through a set of emerging technology breakthroughs such as artificial intelligence (AI), internet of things (IoT), robotics, nanotechnology, and quantum computing, to name a few (Schwab, 2016). By employing these technological innovations, the current industrial revolution, also known under the label of Industry 4.0, promises to address today's main challenges faced by manufacturing industries, including process flexibility (Sethi & Sethi, 1990), training and skill development (Gleeson et al., 2013) and the complexity of the product development process (Elmaraghy et al., 2012).

To remain profitable and competitive, manufacturing factories must keep up with the technological advancements and be able to quickly adapt their products to the market requirements (Miqueo et al., 2020). Companies aim to have a substantial edge in product development to bring new products to market more quickly, consume fewer resources, and deliver higher quality designs (Elmaraghy et al., 2012). Individualized and customized products have become reality. Such individualization and customization require flexibility at production level in a cost-effective manner (Gleeson et al., 2013), a requirement that could potentially be addressed by exploiting Industry 4.0's enabling technologies. (Sethi & Sethi, 1990) and (G. Michalos et al., 2010) support this affirmation, claiming that one of the main challenges that modern assembly systems are faced with, is the cost-driven demand for faster and more secure ramp-up processes, one of the crucial points in the lifecycle of a production system (Brecher et al., 2013).

To address faster *time-to-market* requirements, modern production systems are faced with faster and more *right-first-time* ramp-up processes, which can manage shorter production planning cycles and increasing product variety, as well as the need to include new variants into existing production lines (G. Michalos et al., 2010). A recent survey conducted by (Miqueo et al., 2020) shows that mass customization brings great complexity that needs to be addressed at different levels from a holistic point of view and that developing methodologies for implementing Industry 4.0 remains an open topic of research. The growing amount of product variants leads to an increase of the perceived complexity of shop-floor workers (Brinzer et al., 2017), (Hu et al., 2008), (Toro et al., 2015). Further, to achieve high efficiency and adaptability and respond to ever-changing customer demands, companies should use flexible assembly lines (Barutçuoğlu & Azizoğlu, 2011), which guides the pursuit of greater employee flexibility (Schneider, 2018). (Cohen et al., 2019) claim that the main goal of flexible assembly systems in the Industry 4.0 era is to address the mass customization demand paradigm. As pointed out by (Grote, 2004), adequate management of uncertainty in production processes, driven by customer demand, political and economic factors, is critical for safe and efficient system design. The need of hiring seasonal workers to deal with high demand fluctuations and uncertainty becomes therefore mandatory. In the specific use-case on which we focus our research, related to human-operated assembly lines, (Metters, 1998) evaluates the complexity of seasonal demand environment and concludes that a set of well-established rules to cope with it cannot be provided.

It is evident therefore that manufacturing process require human intervention especially for flexibility reasons. In complex assembly systems, the workers are considered invaluable because of human flexibility advantages, which can manage the dynamic and fast changes caused by complexity (Mattsson et al., 2018). In other words, human labor remains inseparable from modern production systems because of the need of cognitive support to manage the challenges of multi-variant production and increased assembly complexity (D. Li et al., 2018). It is crucial therefore to understand the worker and support the assembly process to stay competitive in the Industry 4.0 era (Grote, 2004), (Mavrikios et al., 2007).

According to Gleeson *et al.* (Gleeson et al., 2013) however, the shortage of skilled workers is a very problematic issue faced by the industry today. The shortage is due to workforce aging and retirement, while, at the same time the skill requirements change because of AI, automation, and other advanced technologies and a lack of understanding and appeal of manufacturing jobs among younger cohorts (Moghaddam et al., 2021). Overall, manufacturing companies predict that the demand for workers will increase over the next few years (NAM, 2020) as the COVID-19 pandemic has exposed the need to produce more goods domestically (Lund et al., 2020), while only few of the remaining jobs will be automated in the near future (Moghaddam et al., 2021). In addition, the increasing adoption of new technologies eventually leads to raising skill requirements for new entry level workers (Berger et al., 2020), increasing therefore the skill gaps in areas such as assembly, maintenance, and inspection (*Closing the Industrial Skills Gap with Industrial AR | PTC*, 2019).

Training is still the most effective way to deal with this challenge. The problem with existing training methods however is that most manufacturing organizations train their employees in various ways, using different means, and achieving different levels of proficiency. Rarely does one find standardized and consistent training programs to develop worker skills; optimal training programs and methods are not generally known. (Mital et al., 1999) found that very few research studies using systematic procedures for developing employee training protocols have been conducted and validated within an industrial setting. Training in manufacturing sectors still seems far from being optimal, an issue that has extremely profound repercussions on the overall flexibility of the manufacturing process, time-to-market and ramp-ups, product customization and quality, aspects that eventually have a critical impact on satisfying the client demand and expectations.

One emerging technology of Industry 4.0 that can potentially address this concern is augmented reality (AR). AR has demonstrated its benefits as a knowledge-sharing tool, among other applications in a variety of domains including education, medicine, tourism, and entertainment (Dey et al., 2018), (Bellalouna, 2020). Studies show that AR training systems can be more efficient in terms of task completion time and error rates when compared to classical training procedures (i.e., paper instructions) (Tang et al., 2003), (Bosch et al., 2017), (Sanna et al., 2015), (Hahn et al., 2015), (Funk et al., 2016). Although AR has been investigated as a guidance tool for manufacturing process since more than two decades (Caudell & Mizell, 2003), only recently, technological advancements enabled a resurgence of the AR use. However, despite the exponential progress that AR has experienced in recent years, no significant breakthrough can

be noted in the industrial environment, due to various challenges (Martinetti et al., 2019), (Masood & Egger, 2019). AR systems have been mostly designed and evaluated in controlled environments, under laboratory settings, as recent surveys show (Merino et al., 2020), (Egger & Masood, 2020), (Bottani & Vignali, 2019). (Palmarini et al., 2018) claimed that AR technology is not sufficiently mature for complying with strong industrial requirements such as robustness and reliability. (Masood & Egger, 2019) identified and classified AR challenges into three main categories: technology, organization and environment and uncovered a gap between academic and industrial challenges. The authors suggested that field studies must be conducted to ensure the successful implementation of AR systems in industrial sectors.

The objective of this work is to follow these recommendations by conducting an industrial case study in a real-world manufacturing factory and elaborate AR expert knowledge sharing methodologies and systems optimally adapted to shop floor environments, with a focus on manual assembly process.

## 1.2 Thesis outline

The rest of the manuscript is structured as follows. Chapter 2 describes the research work carried out in the industrial context at elm.leblanc, specifically conducted to the purpose of the thesis. The main objective of this industrial research was to explore and analyze a real-world industrial assembly environment, identify, and define specific requirements and challenges imposed by similar manufacturing sectors, and finally, elaborate an AR training system optimally adapted to the considered assembly use-case. The work carried out during this case study provided an essential support for repeated evaluation and refinement of the proposed approach in an industrial environment, allowing us to follow a formative research approach.

Chapter 3 proposes a state-of-the-art analysis of AR content authoring methods and tools on one hand, and information conveyance approaches and training systems on the other hand. This survey focused on industrial AR (iAR), but considered topics of general interest as well, presumably relevant to the work conducted in this thesis.

In Chapter 4, we addressed an open research question which claims that there is no agreement in the literature regarding best ways of conveying instructions via AR. We conducted an in-depth analysis regarding AR information conveyance for guidance and training purposes. The first part of the chapter is dedicated to classifying digital assets by their ability and strength to convey information, while considering cognitive aspects and AR authoring efforts. The second part of the chapter explores AR information conveyance methods and systems. Finally, we propose a human-computer interaction (HCI) investigation to identify optimal ways of addressing technological and human challenges in the elaboration of an AR training system dedicated to shop floor context.

Chapter 5 describes the proposed AR authoring approach and details the system implementation. We start with a summary of the key challenges and requirements that the proposed system aims to address. Further, we discuss the authoring proposal by presenting the underlying concepts and principles, illustrate the user interface and detail the AR authoring

procedure. Finally, we present the most relevant visualization and interaction techniques and the key aspects of the authoring proposal.

The AR training proposal is developed in Chapter 6. We present the concept and information representation principles on which the training relies. We discuss the user interface and system behavior and exemplify how the proposed visualization and interaction techniques aim to address human-centered design paradigms and HCI considerations, in addition to effectiveness. We finally present a summary of the key functional and design aspects of our AR training proposal.

In Chapter 7, we present evaluations of the proposed system implementation. The first part of the chapter reports on two field experiments conducted to evaluate the implementation of the AR training proposal. The first experiment assesses the usability of the system and the effectiveness of low-cost visual assets. The second one, an extension of the first, evaluates the potential benefits of using CAD data to convey assembly information via AR, compared to the low-cost approach.

The second part of Chapter 7 evaluates the system implementation of the AR authoring proposal. The first study aims at providing a comparative estimate of the required time for authoring the AR instructions of an assembly workstation, against the most relevant commercial AR authoring tool dedicated to industry, Microsoft Guides. The second experiment reports on a complete evaluation of the proposed AR training system, that considers both the authoring and the training procedures. The main objective of the experiment was to comparatively assess and evaluate (i.e., effectiveness, error rate, time, perceived mental workload and system usability, user feedback) the overall system implementation against Guides.

Finally, Chapter 8 presents conclusions of the thesis's most relevant findings and opens some perspectives of future work.

### 1.3 Contributions

Our publications and standardization works are listed below.

#### Conference proceedings

- [1] T. Lavric, E. Bricard, M. Preda and T. Zaharia (2020), An AR Work Instructions Authoring Tool for Human-Operated Industrial Assembly Lines, *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2020, pp. 174-183, doi: 10.1109/AIVR50618.2020.00037
- [2] T. Lavric, E. Bricard, M. Preda and T. Zaharia (2021), Exploring Low-Cost Visual Assets for Conveying Assembly Instructions in AR, *International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2021, pp. 1-6, doi: 10.1109/INISTA52262.2021.9548570
- [3] T. Lavric, E. Bricard, M. Preda and T. Zaharia (2021), An Industry-Adapted AR Training Method for Manual Assembly Operations, *HCI International (HCII), 2021 - Late Breaking Papers: Multimodality, eXtended Reality and Artificial Intelligence*, vol. 41, LNCS 13095, July 2021, Cham: Springer International Publishing, 2021, doi: 10.1007/978-3-030-90963-5\_22
- [4] T. Lavric, E. Bricard, M. Preda and T. Zaharia (2022), ATOFIS, an AR Training System for Manual Assembly: A Full Comparative Evaluation against Guides, *21st IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, doi: n/a

#### Journal articles

- [1] T. Lavric, E. Bricard, M. Preda and T. Zaharia (2022), A Low-Cost AR Training System for Manual Assembly Operations, *Computer Science and Information Systems (ComSIS)*, 2022, vol. 20, issue 1, pp. 1047–1073, doi: 10.2298/CSIS211123013L.

#### Standardization contributions

- [1] ETSI GR Augmented Reality Framework (ARF) 002 V1.1.1 (2019-07), Industrial use cases for AR applications and services,  
[https://www.etsi.org/deliver/etsi\\_gr/ARF/001\\_099/002/01.01.01\\_60/gr\\_ARF002v010101p.pdf](https://www.etsi.org/deliver/etsi_gr/ARF/001_099/002/01.01.01_60/gr_ARF002v010101p.pdf)
- [2] ETSI GR Augmented Reality Framework (ARF) 001 V1.1.1 (2019-04), AR standards landscape,  
[https://www.etsi.org/deliver/etsi\\_gr/ARF/001\\_099/001/01.01.01\\_60/gr\\_arf001v010101p.pdf](https://www.etsi.org/deliver/etsi_gr/ARF/001_099/001/01.01.01_60/gr_arf001v010101p.pdf)
- [3] ETSI GS Augmented Reality Framework (ARF) 003 V1.1.1 (2020-03), AR framework architecture,  
[https://www.etsi.org/deliver/etsi\\_gs/ARF/001\\_099/003/01.01.01\\_60/gs\\_arf003v010101p.pdf](https://www.etsi.org/deliver/etsi_gs/ARF/001_099/003/01.01.01_60/gs_arf003v010101p.pdf)

---

## Chapter 2. INDUSTRIAL CASE STUDY

---

*“One second is too much!”*

---

---

This chapter describes the industrial case study considered in this thesis, carried out at elm.leblanc, a boiler-manufacturing factory. The main objective was to conduct an in-depth exploration of a real-world industrial assembly environment, to identify the most relevant needs and challenges imposed by similar manufacturing sectors, and finally, elaborate an AR expert knowledge sharing system optimally adapted to similar manual assembly use-cases. Informal meetings and open discussions, interviews, demonstrations, and evaluations conducted with experts from the considered factory over a period of approximately 3 years, allowed us to adopt a formative research approach, test different hypothesis and approaches, and finally identify the most relevant aspects that played a critical role in the design and implementation of the proposed AR training system. By conducting this industrial case study, we aimed as well at addressing a recommendation expressed in multiple research works (e.g., (Dey et al., 2018), (Merino et al., 2020), (Masood & Egger, 2019), (Egger & Masood, 2020)), which suggests that in-depth field studies need to be conducted to ensure the successful implementation of AR systems in industry.

Part of the work presented in this chapter has been published in (Lavric et al., 2020) and (Lavric et al., 2021b).

---

---



## 2.1 Introduction and objectives

The case study presented in this chapter was conducted over a period of 3 years at elm.leblanc, a French manual assembly boiler-manufacturing factory. The main objective of this work was to:

- explore and analyze a real-world industrial assembly environment,
- identify and define specific requirements and challenges imposed by similar manufacturing sectors,
- and finally, elaborate an optimal AR training tool for the considered assembly use-case.

Within this context, designing and evaluating the AR training proposal by directly involving the experts of the factory was a top priority. We consider this industrial work essential, as it allowed us to identify concrete needs, expectations and concerns, topics around which we could design and evaluate different AR authoring and training approaches. We collected data from a series of meetings, interviews, contextual inquiries, and demonstrations with more than 30 experts from seven departments of the factory, as summarized in Table 2-1. A key contribution of this work is represented by the evaluation of commercial AR training solutions, as well as the proposed approaches, in a concrete real-world use-case, with the implication of the potential end-users.

**Table 2-1. Departments and people involved in the considered industrial case study at elm.leblanc.**

Department	Positions
EHS	Manager
Production	Manager, assembly line managers and assistances, workers
Planning	Manager, team leader, engineers
Maintenance	Manager
Quality	Project manager
Logistics	Team leader
IT	Chief information officer, team leaders, project managers, designers

We observe in literature surveys that research work tends to evaluate people from the same area of expertise, typically people that can decide at the management level on the adoption of novel technologies and tools (Masood & Egger, 2019). However, to ensure the success of an AR training system in the industrial context, the implication of people from different areas of expertise impacted by the adoption of such a solution, is mandatory during its design and implementation. This is the reason why, during the case study we involved personnel from shop floor experts to IT engineers and top management leaders. As identified in the early stages of this case study, *on the shop floor, any novel tool must satisfy a set of conditions imposed by different departments before its adoption* (CIO of elm.leblanc). We aimed at identifying and understanding these challenges in detail and use them as guidelines in the elaboration of the proposed AR training system.

The work of this case study was conducted on two main axes, detailed in the following sections:

- exploration of the industrial context (Section 2.2)
- prototyping and preliminary studies (Section 2.3)

A summary and classification of the main challenges and requirements identified during this exploration is presented in Section 2.2.5.

## 2.2 Assembly context exploration at elm.leblanc

***One second is too much!*** One of the most remarkable findings identified early at elm.leblanc, concerns the efficiency of the manual assembly process, conducted by workers at their assembly workstations. During the elaboration of the first prototypes of this thesis (Section 2.3.3), we have found that any shop floor system should be designed by carefully considering time. Requiring the worker to perform an action that is not imperative during the product assembly, should be eliminated even if this would only take one second. As reasoned by the CIO of elm.leblanc, one second spent without a good reason can translate to a notable negative impact on the production year's efficiency. For instance, in a factory producing 100.000 goods per year, one second would translate to approximately 4 working days (7h/day) of wasted production time over the entire production year. A second could lead therefore to an astounding production decrease of 0.47%, considering a 3'30'' nominal assembly cycle.

Further, in this section, we explore the manual assembly context, focusing on the one at elm.leblanc.

Figure 2-1 presents a view of the overall assembly line L2 at elm.leblanc, the one where most of the experiments of this thesis have been conducted.



***Figure 2-1. An overview of the assembly line L2 at elm.leblanc where most of the experiments conducted during this thesis have been conducted.***

### 2.2.1 Research problem validation and refining

Informal meetings and discussions with shop floor experts support the findings identified in the literature, briefly presented in the introduction of this thesis, regarding the main challenges of manufacturing sectors that rely on human workforce. In addition to the complexity and the diversity of the products, we found that the rising unpredictability of the client demand seems to be a major concern in the human-operated manufacturing factories. Reliable predictions

concerning the production output usually cannot be made due to changing requirements and specifications (Brecher et al., 2013). The demand of products is driven primarily by the clients, who are influenced by economic and political factors, marketing, and competition. This is the reason why an accurate prediction of the future demand of specific products is most of the time hard to do, as claims the production manager of elm.leblanc. In addition, the key challenge is to be able to produce the demanded supply and to make it as fast as possible to satisfy customer expectations. One could naively believe that producing in advance could be a simple solution for addressing this concern. Such an approach however comes with multiple disadvantages including the risk of not selling the stock. In this case, storage, depending on the type of the product, can be problematic and costly. Starting to produce as soon as the customer demand is received is the most common strategy in efficiently overcoming this challenge, in the context of factories with high manufacturing diversity. Therefore, the need of hiring seasonal workers to deal with high demand fluctuations becomes a necessity. Statistical data shows that, at elm.leblanc, seasonal workers represent up to 75% of the workforce during the high-demand seasons. The problem with seasonal workers is that generally, they are not trained for the manufacturing process that they will have to perform. To become autonomous and effective, novice workers require individual training procedures, assisted by line managers or other shop floor experts. However, in the context of Industry 4.0, traditional training methods present multiple limitations (Billett, 1995).

### 2.2.2 Human workforce and industrial automation

In this section, we exemplify why human workers are still an irreplaceable resource for specific manufacturing sectors and why full automation is not conceivable yet.

(Cohen et al., 2019) identifies that the role of people in Assembly 4.0 (i.e., assembly systems transformed by Industry 4.0 technology integration) will be increasingly important, calling for future research that considers human workers back at the center of the production systems of the future. The 4<sup>th</sup> industrial revolution is expected to transform the role of the people, but to what extent will assembly workers be affected—are humans to be replaced by machines or empowered by new technology (Miqueo et al., 2020)? Literature findings are somehow mixed regarding this research question.

In manual assembly, the human is the main actor. Discussions with shop floor experts from the factory suggested that in the context of human-operated assembly lines, everything from the conceptual stage of the product to the material feeding and line balancing are key aspects that have direct impact on the productivity. Nevertheless, the uttermost important one is the actual assembly process, which is fully or partially, directly, or indirectly performed or conducted by human workforce. Literature shows as well that the key factors in manual assembly are personnel and time (Warnecke & Hüser, 1995), (Gräßler et al., 2021) and that people still need to be considered central to Assembly 4.0 (Miqueo et al., 2020).

Further, the performance of the assembly system is not only influenced by the design of the system, but also by the characteristics of the worker who performs at workplaces (Gräßler et al., 2021). Human factors can be understood as all characteristics of humans that influence their

actions within socio-technical systems and can be subdivided into physical aspects (e.g., posture), mental aspects (e.g., skills) and psychosocial aspects (e.g., motivation) (Neumann et al., 2016). A pertinent proposal for removing the necessity of dealing with human factors and training procedures would be not relying on human workforce at all, by fully automating existing human-operated assembly lines.

A complete automation of the assembly, by using machinery like robotic arms, would remove the need of human intervention. By replacing human workforce, every aspect directly linked to them such as hiring process, training procedures, unexpected events (e.g., absenteeism), manufacturing errors and others would be eliminated as well. This idea persuades one to believe that full assembly line automation would be entirely beneficial for the manufacturing process. This is partially true and can be applied to a restricted set of use cases, like the automotive one. To establish whether the human worker is a replaceable resource in manufacturing factories, we must consider a key element that is the *workplace flexibility*, defined as the ability of workers to make choices influencing when, where, and for how long they engage in work-related tasks (Hill et al., 2010). Literature shows that flexibility is mainly required to deal with two variables: intrinsic uncertainties and the variability of outputs (Corrêa, 1994). Despite technological advancements of the last decades, the human factor is suggested to be the essential flexibility component in manufacturing processes (Kara, 2004). Assembly systems continue to require manual labor because of specific human qualities, such as the ability to cognition and problem solving, a capability that cannot be yet effectively replaced by technological tools (Wyman, 2017). (Feldmann & Slama, 2001) showed as well that workers in general are considered as major flexibility enablers since they are capable of quickly adapting to changing products and market situations. Manual assembly tasks often require dexterity and sensitivity that cannot always be automated economically (Bughin et al., 2017).

It seems therefore that in Industry 4.0, human workforce will typically remain the most flexible part of production systems, suggesting that systems are not being designed to eliminate, but to support humans with new technologies that could enhance their capabilities and compensate for any limitations (Hermann et al., 2016), (Romero et al., 2016), (Weyer et al., 2015). This is the reason why human workforce remains an undeniable requirement and, for the moment, an irreplaceable resource in manufacturing sectors, particularly for the ones with a high degree of product diversity and complexity. It is therefore imperative to make efforts in providing methods for helping workers learn and adapt as fast as possible to new manufacturing conditions and processes while maintaining the quality of work.

### **2.2.3 Training in manufacturing: importance and limitations**

To become a reliable resource of the production process, a worker must first acquire the required manufacturing skills. Obviously, this is where training comes into play. During training, a worker learns the assembly sequence, the usage of tools and machinery, to perform the assembly operations safely, correctly, and efficiently. In other words, a training procedure empowers a human worker to become autonomous and effective at his workplace. Competent assembly workers commit fewer errors and can keep up with the production requirements,

including time and quality. Moreover, they can detect assembly errors committed by others, preventing cumulative or cascading failures.

So, what is wrong with the traditional training procedures? Due to high staff turnover, demographic change and an increasing number of products and variants, companies need solutions to train their employees in a standardized and efficient way (Geissbauer et al., 2017), which is not particularly the case (Mital et al., 1999). Today, the most common and efficient way to acquire assembly skills is by being trained by an expert in the domain (Oestreich et al., 2019). In the considered context at elm.leblanc, the persons in charge of training are either line managers or experienced workers with a very good knowledge of the workstation in question. The instructor trains a novice worker at a time by performing and explaining the assembly procedure, systematically. The training procedure is performed directly on the assembly line, in real-time, while the line is running. The reasons for adopting a “live” training technique are the following:

- The assembly lines are most of the time running, therefore the availability of the line for other purposes but production is extremely limited,
- Except for the first workstation, the assembly procedure of all other workstations requires a complete assembly of all the previous workstations; therefore, a running assembly line is providing the partially assembled product needed for each next workstation,
- A dedicated environment for training purposes is not conceivable, as this would eventually require a clone of the entire factory. Providing generic workstations is not a viable option either, considering the diversity and the high number of assembly operations of an entire assembly line.

To get a better understanding of the traditional training procedure employed in the considered context, we designed a questionnaire for the line managers of the factory. Table 2-2 presents this questionnaire and the corresponding answers.

**Table 2-2. Questionnaire for line managers et elm.leblanc, designed by us, to identify details regarding the traditional training procedures conducted in the factory.**

Q1: How many trainings are available in the factory?	A: One formation for each workstation (generally 12 workstations / assembly line).
Q2: How many workers are trained per month?	A: It varies. Difficult to anticipate, depends on the season and client demand.
Q3: Who performs the training?	A: Generally, team leaders or experienced workers.
Q4: What are the requirements to perform a training?	A: Workstation experience and trust.
Q5: How much a training session takes?	A: Depending on the workstation, anywhere from 2 to 15 days.
Q6: Which are the main challenges during a training session.	A: Motivation, assimilation capacity, forgetfulness, skills.
Q7: How important is the training from your point of view?	A: Extremely important. Training is the very first thing that a novice worker does after he is hired.
Q8: How many seasonal workers are hired?	A: Up to 75% in high demand periods.

---

Q9. When is the training performed?

A: On demand, we cannot anticipate. High demand periods are not defined by specific dates. They are determined by the outside temperature, which makes it very difficult for us to establish a concrete hiring plan. Sometimes a training decision is taken in the morning for the afternoon.

---

Q10. How is the training performed?

A: It starts directly in front of the workstation. The instructor performs the assembly and explains. The trainee takes over gradually, under supervision. The instructor can have another workstation in charge at the same time.

---

Q11. Is there any digital content to support the training procedure?

A: Yes. Each workstation has paper instructions. However, these are not used for training purposes but rather by experienced workers when a product model change during the shift or when switching to a workstation the worker did not operate in a long time.

---

This inquiry suggests that the most convenient way of performing the training procedure is by assisting the novice worker directly on the running assembly line. Further, we describe the steps of the traditional training procedure, as observed at elm.leblanc. This is relevant to this work, as proposing a novel AR training solution should address the expectations and performance of the traditional one.

1. The instructor and the novice worker are present at the workstation, equipped with security shoes and gloves.
2. First, the instructor performs several assembly cycles while giving explanations; the novice worker observes.
3. After some cycles, the worker is asked to perform the assembly operations that he remembers while the instructor is closely assisting and giving feedback.
4. As the tempo of the line is given by the slowest workstation, during the whole training procedure, the instructor tries to keep up with the nominal cycle time by interfering with the assembly process as often as necessary.
5. The novice worker learns the assembly sequence and becomes autonomous, but not effective from a time perspective. The time required for learning the sequence is dependent on the assembly complexity and on the worker's skills.
6. At this point the instructor is not required anymore, however the worker is still considered novice until he performs the assembly within the nominal time.

Depending on the workstation complexity, on the assembly experience and on the cognitive skills, the training procedure can last from minutes to hours and days.

One of the most important limitations of this training approach is the availability of an instructor. It is important to note that the production department can request for new workforce from one day to another. As the instructor is generally an individual with other responsibilities, conducting a training session can affect the production process. Another potential concern, less quantifiable but more sensible, identified during discussions with experienced workers, can arise in the way the instructor conducts the training. This procedure is influenced by the interaction between the two participants, by the quality of the explanations but also by the interest and the cognitive skills of the worker. In addition, the explanations given by the instructor are not standard and are subjective.

Finally, another relevant concern regarding traditional training methods is represented by product diversification and change. The effectiveness of classical training procedures is diminished by technological advancements, which allow an increasing diversity of the products and leads to more complex assembly process. Product upgrades or new products imply modifying existing assembly operations or defining new ones, affecting the existing assembly procedures associated to them and implicitly the training. It is mandatory for the training responsible to keep up with these changes as soon as they occur.

Transferring expert knowledge in the traditional way presents a set of limitations that makes existing training techniques unsuitable for the dynamic industrial context of today. The most relevant concerns are the following:

- It is not reusable, portable, and scalable as a human instructor is required,
- It is subjective and its effectiveness rely on technical and social skills of the instructor,
- It is not standardized which leads to inconsistency and various training performance.

#### 2.2.4 Assembly organization, process, and resources

In this section we provide an analysis of the typical organization of an assembly line and its components, as this information fundamentally impacts the design of the proposed AR system.

##### 2.2.4.1 Assembly line analysis

###### 2.2.4.1.1 Organization and function

Manual assembly lines are often organized as flow production and represent the workplace where the manufacturing process of the product, the boiler in our specific use-case, is performed (Figure 2-1). In general, an assembly line consists of several consecutive workstations, at which a worker performs repetitively, the same series of assembly operations. At elm.leblanc, most assembly lines consist of 12 workstations. Figure 2-2 illustrates the first workstation (i.e., W201) of the line L2 at elm.leblanc and its main components.

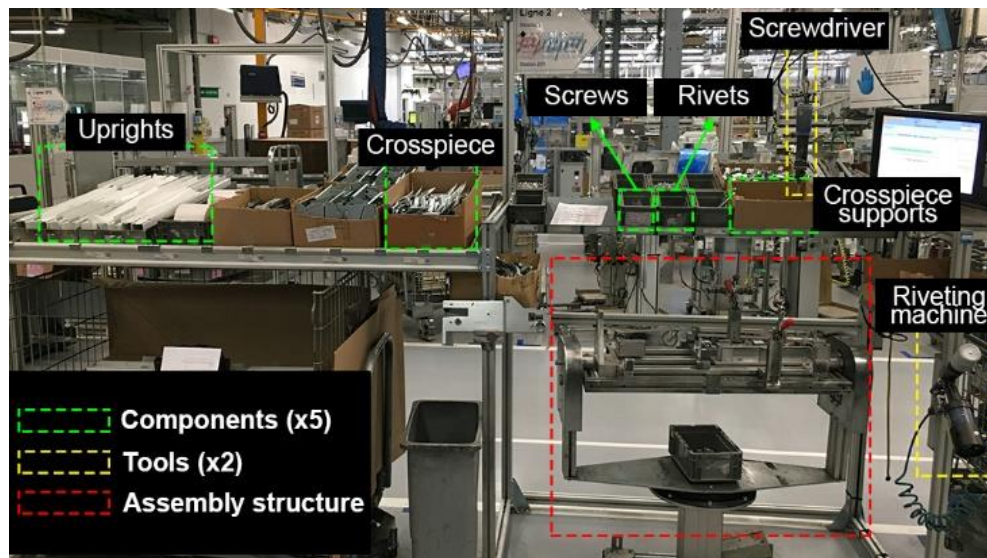


Figure 2-2. The first workstation (i.e., W201) of the assembly line L2 at elm.leblanc. The tools and assembly components are highlighted and labeled.

Any assembly line is dependent on all its workstations, meaning that a line cannot function unless all its workstations are operating. The workstations are positioned in a U-shape. The manufacturing process is sequential, it starts at one end of the U-shape, and it ends at the other one. The assembly process is performed gradually as follows: at each workstation, new assembly components are added sequentially, then, the product, which is partially assembled on a specific cart, moves to the next workstation where the next set of assembly operations is performed. At the end of the line, the product is completely assembled, tested, packaged and ready to be shipped. It is important to note that the assembly operations performed at a given workstation depend on the ones performed at the previous one. Thus, performing a training procedure at workstation N requires the partially assembled product to be in the assembly state N-1. The production tempo is given by the slowest workstation of the line; therefore, the assembly work is balanced between the workstations so that their nominal assembly completion time is very similar. Consequently, a training procedure that takes place at any workstation of a running line inevitably affects the production speed of the entire line, unless an experienced worker interferes during the training.

A realistic objective of an AR training method, from a time perspective, is therefore to minimize the time needed by a novice worker to read the AR instructions, while performing the assembly. Broadly speaking, this means that the conveyance of the AR instructions must interfere as less as possible with the actual assembly process to allow the worker to perform the assembly operations as fast as possible and keep up with the nominal assembly time of the workstation. We must note however, that other industrial sectors have less restrictive requirements, therefore the training sessions may be performed under different circumstances.

A specific scenario arises when the product demand is low, and only half of the usual workforce is required to satisfy it. In a half-team shift each worker oversees two workstations, meaning that the number of workers is half the number of workstations of the line. This information is pertinent because it represents a potential scenario in which a novice worker would have twice the nominal cycle time to perform a full assembly of a workstation during training.

#### 2.2.4.1.2 Product models, assembly components and operations

Assembly line updates and introduction of new product models happen quite often in the factory. As an example, in the year 2021, a new assembly line was created and configured to support the manufacturing of around 30 new product models. The line has around 260 assembly components from which 160 are new. The assembly components are very diverse in size, shape, and weight, from very small components like screws, clamps, cables to medium like crossbars and uprights, and big like boiler balloons (e.g., Figure 2-6 and Figure 2-7). Small and medium components are generally manipulated by hands, while big ones, including the boiler itself, require specific machinery.

To assess objectively the diversity of the assembly operations, we have analyzed and categorized them by counting the number of verbs found in the existing paper instructions of one of the assembly lines. We have identified 374 actions (i.e., verbs) in total, among which 55 are unique. We found that actions asking to “*install*”, “*place*”, “*put*” and “*position*” are interchangeable

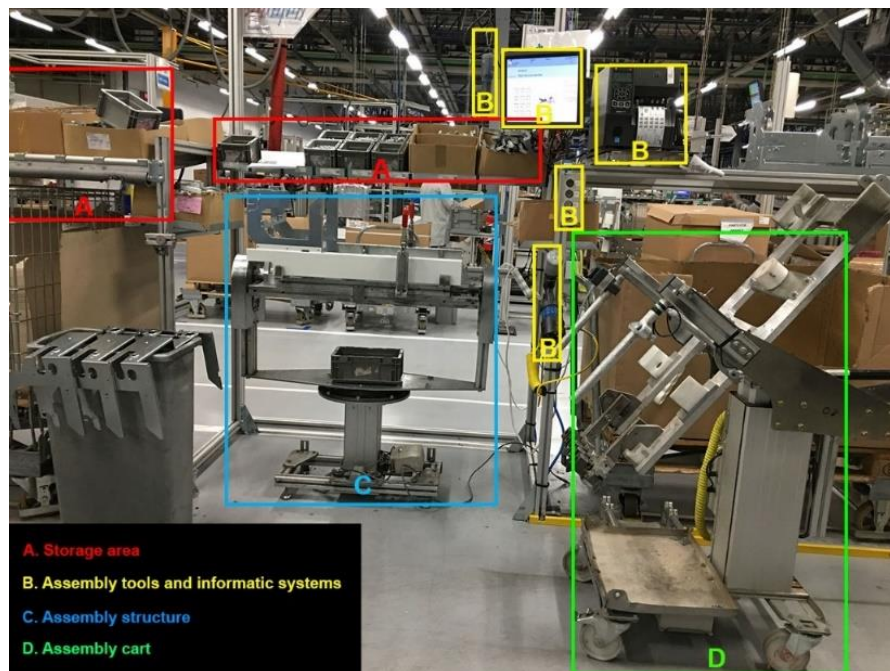


verbs but not similar assembly operations and represent 36.6% from the total. Next most common actions are “screw” with 9.8%, “connect” or “hook” with 7.7%, “stick” with 3.4% and “rivet” with 3.4%. Other actions, each representing less than 3% of the total are “evacuate”, “scan”, “push”, “clip”, “action”, “rotate”, “turn”, “engage”, “prepare”, etc. We note that only actions like screwing, riveting, scanning/flushing, and pressing (a button) are actions that can potentially be represented in a generic way to be reusable independently on the workstation. The remaining actions are workstation and assembly component-specific and represent 82% from the total. For this very high percentage of task-specific assembly operations the literature (Section 2.2.4.3.2), does not provide a standardized way of visual representation in AR. Similarly, there is no general solution for visual information conveyance, adapted to specific assembly components and operations (Gattullo, Evangelista, et al., 2020). We note that one of the main goals of this thesis is to address this research gap by proposing generic and consistent authoring and information representation processes, independent on assembly particularities.

#### 2.2.4.2 Workstation analysis

Another relevant aspect of assembly factories is the workstation organization. This represents the environment where the assembly process is performed by the worker. Workstations are generally composed of four main components, as illustrated in Figure 2-3.

1. A storage area (Figure 2-3.A), where the assembly pieces are stored.
2. Tools and informatics systems (Figure 2-3.B), used by the worker to perform specific assembly operations like screwing, riveting, scanning, and testing.
3. Assembly structures (Figure 2-3.C), where parts of the final product are assembled.
4. An assembly cart (Figure 2-3.D), on which the product is assembled and sent to the next workstation.



**Figure 2-3. Common workstation components and organization: A (red) – storage areas, B (yellow) – assembly tools and informatic systems, C (blue) – assembly structure and D (green) – assembly cart.**

Let us analyze how the workstation components and their organization can impact the design of an AR training system.

#### 2.2.4.2.1 Storage area

The storage (or picking) area is a fixed frame, which stores the boxes with the assembly components of the workstation. These boxes are organized in such a way to minimize the worker's movements and effort. Their position is fixed and only some of them are needed, depending on the product model that is manufactured. During the training procedure, an AR system is supposed to indicate the position of these boxes containing the required assembly components, sequentially, for each picking operation. This technique is also known in the literature as pick-by-AR and can be employed by using visual cues (e.g., spatially registered arrows or virtual tunnels) (Hanson et al., 2017) or SAR-based systems (Funk, Shirazi, et al., 2015), (Battini et al., 2015). The position of the boxes can be indicated during the creation of the AR work instructions or detected, by their label, in real time, using image recognition or other sensor-based techniques. However, the complexity and reliability of image/sensor-based techniques do not meet yet industrial requirements. Even though the storage boxes are generally labelled with reference numbers and fiducial markers, this information is not reliable for image detection purposes because of the small dimension of the labels, the uncontrolled lighting conditions of the environment and various other reasons like label deterioration, occlusion, or even missing label.

Another technique used for identifying the position of an assembly component in real-time is to recognize the assembly component itself. A reliable, real-time assembly component recognition system would be optimal, as this would also address assembly environments that change between the authoring of the AR instructions and the training procedures. However, a real-time image processing-based solution is not adapted to the considered use-case (as detailed in Section 2.3). Thus, the position of some storage boxes is, in some cases, too high for a head-mounted camera (embedded in state-of-the-art AR devices). The height of the worker would also play an important role in such approach. An alternative would be a fixed camera-based system specifically dedicated to component recognition (and potentially other recognition tasks such as identifying and validating workers' operations) during AR training, and eventually authoring. Such approach was not considered viable because of concerns including initial and running costs, system complexity, maintenance, and calibration. In addition, the reliability and efficiency of existing machine vision algorithms for object detection purposes in AR (Huang et al., 2019), (G. Liu et al., 2019), (Paramarthalingam, 2016), (Nishihara & Okamoto, 2015), (Lai et al., 2020), (L. C. Wu et al., 2016) are not yet proven in industrial conditions. It is imperative for the viability of an AR training system to function efficiently without the need of recurrent set-ups, calibrations, and other manipulations of the assembly. We note that in the considered use-case, the picking areas are static, suggesting that hardcoding boxes' positions during the authoring procedure could represent the simplest and most viable resolution to this concern.

#### 2.2.4.2.2 Assembly tools

Most workstations consist of various assembly tools (e.g., screwdrivers, riveting machine, pliers), screens and other informatic components, employed by the workers during the assembly procedure. Such components are stored at fixed locations generally. However, some exceptions exist. An example is represented by vertical screwdrivers, suspended by a cord or a metallic arm. After usage, once released, they can end up in a slightly different position than the initial one. We note that such misalignments are insignificant in the considered assembly environment, therefore we can assume that their position is static as well.

#### 2.2.4.2.3 Assembly structures

The assembly structures are specific areas of workstations where parts of the boiler are assembled individually, prior to being assembled as part of the boiler, on the mobile assembly cart. These structures are generally static, but some can be rotated, as in the case of the workstation on which we evaluated the proposed AR training system implementation (i.e., W201 - Figure 2-3.C). This is relevant for the cases where a fiducial-based, or any other context-unaware, spatial registration of the virtual elements is considered. Such real-world registration on a mobile structure will be broken as soon as the initial position of the assembly structure is changed, unless a real-time calibration is ensured. The lack of reliable, real-time object tracking systems is a well-known concern (L. Liu et al., 2019), notably in difficult environmental conditions. The existing solutions proposed are generally not evaluated in industrial contexts and seem to provide acceptable results only under certain evaluation conditions and laboratory settings (Popelka et al., 2012), (Wagner et al., 2010), (Mooser et al., 2007), (Uchiyama et al., 2012), (Ahmadyan et al., 2020), (Y. Wang et al., 2018), (J. Wu et al., 2018).

#### 2.2.4.2.4 Assembly cart

The assembly cart is a mobile structure on which the boiler is finally assembled. The cart moves from one workstation to the next, on a straight line, guided by a rail. The final position of the cart at a given workstation varies, depending on the worker preferences. We note as well that the worker can also manipulate the height and the rotation (on one axis) of the cart to fit his height and style of work. Unlike the assembly structures therefore, for the assembly carts is not only the rotation which can change but also the ground position. However, relying on superimposed-based AR requires a real-time, precise tracking of the cart during the entire assembly procedure, which represents an important technical concern.

#### 2.2.4.2.5 Conclusions

The assembly workstation is a complex, dynamic environment, with a very diverse set of components and operations. More generally, by considering the manufacturing industry at large and its multitude of domains and products, we argue that a tremendous number of components, tools and manual assembly operations exist. Consequently, the assembly experience acquired elsewhere than the specific workstation on which a novice worker will work, would most probably be imperceptible. This not only highlights the importance of the training, but also suggests that generally, a worker, independently on his previous assembly experience, can be considered novice if he never operated the considered assembly workstation or a very similar one before. This claim is supported by the feedback of our informal experiments conducted in

the factory. One such example is represented by an experienced worker which did not know how to use a riveting machine used by his colleague from an adjacent workstation hundreds of times per day.

Regarding the organization of the workstation in general, we identified that most components are fixed, workstations have generally very few things in common (e.g., screwdrivers), and the storage boxes labeling is not reliable. Even though the overall organization of the assembly lines and workstations is exemplary from a shop floor perspective, the assembly context does not seem to be *AR-friendly*, particularly for an AR system that would need to rely on real-time visual processing and precise spatial registration of the virtual elements.

### 2.2.4.3 Assembly process analysis

#### 2.2.4.3.1 Assembly operations

Concerning the assembly process in similar manufacturing sectors, market demand requires product customization, which leads to frequent updates and new product models. To objectively assess the diversity of the assembly process, we have conducted an analysis of the verbs found in all the existing paper instructions (i.e., all workstations) of one of the assembly lines, L2, illustrated in Figure 2-1. The objective was to assess whether assembly operations could potentially be digitally captured and represented in a specific manner, based on assembly characteristics, trying to standardize their description in AR and optimize both the AR authoring and training procedures.

Table 2-3 presents the list of all the verbs and their number of occurrences identified in the considered paper instructions (12 in total).

**Table 2-3. Manual assembly operations (i.e., verbs) and their occurrences identified in the existing paper instructions of the assembly line L2.**

<i>Fit</i>	95	<i>Grease</i>	8	<i>See</i>	4	<i>Remove</i>	2	<i>Steer</i>	1	<i>Mount</i>	1
<i>Screw</i>	37	<i>Clip</i>	8	<i>Operate</i>	4	<i>Prepare</i>	2	<i>Format</i>	1	<i>Descend</i>	1
<i>Position</i>	20	<i>Bring</i>	7	<i>Unplug</i>	4	<i>Engage</i>	2	<i>Advance</i>	1	<i>Strap</i>	1
<i>Put</i>	20	<i>Tighten</i>	7	<i>Turn</i>	3	<i>Ensure</i>	2	<i>Put into</i>	1	<i>Print</i>	1
<i>Stick</i>	13	<i>Place</i>	6	<i>Disconnect</i>	3	<i>Unscrew</i>	2	<i>Take off</i>	1		
<i>Rivet</i>	13	<i>Press</i>	5	<i>Check</i>	3	<i>Take out</i>	2	<i>Watch</i>	1		
<i>Connect</i>	18	<i>Take</i>	5	<i>Unlock</i>	3	<i>Lower</i>	1	<i>Follow</i>	1		
<i>Plug</i>	11	<i>Return</i>	5	<i>Control</i>	3	<i>Do</i>	1	<i>Adjust</i>	1		
<i>Evacuate</i>	10	<i>Lock</i>	5	<i>Close</i>	3	<i>Flip</i>	1	<i>Open</i>	1		
<i>Scan</i>	10	<i>Drop off</i>	5	<i>Search</i>	2	<i>Assemble</i>	1	<i>Fill</i>	1		

We identified 54 unique verbs and 370 occurrences. We note that some of the verbs are similar and potentially can be interchanged. Some examples are [*fit*, *position*, and *put*], [*check* and *control*] and [*remove* and *take off*]. Repetitive and non-value verbs in assembly instructions are identified in the literature as well (Miller et al., 2012).

To evaluate the comprehensiveness and how generic these verbs are, we combined those which are similar and compare the resulted verb set with one identified in the automotive industry (Miller et al., 2012), presented in Table 2-4.

**Table 2-4. Manual assembly operations identified in an automotive use case.**

<i>Align*</i>	<i>Connect*</i>	<i>Hand Start</i>	<i>Position*</i>	<i>Re-open</i>	<i>Snap</i>	<i>Walk*</i>
<i>Apply</i>	<i>Disconnect*</i>	<i>Insert</i>	<i>Prep*</i>	<i>Restock</i>	<i>Start</i>	<i>Write</i>
<i>Assemble*</i>	<i>Ensure*</i>	<i>Install*</i>	<i>Press*</i>	<i>Retrieve</i>	<i>Take*</i>	
<i>Attach*</i>	<i>Fasten</i>	<i>Loosen</i>	<i>Pull</i>	<i>Return*</i>	<i>Tear off</i>	
<i>Bring*</i>	<i>Feed</i>	<i>Maneuver*</i>	<i>Push*</i>	<i>Rotate*</i>	<i>Tighten*</i>	
<i>Check*</i>	<i>Fix*</i>	<i>Move*</i>	<i>Put*</i>	<i>Scan*</i>	<i>Tuck</i>	
<i>Clean</i>	<i>Flip*</i>	<i>Open*</i>	<i>Read</i>	<i>Seat</i>	<i>Turn*</i>	
<i>Close*</i>	<i>Get*</i>	<i>Operate *</i>	<i>Receive</i>	<i>Secure</i>	<i>Unplug*</i>	
<i>Compare</i>	<i>Guide*</i>	<i>Pick*</i>	<i>Release*</i>	<i>Set*</i>	<i>Use*</i>	
<i>Confirm*</i>	<i>Hand</i>	<i>Place*</i>	<i>Remove*</i>	<i>Slide</i>	<i>Verify*</i>	

Common verbs between the verb sets presented in Table 2-3 and Table 2-4 are marked by (\*).

We found that 78% of the verb set extracted from the analyzed paper instructions is common with the verb set found in the literature, presented in Table 2-4. The ones that were not found in the assembly operations from the automotive example are *clip*, *descend*, *fill*, *grease*, *print*, *rivet*, *stick*, *lock/unlock* and *screw/unscrew*. This comparison shows that assembly operations are described by using a rather wide range of verbs and that a high percentage of them are common among different industrial sectors. However, considering that the creation of the manual assembly instructions does not follow a standard procedure from a vocabulary perspective, we assume that is virtually impossible to provide an exhaustive list of verbs used for describing specific manual assembly operations.

To address the vocabulary concern, in a similar use case, (Scurati et al., 2018) proposed a methodology for converting maintenance actions into standard symbols for AR applications. They identified 21 verbs covering approximately 84% of the maintenance operations found in the analyzed instruction manual and classified them into three categories, based on their complexity:

- High level verbs: move, disassemble, assemble, install, remove, replace, hold,
- Low level verbs with agreement rate > 0.2: examine, fill, clean, cut,
- Low level verbs with agreement rate < 0.2: tighten, loosen, lift, push, pull, turn clockwise, drain, turn counterclockwise, close, open.

We found that only approximately 13% of the verb set identified in our case study overlaps with the one identified in the considered research work, suggesting that a very low percentage of the verbs used to describe the industrial operations are common across different industrial processes. The value reported from the previous comparison (i.e., 78%) indicates that assembly environments have a much higher verb set commonality even among different industrial sectors (e.g., boilers vs. automotive).

Nevertheless, by considering the diversity of the manufacturing operations, components, machinery, and tools identified in our case study, we believe that trying to provide a standardized way of capturing and conveying operation-specific assembly information in AR, represents a challenge that cannot be fully addressed in this work. The diversity and complexity

of industrial manufacturing potentially explains the lack of dedicated AR standardized methods and guidelines for describing and conveying manual assembly expertise in these sectors This issue is further analyzed in the following section.

#### 2.2.4.3.2 Existing standards and APIs

Manufacturing factories use standards for various reasons, from planning and design to quality management and line optimization. We observe that the existing paper instructions available at elm.leblanc (e.g., Figure 2-4.a) follow a template for describing the assembly operations by combining text description, reference numbers, images, icons, and cues. Let us note that paper instructions of other manual assembly factories (e.g., Adecam Industrie) have a completely different format, suggesting that each factory adopts specific ways of describing assembly operations. This may explain the lack of open standards dedicated to describing manual assembly operations in industrial sectors, at a large scale. Let us further analyze if existing AR standards tackle this issue.

AR standards such as Augmented Reality Application Format (ISO/IEC - ARAF, 2017) and Augmented Reality Markup Language 2.0 (OGC - ARML 2.0, 2012) are XML-based data formats which provide scene description elements for describing AR content, mechanisms to connect to local and remote sensors and actuators, to integrate compressed media (image, audio, video, graphics) and to connect to remote resources. Such standards provide a way of creating AR applications and services in a standardized manner to address AR interoperability concerns. In addition to the standardized data description, they provide scripting functionality to allow dynamic access to the properties of the scene elements, event handling and user interactivity. However, they do not provide methodologies and guidelines for describing, authoring, and conveying information via AR.

The W3C (WebXR, 2018) consortium proposes a group of standards which are used together to support rendering 3D scenes to hardware designed for presenting virtual worlds (VR), or for adding graphical imagery to the real world (AR). WebXR is a JavaScript application programming interface that enables applications to interact with AR and VR devices in a web browser. As for ARAF and ARML, these set of standards address and implement AR concepts and functionalities like spatial tracking and anchors, rendering, viewpoints, interaction, and sensors. However, no specifications regarding content authoring or information presentation in AR are provided.

An in-depth analysis of existing AR-related standards, addressing both software and hardware technologies, including the ones presented in this section, are discussed in the ETSI ARF group, in an AR standards landscape review (Arf, 2019). We observe significant work being conducted for standardizing AR system architecture, data representation, communication protocols and user interaction as well as domain-specific standards with the main objective of providing interoperability between AR systems and components, services, and application interfaces. On the other hand, AR content authoring, information description and visualization seem to be topics out of the scope of the ETSI ARF standardization group.

In addition to AR-related standards, we have also investigated the so-called methods-time measurement (MTM (METHODS-TIME MEASUREMENT)). MTM is an industrial standard that is not specifically dedicated to AR but is widely used in various manufacturing sectors. MTM analyzes the methods used to perform any manual operation or task and, as a result yields the standard time in which a worker should complete that task. MTM-UAS, a version of MTM used at elm.leblanc, was specifically developed for small batch production where speed of analysis has priority over methods and operational detail (*MTM-UAS / MTM*, n.d.). The basic operations of this standard consist of 7 activities including *get and place*, *place*, *handle tool*, *operate*, *motion cycles*, *body motions* and *visual control*. These activities make 77 unique codes, used to describe manual assembly operations from a completion time perspective. Unlike the verb set presented in the previous section, we note that these activities have a meaning that is more general, presumably not adapted for describing specific assembly operations in AR. We found during our assembly line experiments (Section 2.3) that the same MTM codes are used to describe assembly operations which are not of the same type and require completely different manipulation gestures and operations. For instance, the code “PB1” of the “get and place” activity is used to describe the installing of a component on the assembly structure as well for indicating the positioning of a screw on the screwdriver. By considering the number of unique verbs (i.e., 54) used for describing assembly operations as identified in the previous section and the number of basic operations (i.e., 7) proposed by the MTM-UAS standard, we assume that MTM-UAS is too ambiguous for providing a formalized and explicit description of complex assembly operations in AR, especially to novice workers. We note that the TMU values of the assembly operations could potentially be used as time delimiters for an automatic capturing of the assembly expertise. However, as identified in the related work (Chapter 3), such authoring approaches are still far from industrial requirements; therefore, the potential utility of such information in the context of authoring AR instructions is for now negligible.

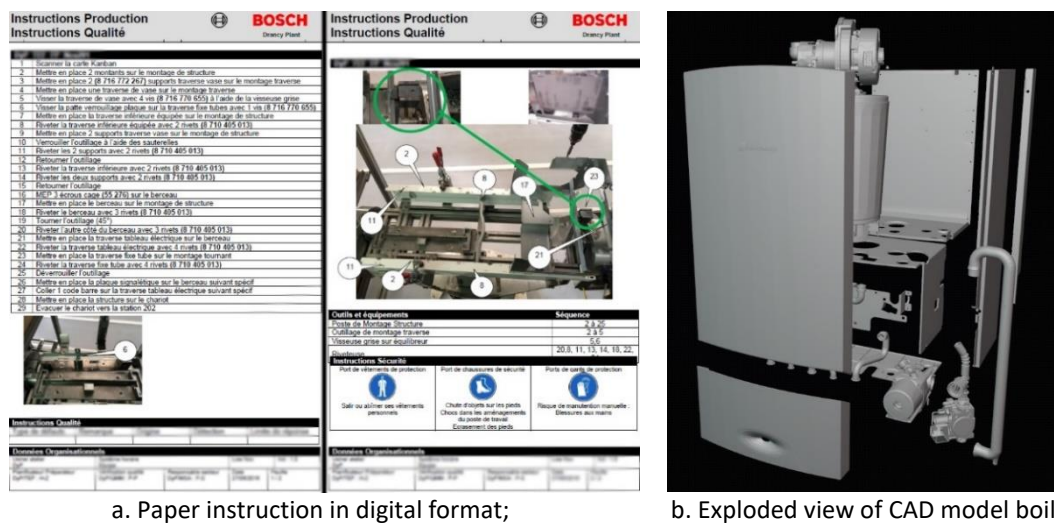
Finally, we identified a standardization attempt proposed in the literature that tries to address industrial manufacturing. The work, conducted in (Scurati et al., 2018), aims at identifying the most frequent maintenance actions used in instruction manuals and convert them into graphical symbols, which then can be used to convey information in AR guidance systems. This approach could remove language barriers and decrease authoring complexity. Three 2D symbol vocabularies have been proposed, based on the *guessability* and homogeneity measured from 192 persons. Vocabulary 1.1, a subset of the first among the three, reported the highest *guessability* and was selected to be used in an AR collaborative maintenance use case. The study does not report on evaluation results (e.g., time, error rate, usability, cognitive load), therefore the usability and suitability of the proposed vocabulary remains difficult to estimate. We observe however that the average confidence score reported by the users for the entire set of symbols in the chosen vocabulary is of 31.7%, which arguably represents a very low value. This indicates that prior to conveying information by using the proposed approach, the participants need first to get familiar with the 2D symbols of the vocabulary. This suggests that even simple and supposedly intuitive 2D symbols require learning, potentially explaining the lack of unique standardized 2D symbol set in manufacturing sectors.

### 2.2.4.4 Existing digital resources

In this section, we discuss existing digital data of the factory that could impact the elaboration of the AR training system.

In the considered use case at elm.leblanc, paper instructions exist for each workstation. However, they are rarely used for training novice workers, but rather provide support to experienced workers for job rotation purposes. This claim is supported by our own inquiry summarized in Table 2-2, and identified in the literature as well (George Michalos et al., 2010). A second digital resource available in the factory and relevant to our work is represented by the 3D CAD models conceived for most of the boilers and their individual components.

Figure 2-4.a illustrates an example of one of the existing paper instructions at elm.leblanc. An exploded view of a CAD model (so-called product model) of a boiler is illustrated in Figure 2-4.b.



a. Paper instruction in digital format;

b. Exploded view of CAD model boiler.

**Figure 2-4. Relevant existing digital data at elm.leblanc.**

The product models consist of detailed CAD parts of each assembly component of the boiler, except wiring and other flexible components. CAD models of tools, machinery, assembly structures and other parts of the workstations do not exist. In addition to storing, file versioning, fetching and manipulating concerns, the existing CAD models are created in a proprietary format (i.e., AutoCAD DXF), not compliant with popular AR tools and libraries. Prior to using them in AR, such models require file conversion and mesh simplification, as discussed in Section 2.3.3. In a general manner, we may state that authoring, storing, fetching, conversion and simplification of CAD models are topics that raise important concerns regarding the acceptability and viability of an AR solution that relies on such complex 3D content. It is preferable that the proposed AR training system is not dependent on but supports the usage of 3D CAD data.

### 2.2.5 Industrial requirements and challenges

Table 2-5 presents a list of the most relevant aspects collected during our case study conducted at elm.leblanc, along with the department where the quotes or comments were obtained from. The information was collected from semi-structured interviews and open discussions, in the context in which an AR training system would replace the existing classical training procedures.



**Table 2-5. Requirements and challenges stated by experts from the factory at elm.leblanc.**

<b>CIO</b>
New technological tools can be deployed on the shop floor only if the two following prerequisites are addressed. First, the solution must be mature from a technological and business perspective: it must be easy and fast to use as well as being efficient. Second: the management must be done carefully with all stakeholders. Teams must understand the benefit first and see concretely the results on the shop floor. If the tools fit their needs and they find value in it, then it is a win.
To be sustainable, the solution must be easy to use and maintain. The less people needed to manage an IT solution, the better. People come and leave. Knowledge transfer is a challenge.
Costs obviously are important. A minimal investment to prove that the system increases the KPI. Budgets for innovation in manufacturing factories are limited.
Line managers and workers should author the assembly instructions. They are the shop floor experts.
Full control of the solution and ability to adapt it to different use cases would increase the acceptance rate.
<b>Hygiene &amp; Security - Manager</b>
Safety is our priority. Manufacturing factories have very strict safety procedures. If anything can potentially put a worker at risk, it will be rejected.
The repetition of gestures increases the risk of illness. We must avoid asking workers to perform extra gestures or movements. The assembly sequences are designed to minimize physical ergonomic risks.
<b>Production</b>
Simplicity and effectiveness are much more important than fancy. In manufacturing, we do not use tools because they are nice but because they help workers perform better and safer.
Assembly line workers are here for one reason. Equipping them with technological tools is not easy. It must be extremely easy to use, intuitive and safe.
We usually do not have extra time. Workers neither. Our objective is to keep the line going and provide quality products. For a new tool to be adopted, its simplicity and efficiency must be proven right away.
It must be so simple that everyone can use it. It should help us better train our temporary workers, as there is a huge improvement potential in this area.
<b>Planning</b>
The solution must be contextual and adaptive, simple to use and to understand, especially for shop floor workers and people without technical knowledge.

These findings, along with the ones reported by the preliminary studies presented in this chapter, allowed us to identify and categorize the most significant organizational and technical requirements and challenges that an AR training system should address, to be useful and eventually adopted in similar industrial context. These aspects concern existing resources, technical setup, assembly process and environment, shop floor safety, user acceptance and viability. A summary description of these requirements is further provided.

1. **Existing resources.** Ideally, the proposed solution is not dependent on existing assets like paper instructions or CAD data. Their reliability and utility are uncertain, especially for AR usage, therefore, the proposed training system should not rely on such contents. To support viability, the system should be as independent on existing resources and external services as possible.
2. **Technical setup.** The system should be mobile, flexible, and easy to install and maintain. A standalone application running on a mobile device, easily accessible is recommended. Integration with external services, system calibrations and other recurring processes are undesirable. The dynamics of the considered assembly environment suggests that the technical setup represents an important viability concern as well.

3. **Assembly process.** Many assembly operations require both hands, therefore the user interface must support hands-free interaction. It is preferable that the authoring of the AR work instructions is fast and reliable, considering the limited availability of the lines. Both the authoring and the training procedures should not require changes in the assembly process. Wired devices are not appropriate, as some assembly workstations cover distances up to 5m.
4. **Assembly environment.** The proposed solution should not be dependent on the assembly environment and should be functional and effective in different assembly use-cases. The system must adapt to new scenarios without requiring technical knowledge, considering that the end users are eventually shop floor experts.
5. **Safety.** Worker safety is an essential requirement in any shop floor environment. Providing dedicated workstations for training purposes is not conceivable in similar manufacturing sectors, therefore, it is essential that a new AR training method does not introduce any potential safety risks. The AR system should interfere as little as possible with the user, especially because studies regarding the physiological and psychological impact of AR systems are not yet conclusive.
6. **User acceptance.** The acceptance of the system is mostly driven by its utility and ease of use. Its potential benefit must be proven to shop floor experts immediately. The proposed solution should adapt to different assembly scenarios and not require expert knowledge. Ideally, the performance is not affected by the education level.

## 2.3 Prototyping and preliminary studies

### 2.3.1 Informal field studies

We conducted two informal field experiments on a real-life manual assembly line, denoted by L2. The first one consisted in familiarizing with the assembly line and its operations, while assisted by a line manager. In addition to discovering the assembly context, the objective of the study was to get a deeper understanding of the assembly process, identify key assembly operations and difficulties, and understand how the training procedure is conducted. The most relevant findings identified during this study, from the perspective of a novice worker, are:

1. Most of the assembly operations are specific. The most common task is screwing and can be found at multiple workstations of the assembly line.
2. Few assembly operations were difficult to perform, not considering time constraints.
3. The assembly sequence can be slightly adjusted at some workstations, in the sense that the order of tasks can be interchanged. This approach together with some other techniques (e.g., assembly preparation) are adopted by workers to gain time.

The second assembly line experiment consisted in performing the assembly of one of the most produced boilers in the factory (Figure 2-5), from scratch, by only using the corresponding existing paper instructions. The experiment has been conducted in cooperation with two other PhD students at elm.leblanc over a period of 3 weeks. The main objective of the study was to familiarize with the considered manufacturing context and assess the usability of the paper

instructions. We aimed as well at identifying critical assembly components that may have an impact (e.g., small, flexible assembly components) on the elaboration of the AR training system.



**Figure 2-5. Field experiment: assembly of a boiler in progress.**

To analyze the manufacturing process in detail in a later stage, in addition to the paper instructions, we have recorded our assembly work experiment by adopting two approaches.

First, we have used a head-mounted camera (i.e., GoPro) for visualizing what the worker sees during the assembly process and potentially identify if a first-person view (FPV) video demonstration can be used for automatically generating AR instructions. We figured that FPV demonstration videos alone are not sufficient for effectively capturing manual assembly expertise. We note, however, that we did not perform an objective evaluation of the feasibility of this approach from a technical perspective, assuming that literature findings are sufficient. On the other hand, we realized that FPV video demonstrations could represent an effective visual asset for conveying assembly information in AR.

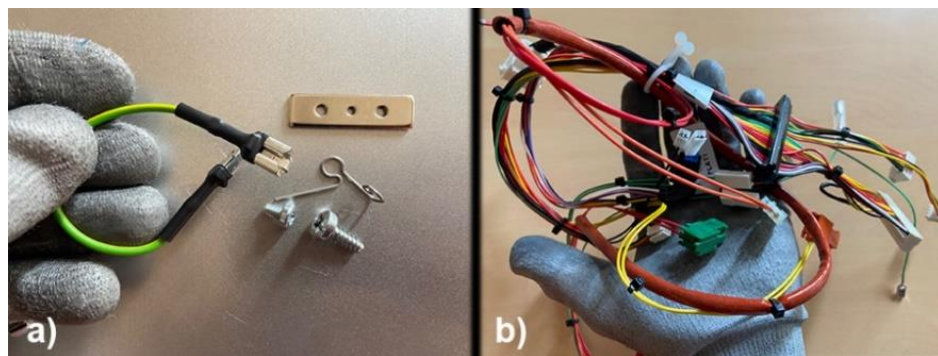
The second method consisted in recording, with the help of a smartphone, the assembly work executed at first sight. This technique was useful for capturing first impression remarks and other assembly details that we could further analyze offline.

The most relevant findings of this field experiment are listed below:

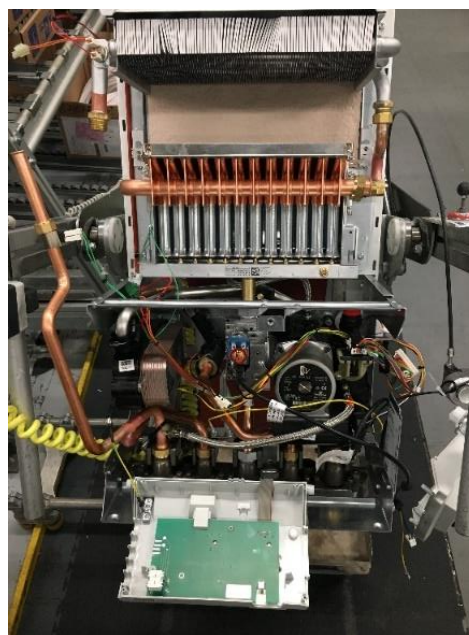
1. The paper instructions are impractical, especially at a first encounter with the workstation. The text instructions are generally missing important details, while the illustrations (e.g., assembly photos) are not always sufficient for a novice worker to understand the assembly process. For instance, the very first instruction of the line, “Place the upright on the assembly structure”, does not indicate neither the reference of the component to be assembled, nor where to be assembled and how. Similar issues have been identified for all workstations considered in the experiment.
2. The assembly experience acquired at a given workstation does not lead to faster understanding or better performance at another one, due to the specificities of each. This raises a problem in terms of re-usability of the AR content.

3. A third notable finding concerns small (Figure 2-6.a)) and flexible (Figure 2-6.b)) assembly components like screws, clips, and cables. The most important concerns are:
  - Information conveyance method - it is not clear what is the best method for conveying this type of assembly information via AR, neither what are the most suited visual assets for representing them.
  - Spatial registration - a very accurate spatial registration should be required in the case of using superimposed CAD models representations.
  - Specificity of cables - cables are flexible, and they generally allow slightly different installing positions; it is not clear how this information can be efficiently digitized and conveyed via AR.
  - Hidden assembly locations - some assembly operations take place in locations that are not directly visible to the worker, potentially rising occlusion concerns in AR.

Figure 2-6 and Figure 2-7 illustrate some examples of assembly components and states that are difficult to represent with CAD data in AR.



**Figure 2-6. Assembly components: a) small (e.g., screws, clips) and b) flexible (e.g., cables).**



**Figure 2-7. Assembly complexity: a boiler assembly state difficult to represent digitally in AR environments.**

The following conclusions summarize our findings:

1. The existing paper instructions are neither reliable nor practical. Their utility in generating or creating AR instructions is questionable.
2. Most assembly operations are unique except screwing and riveting. The potential benefit of assembly experience acquired elsewhere is very low.
3. Many assembly operations consist in manipulating small or flexible components like screws, clips, and cables. It is not clear how this information could be efficiently captured and represented in an AR training system.

To push further our investigations, we have set up a second experimentation, specifically dedicated to AR training methodology and described in the following section.

### 2.3.2 User study: AR training methodology for assembly operations

We have conducted an AR project with 14 students from a master's program at Telecom SudParis. The main objective of this work was to gather ideas from students, considered novice workers, regarding assembly information representation and conveyance in AR. From a technical perspective, we wanted to assess the potential of the state-of-the-art AR tools including HoloLens 1, Unity3D (*Unity 2019.4.10*, 2019) and HoloToolkit – i.e., the predecessor of MRTK (*Mixed Reality Toolkit | Microsoft Docs*, 2020), under real world conditions. Students had to propose and implement a methodology for delivering the assembly information in AR, starting from existing paper instructions, and using the corresponding CAD models of the assembly parts. Student's feedback and ideas have been collected during the entire period (i.e., 2 months) of the project. The study was concluded with a formal presentation where each proposal was described, demonstrated, and finally discussed in a Q&A session conducted by the CIO of elm.leblanc. The most relevant conclusions of this study are the following:

1. Assembly representation: combine visual and audio, especially for describing assembly details that might not be obvious to novice workers.
2. Multimodal interface: combine gestures and voice commands to allow users interact with the application by using their preferred interaction method.
3. Graphical elements positioning: consider peripheral space for the elements used intermittently. The AR information should be displayed in front of the user and only when needed, avoiding unnecessary interference with the assembly.
4. Guidance levels: the application should propose different guidance, based on the user profile and experience.
5. Statistics & feedback: collect statistics and provide feedback to training responsables.

Among the difficulties encountered and reported by the students we note:

- the technical implementation was found very complex, particularly the authoring of animated AR instructions,
- the interaction techniques (e.g., "air-tap") proposed in HoloLens 1 have been found insufficient and impractical,
- the lack of a standard AR conveyance interface and predefined dedicated toolsets to create AR instructions made the development a hassle.

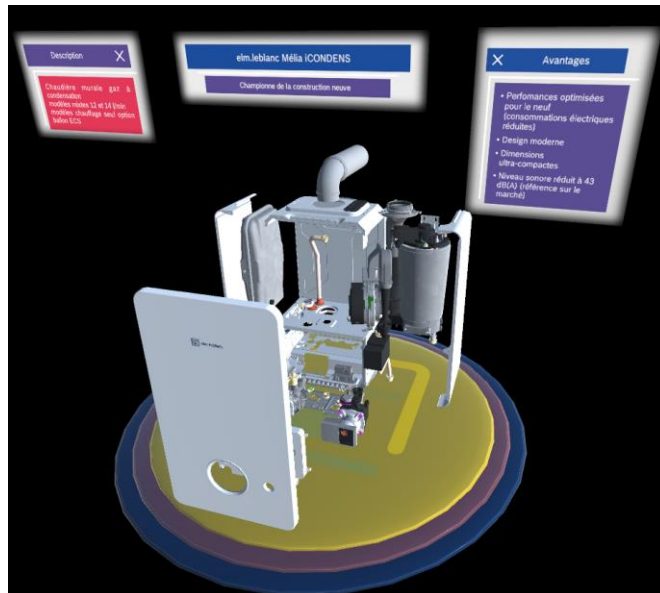
### 2.3.3 Prototyping and evaluations

To further explore state-of-the-art AR technology and assess their pertinence from a shop floor perspective, we have proposed the implementation and evaluation of two AR prototypes, to:

- identify organizational and technical challenges for authoring AR instructions by using existing digital content - i.e., CAD models and paper instructions,
- demonstrate and collect feedback from the shop floor experts potentially concerned by the development and deployment of such solution in the factory.

#### 2.3.3.1 AR Boiler Showroom

To assess 3D modelling challenges and evaluate the features of state-of-the-art AR devices (i.e., Hololens 1), a 3D boiler showroom application has been developed. The application makes it possible to visualize and inspect in AR one of the most popular boilers manufactured at elm.leblanc, in both normal, and exploded views. Users can switch between the two modes by interacting (i.e., click – “air-tap”) with the 3D model of the boiler and visually inspect it by moving around.



**Figure 2-8. AR Boiler Showroom. An AR application for visualizing and inspecting in AR the corresponding 3D model of a boiler produced at elm.leblanc.**

The main difficulties encountered during the development of this application are the following:

- existing CAD models are not AR-friendly: their high accuracy leads to prohibitively important file sizes (around 500 MB on average) Moreover, they are stored in a proprietary format (i.e., AutoCAD DXF), not compliant with the most popular tools used for AR development like Unity 3D. To address such concerns, a manual intervention is required to simplify and convert the considered model into a compliant file format (e.g., OBJ, FBX),
- specific expertise is required for finding and retrieving the right version of the CAD model from the storage system, corresponding to the considered boiler.

### 2.3.3.2 TrainingAssistant

A second prototype is another Hololens 1 application, in which we “digitized” an existing paper instruction as an interactive replica in AR, with the following objectives:

1. Explore and identify AR interaction and visualization techniques adapted to shop floor workers, which do not have any AR expertise.
2. Design an interface that potentially can serve as a template for displaying AR instructions for manual assembly operations.
3. Identify benefits of displaying assembly instructions in AR, compared to paper.

To accommodate shop floor workers, we have decided to design the UX of the application as simple as possible. To navigate through the AR instructions, the user selects them from the AR replica of the paper instruction by using the head gaze, indicated by a virtual cursor in the AR scene. Performing a selection requires the user to keep the head gaze cursor over the button of the corresponding instruction for 3 seconds. This interaction technique addresses the hands-free requirement, an essential condition identified early on in this research work and replaces the native *click* interaction technique proposed in Hololens (i.e., “air-tap”). This interaction technique has been later implemented in Guides (*Microsoft Mixed Reality / AR Guides / Microsoft Dynamics 365*, 2019), by using eye gaze.

The prototype offers four main views:

- a welcome screen that asks the user to identify the workstation by scanning the corresponding QR code (Figure 2-9.a),
- an interactive replica of the corresponding paper instructions (Figure 2-9.b), displayed to the left-side of the workstation,
- a panel with the line and workstation statistics (Figure 2-9.c), displayed to the right-side of the workstation,
- assembly animations that complement some of the instructions (Figure 2-9.d).



**Figure 2-9. AR TrainingAssistant screenshots. a) Welcome screen, b) “Instructions” screen, c) “Statistics” screen, d) A 3D animation of an assembly operation.**

Informal field experiments have been conducted with six people from the factory, including shop floor workers. Their feedback was categorical: unanimously, they found that the proposed AR approach is not adapted for effectively conveying manual assembly information. For this reason, we did not consider any further evaluation. The main concerns identified are the following:

- as for the paper instructions, the AR interactive replica does not indicate the positioning of the components and assembly locations, which makes the instructions impractical for a novice worker,
- the positioning of the two panels (i.e., *Instructions* and *Statistics*), at the sides of the user, visually obstruct the assembly line. In addition, the user is required to perform repetitive gestures for reading the AR instructions,
- the *click* (i.e., “air-tap”) interaction was required in few cases, to allow a subjective evaluation between the two modes, eye gaze vs. air-tap. The latter was found slow and counterintuitive even for people with AR experience, suggesting that better interaction techniques should be proposed, particularly for people without AR experience,
- overall, reading the AR instructions took longer than reading the printed instructions while the effort seemed to be higher as well due to the unfamiliarity of the user interface.

The sole positive remark concerns the 3D animations (Fig 2. d). Participants were excited about this visual information conveyance method and found it easy to understand.

### 2.3.3.3 *On-the-Fly, In-Situ Authoring Tool (ATOFIS-BETA)*

The authoring process is a key issue that needs to be addressed in an early stage, with a major impact on the AR information conveyance and implicitly on the training.

To consider such a requirement, we have developed a first authoring tool proposal, so-called ATOFIS-BETA, which adopts a one-step, *in-situ* authoring paradigm that takes place directly and solely inside an AR headset.

ATOFIS-BETA was developed for Hololens 2, by using Unity3D 2019 (*Unity 2019.4.10*, 2019) and MRTK 2.3.0 (*Microsoft MRTK v2.3.0*, 2020), (*Mixed Reality Toolkit | Microsoft Docs*, 2020). The technical implementation relies on features provided by Hololens 2 including hand tracking, eye tracking, 6DoF tracking, voice, and AR capture.

We note that the proposed authoring is adapted to assembly procedures that can be described by the “*grab and place*” action, as defined in (MTM-UAS | MTM, n.d.)(MTM-UAS | MTM, n.d.). We observed that many assembly operations found during our case study follow this principle, where the worker is required to first pick a component from the storage area and then execute an assembly operation by using the considered component.

To create AR work instructions, the author follows the following steps:

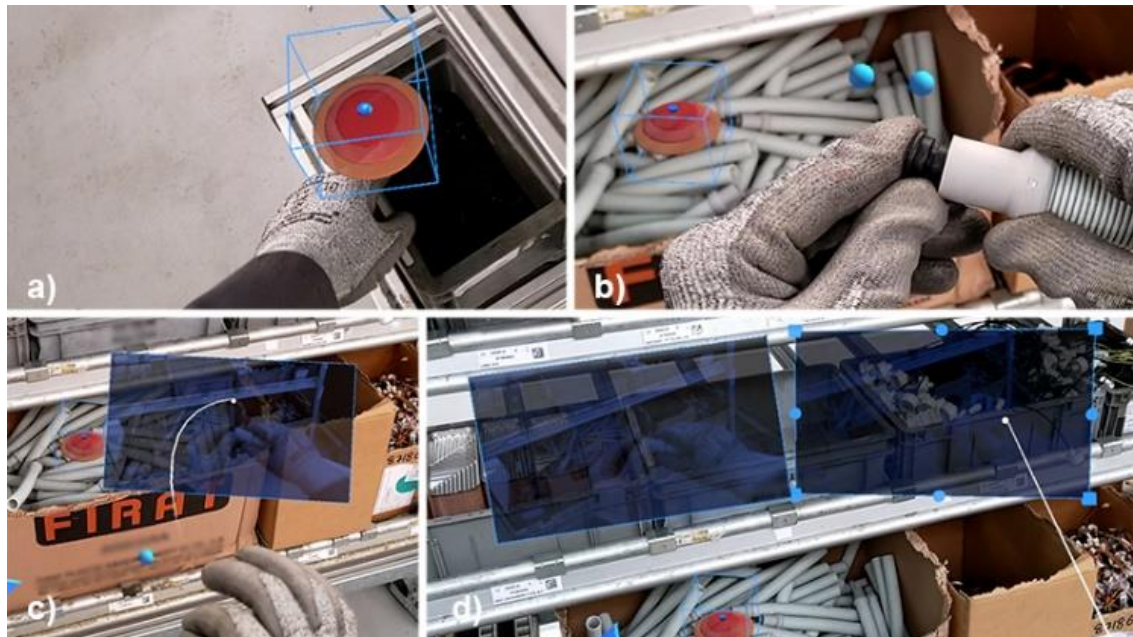
1. Go to workstation, open application, select application mode (i.e., authoring).
2. Go to the location of the assembly component.
3. Indicate the location of the assembly component in the real-world environment by using the index fingertip and place a location anchor (i.e., visual cue) by using vocal command. This indicates the location of the component to be assembled and the beginning of a [*pick – execute*] assembly pair. Once the anchor added, the AR elements corresponding to the previous AR instruction are hidden to avoid visual interference and UI clutter.
4. (Optional) Capture photos and videos illustrating the picking operation by using vocal commands. Manipulate (i.e., size, rotation, position) them by using natural hand



interaction, for contextualization purposes. All the captured media, until the second anchor is placed, corresponds to the first anchor, and describes the assembly picking procedure.

5. Go to the location of the assembly operation.
6. Perform the same actions described in Step 3, for indicating the assembly location. The second anchor indicates the physical location of the assembly operation and, at the same time, the end of a *[pick – execute]* assembly pair.
7. Same as Step 4, for demonstrating the assembly operation. All the captured media, until a new anchor is placed, corresponds to the second anchor, and describes the assembly procedure. The author is authorized to capture an arbitrary number of photos/videos.
8. Loop through steps 2-7 for authoring a new AR instruction.

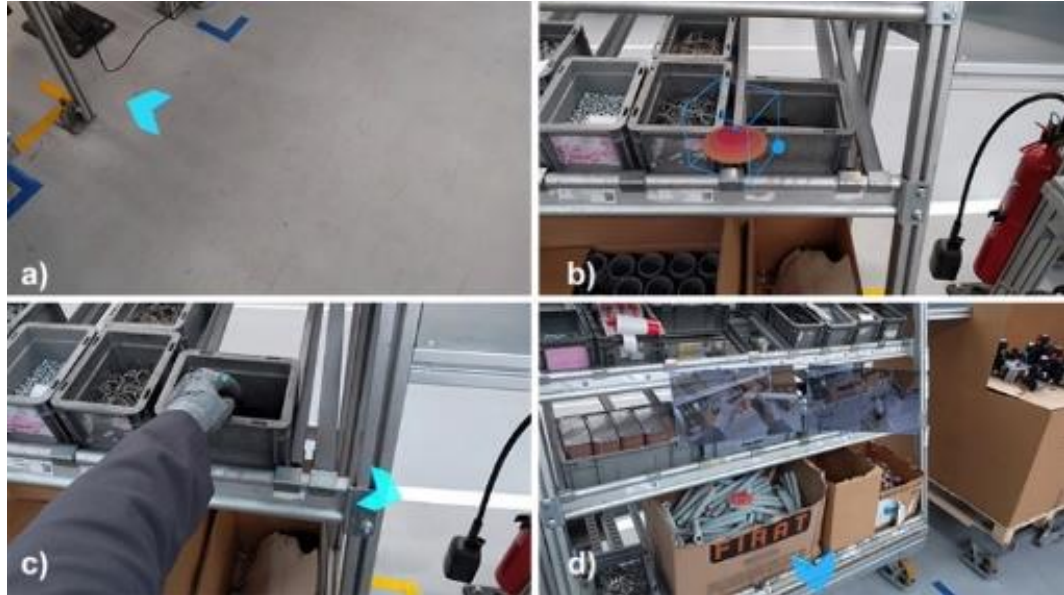
The authoring application interface is minimal. The author uses natural gestures and vocal commands to place the location anchors and capture photos and video demonstrations. No other visual elements (e.g., menus, buttons) are used. Figure 2-10 illustrates a series of screenshots from the authoring process of an assembly operation, including placing a location anchor for indicating the physical location of the picking operation (Figure 2-10.a)), a photo capture illustrating the assembly operation (Figure 2-10.b)), photo manipulation by using far interaction technique (Figure 2-10.c)) and finally a representation of multiple elements used to describe the assembly task, contextualized, and arranged in a certain manner (Figure 2-10.d)).



**Figure 2-10. ATOFIS-BETA authoring workflow illustration: a) Location anchor placement – point with index fingertip & use vocal command, b) Photo capture – vocal command, c) Photo element manipulation – hand gestures, d) Capture and manipulate additional photos/videos.**

Let us also briefly explain how the information is conveyed during the training and how the trainee interacts with the application. First, the trainee turns his head in the direction indicated by the directional arrow to identify the assembly component location (Figure 2-11.a)). Once the picking location enters in the field of view (FoV) of the trainee, the directional arrow is hidden.

The location anchor, now in the FoV, indicates the physical location of the assembly component (Figure 2-11.b)). The trainee grabs the assembly component indicated by the location anchor. Once touched, the location anchor is hidden, and a new indication arrow points towards a new assembly location, which indicates the next component to be picked (Figure 2-11.c)). Following the same principle, the trainee executes the remaining assembly tasks.



**Figure 2-11. ATOFIS-BETA training procedure illustration: a) Directional arrow indicating the picking location, b) Location anchor indicating the assembly location, c) The trainee picks the assembly component: the location anchor is hidden, and a new indication arrow is displayed, d) The trainee identifies the next anchor, visualizes the augmentation media, and performs the assembly operation.**

Concerning the interaction techniques retained, the vocal and gestures commands used during the authoring process are summarized in Table 2-6.

**Table 2-6. Proposed vocal commands and gestures used during the authoring**

Vocal commands	Authoring result
<i>place anchor</i>	A predefined 3D model is placed at the location indicated by the forefinger.
<i>take photo</i>	Captures a photo and presents it in front (1.5m away) of the user.
<i>record video</i>	Starts recording a video.
<i>stop video</i>	Stops the video recording and displays the result in the same way as with the photos.
<i>restart</i>	Restarts the authoring process by removing all instructions and their associated AR elements.
<i>training mode</i>	Switch application to training mode
<b>Gestures</b>	
point with index fingertip (use with <i>place anchor</i> )	Indicates the physical location of a new anchor that will to be generated and spatially registered.
object manipulation	Scale, rotate and move anchors and captured media.

The key elements of the authoring procedure and of the information conveyance approach are respectively illustrated in Figure 2-12 and Figure 2-13.

HMD feature	User action	Authoring Result
Voice AR capture	<i>take photo</i> <i>record video</i> <i>stop video</i>	Capture video and photos
Vocal commands Finger tracking	Indicate with fingertip + <i>place anchor</i>	Place anchor
Hand gestures	Hand gestures: pinch, drag, release	Manipulate AR elements: scale, rotation, position transform

**Figure 2-12. ATOFIS-BETA authoring: summary of authoring key elements**

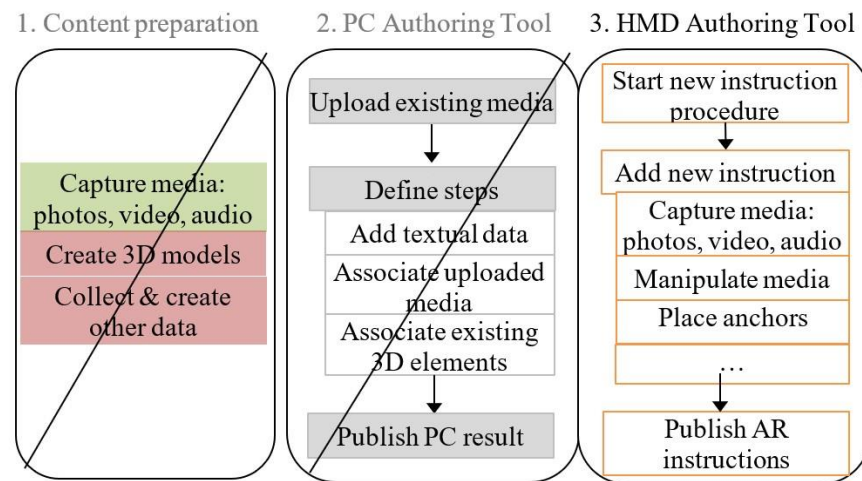
HMD feature	User action	Result
6 DoF tracking	Follow arrow indication	Visualize location anchor Visualize AR elements
Eye tracking	Look at video / away	Play video / pause
Hand tracking	Touch anchor	Remove Anchor Display indication arrow Display next anchor + associated AR elements
Voice	<i>next</i> <i>hide / show</i>	Update AR content for next instruction Hide / show all AR content

**Figure 2-13. ATOFIS-BETA information conveyance: summary of training key elements.**

#### 2.3.3.4 Field experiment and comparative evaluation

To validate the proposed prototype, we have performed a comparative evaluation of ATOFIS-BETA with the most popular AR authoring tool available today, Microsoft Dynamics 365 Guides (*Microsoft Mixed Reality / AR Guides | Microsoft Dynamics 365, 2019*). This study has been carried out under the conditions of a real-life field experiment.

Figure 2-14 illustrates the differences, in terms of authoring stages, between Guides and ATOFIS-BETA. Guides authoring requires three successive stages, including content preparation, PC authoring and HMD authoring. In contrast, ATOFIS-BETA is comprised of a single stage, the HMD authoring.



**Figure 2-14.** ATOFIS-BETA vs. Guides authoring workflow and stages: the crossed-out blocks are no longer required by ATOFIS-BETA.

The preliminary field study considered evaluates the two authoring procedures from a time perspective. Three participants (2 PhD students one R&D engineer) two females and one male aged between 25 and 27 years old, have been asked to perform the evaluation. The participants have basic AR knowledge and already tested AR HMDs devices in the past but have no assembly experience. The experiment consisted in authoring 15 AR work instructions of an assembly workstation twice, by using each of the considered authoring systems. The participants have been guided by the evaluator during each stage of the procedure, to ensure the creation of the same AR work instructions between the two systems from a content perspective (e.g., *instruction 1* – 2 photos, *instruction 2* – 1 video, *instruction 3* – 1 photo and 1 video, etc.). Each AR instruction defined a [*pick and execute*] operation therefore two location anchors have been added for each such pair.

The user evaluation was performed as follows:

1. The evaluator presents the context and the objective of the experiment.
2. The evaluator explains the authoring procedures with both systems, Guides (i.e., PC & HMD) and ATOFIS-BETA (i.e., HMD).
3. The participant learns the 15 considered assembly operations guided by the evaluator.
4. The participant authors the AR work instructions by using each of the two systems.

Table 2-7 reports the time needed for the presentation and explanation of the two systems, while Table 2-8 summarizes the effective authoring times obtained by the users.

**Table 2-7.** ATOFIS-BETA vs. Guides: presentation times.

Participant	Guides			ATOFIS-BETA
	PC authoring	HS2 authoring	Total	HS authoring (total)
1	4'12"	14'24"	18'38"	11'10"
2	5'03"	10'40"	15'43"	12'07"
3	5'03"	9'50"	14'53"	11'15"
Avg.	4'46"	11'38"	16'24"	11'30"

**Table 2-8. ATOFIS-BETA vs. Guides: reported authoring times.**

Participant	Guides			ATO FIS-BETA	
	<i>Photo / video capture</i>	<i>Authoring PC</i>	<i>Authoring HS2</i>	<i>Total</i>	<i>Total</i>
1	5'26''	11'23''	15'17''	32'06''	9'13'' (3.57x)
2	4'11''	7'40''	9'56''	21'47''	7'22'' (3.03x)
3	3'07''	9'35''	8'38''	21'20''	7'26'' (2.94x)
Avg.	4'15''	9'33''	11'17'	<b>25'4''</b>	<b>8'01'' (3.13x)</b>

We can observe that ATOFIS-BETA makes it possible to obtain significant gains in time, for both presentation and authoring.

The reported authoring times, presented in Table 2-8, indicate that multiple authoring stages make the overall authoring process longer. By removing the preparation stage on the PC and performing the media capture *in-situ*, during the authoring process, our proposal becomes 3 times faster than Guides. The HMD authoring with ATOFIS-BETA was faster as well, even though photos and videos have been captured during this stage, potentially thanks to the placement technique of the location anchors and to the automatic handling of the indication arrow.

The users then graded on a scale from 1 to 10 the complexity of the system (Table 2-9).

**Table 2-9. ATOFIS-BETA vs. Guides: systems complexity.**

Subject	Guides	ATO FIS-BETA
1	6	2
2	7	3
3	7	4
Average	<b>6.66</b>	<b>3</b>

ATO FIS-BETA was perceived as three times less complex than Guides, a claim supported by the presentation times reported in Table 2-7 as well. The higher value of the level of complexity of Guides was mainly influenced by the required authoring steps. ATOFIS-BETA's complexity was given by the interaction techniques and by the fact that authoring an AR instruction required following the "*grab*" (first location anchor) and "*place*" (second location anchor) rule. Participants however appreciated the proposed authoring approach as the AR work instructions can be created during the execution of the assembly work, with a minimal user interface interference. The media capture during the authoring process was considered key.

ATO FIS-BETA was preferred in all cases since all participants have found it easier to use, faster, more practical, and less boring. They particularly appreciated the fact that the AR authoring is performed at the same time with the assembly procedure. However, two main limitations of the experiment should be noted. Firstly, the very reduced number of participants (3) suggest that the reported results are solely informative. Secondly, the authoring technique was designed particularly to excel at speed, compromising in a certain manner user interface design and functionality. Nevertheless, the overall reported results suggest that improving the proposed authoring approach with a user-friendly interface and a more generic way of describing assembly operations in AR (potentially including text and CAD data), should be considered further.

## 2.4 Conclusion

The main contribution of this industrial case study is represented by the in-depth analysis conducted in a human-operated boiler-manufacturing factory, at elm.leblanc, assisted by the direct implication of people affected directly (e.g., shop floor workers) and indirectly (e.g., management) by the potential adoption of an AR-based training system. The objective of the study was to obtain a comprehensive picture of the industrial needs and requirements, from both technical and organizational perspectives, that an AR training system should fulfill to optimally address and be adopted in such challenging environment.

An analysis of the manual assembly environment, together with inquiries, open discussions, semi-structured interviews, prototype development, and finally informal field experiments allowed us to conduct formative research to determine the critical aspects that should be considered in the elaboration of an AR training system adapted to industry. We found that broadly speaking, from a shop floor perspective, effectiveness and viability are the key success factors. These are not new to the literature (Masood & Egger, 2019); however, this analysis allowed us to validate and categorize expectations and challenges of manual assembly (presented in detail in Section 2.2.5), obtain a deep understanding of their root cause and eventually help in elaborating an AR training system that overcomes these concerns by finding the best technological compromises while considering environmental and human factors.

In addition, the findings of this industrial case study suggested that an analysis of the state of the art is not only required but should be conducted by questioning the extremely low adoption of AR-based training systems in industrial sectors. Such analysis is essential to identify if and how existing AR training tools proposed in the literature are addressing industrial concerns, which are the most relevant scientific locks, how significant are they in relation to considered industrial case study and finally identify how they can be addressed in this work.

## Chapter 3. RELATED WORK

---

*Most AR systems are elaborated and evaluated in controlled setups, reason for which they generally fail to address industrial needs and expectations.*

---

---

In this chapter, an overview of AR content authoring and training tools and methodologies is proposed. The analysis focuses on research works dedicated to creating and conveying AR step-by-step work instructions for manual assembly in industrial sectors but also investigates AR topics and concerns that are of general interest. AR content authoring and information conveyance in AR are generally explored separately in the literature. However, the holistic nature of an AR training system strongly suggests that the two functions should be considered and analyzed as a whole, notably for real-world industrial usage. We address this recommendation throughout this state-of-the-art analysis, by highlighting potential implications that one function may have on the other.

---

---

### 3.1 Content authoring in AR

An authoring tool, also referred to as content authoring tool or authoring system, is a high-level visual software that enables one to create content or applications without requiring programming or other technical expertise. They are most often template-based and use WYSIWYG (*What You See Is What You Get*) interfaces, to provide an easy and fast way of creating content by using familiar visual metaphors. Authoring tools can be traced back to the 1960s when computer-assisted instruction was viewed as an economically viable way to distribute teaching expertise (Dabbagh, 2002). Historically speaking, authoring tools have been created for instructional designers, educators, and teachers to create interactive multimedia and hypermedia learning environments for the Web technology (Dabbagh, 2001). However, authoring tools can be considered today within any digital content technological field, including AR. Authoring tools for AR applications is not a novelty. However, presumably because the AR technology has not yet become mainstream, existing AR authoring tools are rather simple, provide narrow functionality and are generally dedicated to entertainment, except for some commercial solutions, such as Guides, Vuforia (*Vuforia Expert Capture* | PTC, 2022) and Taqtile Manifest (*Manifest® AR Platform Makes Everyone an Expert* | Taqtile, 2022).

#### 3.1.1 State of the art surveys and classification

Before diving into specificities regarding the design, implementation and functional details of authoring methodologies and techniques proposed in the literature, we first aim to obtain a general understanding of the status and the potential of AR content authoring, by analyzing the various taxonomies proposed in the literature.

Generally, AR authoring tools are classified from low to high-level, according to their programming and content design characteristics. In this sense, development tools for building AR applications can be broadly organized into two different approaches: AR authoring for programmers and AR authoring for non-programmers. (Mota et al., 2015) categorized 19 commercial and academic AR content design tools by authoring paradigms (i.e., stand-alone and AR-plugin) and deployment strategies (i.e., platform-specific and platform-independent). We observe that most of the analyzed tools are stand-alone and platform-independent and that the plug-in approach potentially offers more features than the stand-alone approach. The proposed taxonomy however does not investigate neither the scope of these tools, nor corresponding features and characteristics. In addition, no assessment or functional evaluation between these authoring tools is provided. It is difficult therefore to determine which of the dataflow models is most adapted to certain authoring scenarios and requirements.

(Nebeling & Speicher, 2018) tackled this problem and classified AR authoring tools by their level of fidelity in AR/VR and by the required skills and resources involved. Five classes have been identified, including (1) basic mobile screens and interactions, (2) basic AR/VR scenes and interactions, (3) AR/VR focused interactions, (4) 3D content and (5) 3D games and applications. The authors concluded that AR authoring tools must still address numerous challenges for making people without programming knowledge able to author complex AR experiences. The related challenges however are not clearly listed; neither a set of requirements provided, making



it hard to identify what functionalities the AR authoring tools are missing or which of the existing authoring functions proposed by such systems need improvements.

More recently, (Bhattacharya & Winer, 2019) have introduced a different taxonomy for classifying existing AR authoring tools, based on their intended design, as follows: linking system (Haringer & Regenbrecht, 2002), (Grimin et al., 2002)), AR previewer (Zauner et al., 2003), (G. A. Lee & Kim, 2009), virtual registration (Knöpfle et al., 2005),(Moura et al., 2011), hybrid methods (S. K. Ong & Zhu, 2013), context-aware (Zhu et al., 2013), knowledge based (Jo et al., 2014) and finally, third party packages (Park et al., 2016), (Mossel & Venditti, 2013). Further, a classification of AR authoring tools interfaces was provided, based on how the virtual content is authored. This authoring classification included desktop GUI based (Tsihrantzis et al., 2010), mobile AR (Markouzis & Fessakis, 2015), (Yang et al., 2016), HMD with 2D/3D camera sensor (X. Wang et al., 2015), hybrid (Ha & Woo, 2007), (Yu et al., 2016) and finally, demonstration-based (Bhattacharya & Winer, 2019).

We observe that the literature tends to categorize AR authoring tools by using multiple taxonomies. However, a set of well-established requirements and guidelines for elaborating use-case-specific AR authoring tools dedicated to industrial sectors is not provided, potentially because of the lack of field experiments and real-world evaluation. This could explain why 48% of the AR solutions do not target specific use cases or industrial sectors, but rather seem to be designed with a general-purpose objective (de Souza Cardoso et al., 2020).

Let us further analyze the main AR authoring tools and techniques proposed in the literature that we consider relevant to our research work.

### **3.1.2 Methods and systems**

We specifically focus on AR techniques targeting a potential industrial usage and dedicated to creating AR step-by-step work instructions.

#### *3.1.2.1 Fully automatic authoring*

The best AR authoring system is the one that can generate AR work instructions automatically, without human intervention and ideally without prerequisites and processing stages. Such a system should be able to connect and extract the necessary information from various data content providers (DB, live data feed, etc.) and generate the corresponding AR work instructions automatically. (Palmarini et al., 2018) point out that 64% of the AR contents are manually generated, suggesting that a rather high percentage of AR content (36%) is generated in a fully or semi-automatic manner.

One such research work (i.e., AREDA) has been conducted by (Bhattacharya & Winer, 2019). Authors propose an AR authoring system able to automatically generate AR work instructions from a video recording of an expert performing the assembly. This approach is called demonstration-based authoring, and involves two stages, which are demonstration and refinement. The demonstration phase requires several procedures including the capture setup, background calibration, area calibration, skin calibration and finally the processing of the recorded demonstration. Authors claim that these steps are easy to perform and do not require

computer vision (CV) or 3D graphics expertise. The refinement phase proposes a set of “layers” that allow the author to enrich the AR work instructions with 2D text, images, or additional video elements. The technical limitations concern the complex system setup, and the lack of a user evaluation, which questions the usability of the system particularly in a real-world industrial context. In addition, authors show that the presence of relatively complex assembly setups and/or small components to be assembled represent a significant challenge for the proposed authoring system. We underline however some potential advantages of such an authoring approach, as the author of the AR instructions (i.e., an assembly expert) is only required to perform a calibration setup and the assembly procedure itself, while the system captures and generates the corresponding AR work instructions automatically.

#### 3.1.2.2 *Video retargeting*

Another automatic generation technique of AR work instructions, so-called video retargeting (Mohr et al., 2017), (Eckhoff et al., 2018), generates interactive 3D tutorials from existing 2D video tutorials. This authoring technique relies on computer vision to detect and track objects in the video (Mohr et al., 2015) and consequently generate instructional 3D animations (Eckhoff et al., 2018). However, the analysis techniques remain limited to the performances of the considered computer vision algorithms, which restricts the capabilities of the obtained instructions.

Moreover, most retargeting approaches are only demonstrated on assemblies which consist of large detectable parts, making them impractical for our research context. The video retargeting method proposed in (Yamaguchi et al., 2020) claims to address the detection of small parts. However, the related limitations are significant: (1) larger parts are still needed for tracking, (2) the system can only process videos where objects are continuously manipulated in a series of video frames without interruptions, (3) the corresponding CAD model together with a detailed assembly video must be provided, while (4) lighting, background contrast and camera calibration play an important role in the detection performances. It seems therefore very unlikely that such an authoring approach could provide acceptable results in a real-life industrial environment.

#### 3.1.2.3 *Product disassembly*

Another automatic AR work instructions authoring approach, so-called AR-guided product disassembly (ARDIS) and dedicated to maintenance and remanufacturing purposes is proposed in (Chang et al., 2017). The system consists of two major modules, disassembly sequence planning and automatic content generation. A third one represents the AR interface and conveys the work instructions for guidance purposes. The product disassembly sequence planning requires the definition of an assembly sequence table (AST) in which the contact and the translation functions between all parts of the product are defined. A disassembly sequence table (DST) is obtained from the AST and other additional constraints, which finally generates the optimal disassembly sequence of the product. The second module, the automatic content generation, uses visual data (e.g., text, 2D graphics and images, and CAD models) to generate the corresponding AR disassembly instructions. Unfortunately, the description of the overall authoring process is not complete, and the procedure does not seem to be fully automatic. We note that the definition of the AST requires human input for specifying relations and constraints

between the product's components. Further, it seems that the textual and 2D/3D graphics used for visually describing the generated AR disassembly instructions need to be provided to the authoring module beforehand. The correspondence between such media assets and the generated assembly steps also requires manual human intervention. In a general manner, the pertinence of the system in an industrial setup remains highly questionable.

#### 3.1.2.4 *Printed documents to AR instructions*

(Mohr et al., 2015) propose a system that automatically converts printed technical documentation (such as handbooks) into 3D AR instructions. The system identifies the most frequent forms of instructions found in printed documentation, such as image sequences, explosion diagrams, textual annotations and arrows indicating motion. The analysis of the printed documentation requires the documentation itself and a CAD model or a 3D scan of the corresponding real-world object. The authoring is semi-automatic, as some form of user input is required. The output is an interactive AR application, presenting the information from the printed documentation in 3D, superimposed on the real object. The proposed approach has been designed for basic manipulation operations (e.g., place, rotate, slide), while one of the system prerequisites is a 2D assembly diagram of the product model.

Other similar research works retargeting technical documentation to AR step-by-step instructions have been conducted in (Zauner et al., 2003), (Gupta et al., 2012), (W. Li et al., 2004) and (Shao et al., 2013). Such authoring methods are not adapted to use cases where 2D diagrams (e.g., annotated, action, explosion, structural) and technical drawings do not exist, for complex assembly operations and for small or deformable parts. Let us observe that the use of existing digital data or paper instructions as raw input raises various concerns. Firstly, as the work instructions are not described in a formalized manner across different industrial sectors and factories, expert intervention is required for adapting the AR authoring system from one use case to another. Secondly, the quality of work instructions is generally sparse. The main categories used to classify quality issues include intrinsic and representational problems, unmatched, questionable, and inaccessible information (Haug, 2015). (D. Li et al., 2018) claimed that work instructions are often insufficient and unused.

In a recent related study, (Gattullo et al., 2019) identify that the implementation of AR documentation in industry is still challenging because specific standards and guidelines are missing. They propose a methodology for converting traditional documentation to AR-ready documentation, as a prerequisite for creating the corresponding AR work instructions. However, redesigning and recreating existing paper instructions prior to authoring the corresponding AR work instructions contradicts industrial expectations.

#### 3.1.2.5 *Machine learning-based techniques*

The recent advances in artificial intelligence methodologies, and in particular deep learning techniques, have also impacted the field of AR authoring procedures.

Recently, (Mourtzis et al., 2020) introduced a framework for automatic generation of AR maintenance instructions based on convolutional neural networks (CNN). The framework consists of three main modules: a spatial recognition module, a CNN module used for identifying

(predicting) the model of the assembly device from a video stream, and an AR instructions generation module. A preliminary experiment conducted in a controlled environment shows that the generated AR instructions presented in a Hololens 1 device allowed the 10 participants to accomplish a disassembly task in 60 minutes, compared to 25 minutes required by an experienced technician. The limitations of the proposed system relate to the computer vision module which does not provide results in real time. In addition, the computational time increases exponentially with volume of the input data. Secondly, the CNN was only able to reach an accuracy of approximately 80% in recognizing simple features from an image dataset, with a fluctuation of  $\pm 10\%$ . This represents a rather poor performance even for laboratory evaluation settings. Finally, the prerequisites of the system, including the availability of large data sets of thousands of images required for training the CNN and of 3D CAD models for generating the AR instructions, question as well the usability of the proposed system in an industrial setup.

#### 3.1.2.6 *Motion-capture*

Another method for capturing assembly expertise, briefly discussed in the literature, is represented by human motion capture (MOCAP) systems (Pilati et al., 2020), (Sato & Cohen, 2010). However, motion capture systems are expensive and difficult to setup and operate in an industrial context, while the generated motion cannot be edited or adapted to different task constraints (SPA et al., 2014). Globally, the MOCAP approach is not yet sufficiently explored to be effectively employed in AR content authoring process under challenging conditions.

#### 3.1.2.7 *Context-aware and adaptive authoring*

Finally, one of the most promising approaches for creating AR work instructions is represented by the so-called context-aware and adaptive AR authoring.

The ARUM system introduced in (Erkoyuncu et al., 2017) attempts to improve the maintenance efficiency through adaptive operational support. ARUM comprises two platforms, a first one that enables the automated authoring process and a second one that allows maintenance experts to interact with the information frameworks used for generating the AR contents. Thus, the authoring steps are automatically generated by the system, while maintenance experts focus on information representation, defining its format and sequence of operations. A preliminary study, conducted with 6 participants shows that ARUM is 4 times faster than Vuforia during authoring, and 2 times faster than paper-based instructions during maintenance. However, the authoring process involves a multi-step procedure, the usage of a desktop application is mandatory and appropriate 3D content needs to be available.

A similar work dedicated to assist maintenance workers with context-aware AR has been introduced in (Zhu et al., 2013), (Zhu et al., 2014). The system, so-called ACARS, can analyze the context and provide relevant information to the user in real time by conveying textual descriptions, indication arrows, product models and animations. It consists of five modules - i.e., Context Management, AR-based Visualization, Database, Offline Authoring, and On-site Authoring, and requires the use of a desktop PC, programming skills, and domain experts. We note that in addition to the functional complexity of the proposed authoring framework, the study did not report on the performance of the authoring process (e.g., time, error rate, mental

workload). The system has been evaluated solely in a preliminary field experiment with 8 participants and reported better qualitative scores when compared to paper-based and traditional AR-assisted approaches. The subjective feedback revealed multiple concerns, including tracking and registration accuracy, rendering issues, and non-intuitive interaction.

(Geng et al., 2020) propose a systematic design method for authoring adaptive and customizable AR work instructions for complex industrial operations. The proposed authoring workflow consists of 3 separate stages, i.e., planning, enriching and post-processing, and requires 2 experts (an engineer and an assembly worker). The authoring starts with the planning stage which is conducted by an engineer using a desktop application, then continued by the worker on the field using AR equipment and finally completed by the engineer with the post-processing stage. The reported authoring time for an assembly scenario consisting of 19 operations is approximately 120 minutes. Even though the acceptability score (73/100) and the perceived mental workload (51/100) reported by the participants are acceptable, it seems that the proposed authoring method is globally complex and slow. We note that 50% of the participants (i.e., 5 out of 10) rated the system with a score lower than 70, which is considered the minimal threshold value of a system to be considered acceptable. Moreover, the large size of the ontology entities used during the authoring adds complexity to the authoring system.

The CAPturAR method proposed in (T. Wang et al., 2020) is an *in-situ* programming tool that supports users to author context-aware applications by referring to their previous activities. The system consists in a customized AR head-mounted device with multiple camera systems that allow for non-intrusive capturing of the user's activities. During authoring, the captured data is reconstructed in AR with an animated avatar and virtual icons that represent the surrounding environment. The proposed framework employs an object-orientated programming metaphor, allowing the user to define *events* with *human actions* and *context attributes*, and finally create context-aware applications (CAPs) by connecting *events* with IoT functions. From a hardware perspective, the system is composed of a VR headset and a touch controller, a stereo camera for video see-through AR and object detection, a fisheye camera for action recognition and an AR HMD connected to a backpack computer. A preliminary, remote user study with 12 participants has been conducted in a laboratory setting to evaluate the event definition precision and the overall usability of the system. The authored CAPs consisted in activities like reading books, heaving meal, taking a pill, and drinking coffee. While all participants managed to successfully complete the authoring tasks and the obtained usability scores indicated a good usability of the proposed user interface, the study did not report neither on the task difficulty and authoring time, nor on the usability of the authored CAPs. Even though the proposed system has been evaluated on four different scenarios, including a sequential task tutorial, relevant to our research, the authoring technique, and the hardware configuration question its suitability for industrial environments, particularly its effectiveness for authoring manual assembly operations.

The literature analysis presented in this section mainly focused on AR content authoring approaches. Let us now analyze how the aspects related to training and information conveyance in AR are covered in the state of the art.

## 3.2 AR training: information conveyance and guidance

In this section, we propose a state-of-the-art survey on AR training and guidance systems and methods, focusing on those designed to support workers in manual sequential assembly operations. We aim to identify how AR is used in similar research works, including information representation (e.g., visual contents like text, video, 3D, etc.), visualization hardware (e.g., HHD, HMD, SAR, etc.), registration methods (e.g., head, world, etc.) and picking techniques (e.g., pick-by-light or pick-by-vision).

### 3.2.1 AR training versus traditional training

Numerous studies have tried to demonstrate the benefits of AR compared to traditional training procedures. Most often, such studies compare AR approaches with paper-based instructions by measuring the time required to complete a task, the number of committed errors and the perceived mental workload reported by the participants (Jasche et al., 2021). (Azuma, 1997) shows that superimposing 3D animated drawing can ease the assembly processes compared to traditional user manuals. (Vanneste et al., 2020) compare the effects of verbal, paper-based, and AR instructions on the assembly performance and show that AR outperforms traditional guidance methods in terms of productivity, quality, stress, help-seeking behavior, perceived task complexity, effort, and frustration. The field study on AR-assisted assembly reported in (Koumaditis et al., 2019) indicates improvements in physical/temporal demands, effort, and task completion time. (Smith et al., 2021) study the effects of a mobile AR fault diagnosis application on the performance of novice workers compared to a group of experts with no AR support. The study reports significantly better performance in the AR group in terms of assembly time, accuracy, and cognitive load. (Polvi et al., 2018) compare the effects of AR guidance versus 2D images in an inspection use case and report as well significant improvements in completion time, error rate and cognitive load. (Tang et al., 2003) demonstrate that AR significantly improves the performance and relieve the mental workload on assembly tasks, in comparison to printed manuals or images displayed on LCD or HMD. Quite a few other studies have evaluated the effectiveness of conveying instructions via AR (see-through or projected), demonstrating benefits when compared to traditional ways of training or guidance such as paper instructions, desktop monitors or tablet displays (Loch et al., 2016), (S. Henderson & Feiner, 2011), (Ceruti et al., 2019), (Lai et al., 2020). Other studies suggested that AR leads to better completion times only when participants are required to execute difficult, complex assembly tasks (Blattgerste et al., 2017), (Funk et al., 2016), (Radkowski et al., 2015), (Smith et al., 2020).

The literature reveals however a few contradictions and drawbacks when comparing AR-based training approaches with traditional ones, particularly when it comes to industrial sectors. Thus, the studies reported in (Funk et al., 2016), (Werrlich et al., 2019), (Gavish et al., 2015) show that, in the considered uses cases and application conditions, classical training methods still perform better than AR in some respects and notably in terms of task completion time (Gavish et al., 2015). The comparative study between paper-based and head-mounted AR instructions conducted in (Werrlich et al., 2019) reports improvements in the error rate, but significant longer assembly completion times when using AR.

### 3.2.2 Use cases in manufacturing sectors

Industrial studies have investigated the suitability and effectiveness of AR applications for manual assembly and maintenance for more than two decades. As identified by (Westerfield et al., 2015), earlier research in utilizing AR technology for training has largely involved procedural tasks, where the user follows visual cues to perform a series of steps, with the purpose of maximizing user's efficiency.

Generally, the literature demonstrates the effectiveness of AR technology in improving performance and training time on various manufacturing tasks including assembly (Vanneste et al., 2020), (Lai et al., 2020), (Arbeláez et al., 2019), maintenance (Erkoyuncu et al., 2017), (Siew et al., 2019), and inspection (Urbas et al., 2019), (Polvi et al., 2018), (Murithi & Lin, 2020). One of the first AR industrial exploration studies has been conducted by (Caudell & Mizell, 1992), within the context of aircraft manufacturing. Authors developed an industrial AR application dedicated to the assembly of aircraft wire bundles. The information visualization in AR is based on wireframe graphics displaying the path of the cable to be added to the bundle.

(Reiners et al., 1999) conducted another early investigation that involved the use of AR to assist a car door lock assembly. The proposed system employed 3D CAD models, namely the ones of the car door and the internal locking mechanism, for guiding users through the assembly process in a procedural, step-by-step manner. Voice commands are here used to move from one assembly step to the next. However, the implementation was not stable enough for novice users to gain tangible benefit from the proposed AR system in comparison with traditional guidance.

Such early research works led to the formation of several research groups dedicated to exploring the use of AR in industry. One of these groups, so-called ARVIKA, have found that the use of AR in industrial contexts can be beneficial, as the expensive nature of AR systems is often offset by the reduced development time and product quality improvement. One example is represented by a use case where design engineers were able to rapidly evaluate ergonomic aspects of different aircraft cockpit prototypes, by overlaying virtual layout elements over real cockpit mockups, leading to significant improvement in the design process (Friedrich, 2002). A similar research group, called STAR, proposed an AR system that allows a technician to capture and transmit real-time video of his working environment to an off-site specialist (Raczynski & Gussmann, 2004). The specialist annotates the video with drawing and text, visual elements that appear then spatially registered in the worker's augmented view. This was one of the first AR remote collaboration proposals, which turned out to be an effective mean of communicating physical procedures by an assembly expert that is off-site, to one or multiple trainees that are on-site, eventually in different locations.

(Fournier-Viger et al., 2009) proposed an AR application to support military mechanics conducting routine maintenance tasks inside an armored vehicle turret. The field experiment reported that AR allowed users to locate components 56% faster than when using traditional head-up displays and 47% faster than when using monitors.

Numerous recent studies like the European research projects STARMATE (Schwald et al., 2001), (Schwald & de Laval, 2003), SKILLS (Webel et al., 2013) and SYMBIO-TIC (Syberfeldt et al., 2015),

(Holm et al., 2017), and companies such as Honeywell (Marr, 2018), Porsche (Palladino, 2017), and Mercedes-Benz (O'Donnell, 2018) also reported up to 50% improvement in production time, with over 80% reduction in the error rate by using AR-based training solutions. Industrial reports show that AR has started to be adopted as a novel experimental training technology for faster training and upskilling of manufacturing workers on complex tasks with the potential to reduce new hire training time by up to 50% (Conway, 2019).

### 3.2.3 Non-industrial AR training research works

The non-industrial applications are most often performed in controlled environments, under laboratory settings. One of the first studies in this area has been conducted almost two decades ago by (Tang et al., 2003). It compares assembly operations of toy blocks with four different instructional modes including traditional printed manual, instructions displayed on an LCD monitor, static instructions displayed via a see-through HMD and spatially registered 3D models displayed as well in an optical HMD device. The experimental results demonstrated that AR can significantly (82% reduction in the error rate) improve the performance and relieve mental workload on assembly tasks, in comparison to printed manuals or images displayed on monitors or in HMD. The study shows that the AR approach is particularly useful for diminishing cumulative errors, i.e., errors resulting from previous assembly mistakes. However, AR does not appear to have a statistically significant time advantage compared to the other considered approaches. A similar study conducted by (Robertson et al., 2008), have found that participants assembled toy blocks more quickly when using 3D registered AR content in comparison with 2D non-registered AR and graphics displayed on a HUD.

The study in (Baird & Barfield, 1999) involve various components to be assembled on a computer motherboard. To execute the required assembly tasks, participants use different instructional media including printed material, slides on a computer monitor, and screen-fixed text on opaque and see-through HMDs. The HMD display reported significantly faster assembly times and lower error rates when compared to the other methods. These findings are also supported by research works in various domains including furniture (Zauner et al., 2003) and medical assembly (Nilsson & Johansson, 2007).

The potential benefits of AR have been illustrated in numerous recent research works as well, including proposals and comparisons between innovative visualization methods and systems like spatial augmented reality. The study conducted in (Funk et al., 2016), compares four visualization methods including paper, smart glass (HMD), smartphone (HDD), and *in-situ* projection (SAR). The evaluation experiment consisted in executing a series of Lego Duplo tasks. The results of the study reported no significant difference regarding the completion time between *in-situ* projections and paper instructions. For the HMD and HDD a significantly slower time was reported compared to the paper instructions. Regarding the error rate, more errors were reported by HMD participants compared to those using HDD devices and SAR. Finally, for the cognitive load, SAR was perceived to be the lowest, while HMD was the highest.

(Blattgerste et al., 2018) investigated different *in-situ* instructions for assembly tasks by comparing four visualization types including 3D *in-situ*, 2D *in-situ*, 3D wire, and side by side. The



experiment involved 24 participants and consisted as well in solving a standardized LEGO Duplo sequence. A faster completion time and lower error rate have been reported by participants using 3D *in-situ* visualizations, but no significant difference in terms of perceived mental workload. In another study conducted in (Blattgerste et al., 2017), AR devices including Microsoft HoloLens, Epson Moverio BT-200, and smartphone have been compared to paper-based instructions. Here again, as in (Funk et al., 2016), the paper-based instructions reported the fastest task completion time, while the Microsoft HoloLens reported the lowest error rate, at the price of a significantly higher cognitive load.

In a very similar evaluation environment (i.e., LEGO Duplo tasks), (Smith et al., 2020) analyze the effect of different interaction modalities (i.e., touch and voice) and visualization modes (i.e., 3D model, text annotation, and *in-situ* video) regarding task completion, error rate and perceived mental workload. The *in-situ* video leads to the fastest completion time, followed by world registered CAD models and finally text annotations. *In-situ* videos reported the lowest mental workload as well while text annotations the highest. The method by which the participants interacted with the AR content (i.e., touch versus voice) had little to no effect on the task performance. Authors concluded that *in-situ* videos presented in AR represent a very effective way to convey procedural instructions for assembly tasks.

(Radkowski et al., 2015) investigated the effect of abstract (e.g., lines, colors, shapes) versus concrete visual features of 3D content for different degrees of difficulty in manual assembly tasks. The study involved 33 participants and evaluated abstract and concrete *versus* paper-based visualization. The experiment consisted in assembling a mechanical axial piston engine with a total of 16 manual assembly process steps, among which two were rated with a high degree of difficulty while the others with a low degree. The experimental results indicate that the abstract visualization leads to two times longer completion times and higher error rates, when compared to concrete visualizations. The concrete AR visualization and paper-based instructions present equivalent completion times. The study suggests that concrete visualization is more suitable for relatively simple tasks.

(G. A. Lee & Hoff, 2020) proposed a technique to enhance the use of instructional videos, based on the observation that video recording and sharing has become easy and popular. However, as video clips are limited to two-dimensional representation of the task space, it is hard for the viewer to match the objects in the video to those in the real world. Following the instructions in the demonstration video becomes therefore difficult, especially when the task involves navigation in a physical space. To overcome this problem, the authors propose to augment the task instruction videos with AR visualization of spatial cues. Thus, virtual world-registered circles indicate the physical position of the task in the real world and screen-registered arrows indicate the view focus of the video. The experimental results show that video with AR spatial cues help participants to better understand where the instructions are referring to, and thus reduce mental effort while improving the task performance.

The approach proposed in (Westerfield et al., 2015) investigates the combination of AR with an Intelligent Tutoring Systems (ITS) – a system that provides customized instructions and feedback

to learners. The proposed system assists the training procedure by combining AR graphics with adaptive guidance from the ITS. The system improves the task performance by 30% when compared to the same AR training system without intelligent support. The authors found AR as the ideal tool for situations which require objects manipulation such as manual assembly.

### 3.3 AR adoption in the industry

We observe that, despite the exponential progress that AR has experienced in recent years, no significant breakthrough can be noted in industrial sectors, due to various significant organizational and technical challenges (Martinetti et al., 2019), (Masood & Egger, 2019). AR systems have been mostly designed and evaluated in controlled environments, under laboratory settings, as recent surveys show. Thus, in (Dey et al., 2018) authors found in an AR usability study that only 54 out of 369 are field studies. (Merino et al., 2020) note as well that most user studies reported on MR/AR are conducted in laboratory settings. (Egger & Masood, 2020) identified that only 30 out of 291 papers with a focus on AR in manufacturing have an industrial context and call future research to focus on AR in practice. Furthermore, it seems that requirements identified in the academic world differ from the ones identified in the industrial context (Masood & Egger, 2019). (de Souza Cardoso et al., 2020) identify that a considerable number of developed applications (48%) are not focused on specific industry, but rather developed for a general segment and that 97% of the studied AR applications are tested to ensure their viability and compare AR pros and cons against traditional methods, while only 5% are implemented in production. (Palmarini et al., 2018) claim that AR technology is not sufficiently mature for complying with strong industrial requirements such as robustness and reliability. Another recent study (Masood & Egger, 2019) identifies and classifies AR challenges into three main categories: technology (e.g., tracking/registration, authoring, UI, ergonomics, processing speed), organization (e.g., user acceptance, privacy, costs), and environment (e.g., industry standards for AR, employment protection, external support). Authors point out a significant gap between the challenges identified in the academic world versus the industry-related ones, highlighting the need of more field studies.

(Kim et al., 2018) define a set of relevant AR-related research topics that should be considered when developing an AR system for the industry, including interaction techniques, user interfaces (UI), AR authoring, visualization, multimodal AR, AR applications and evaluation. When it comes to AR for manufacturing industries, (X. W. S. K. Ong & Nee, 2016) reveal challenges like time-consuming authoring procedures and appropriate guidance for complex, multi-step assembly tasks. Further, optimal ways for conveying work instructions in Industrial Augmented Reality (IAR) is still an open topic of research (Gattullo, Evangelista, et al., 2020) that should be considered as well.

Industrial AR surveys show that most AR solutions are designed outside the context of their expected usage, without the direct involvement of the end users. Consequently, such systems fail to answer effectiveness and usability requirements imposed by real world use cases, explaining the low adoption rate of AR solutions in industrial sectors.

(Moghaddam et al., 2021) support this hypothesis as well, by claiming that several AR fundamental questions need to be addressed, including best modes of task information delivery through AR, task-specific effectiveness of AR versus traditional assistance, and identifying potentials and pitfalls of using AR technology as assistive tool in industrial sectors. The authors observe that the state-of-the-art in industrial AR is focused on AR content creation and authoring (Erkoyuncu et al., 2017), (Zhu et al., 2013), (Angrisani et al., 2020), (Gattullo et al., 2019), (T. Wang et al., 2020), (Cao et al., 2019), object tracking and registration (Lai et al., 2020), (Y. Wang et al., 2018), (X. Wang et al., 2016), effectiveness of various modes of AR (e.g., head-mounted, hand-held, projector, haptic) (Vanneste et al., 2020), (Polvi et al., 2018), (Blattgerste et al., 2017), (Danielsson et al., 2018), and AR for remote assistance (Mourtzis et al., 2017). Further, they argue that to support manufacturing workers in performing tasks that involve both complex manipulation and reasoning, AR requires addressing three fundamental questions:

1. What is the most effective way of conveying assembly task information to the worker? In other words, how information conveyance impacts efficiency, error rate, learning, independence, and cognitive load?
2. What are the affordances of AR as a training tool prior to task performance versus an assistive tool during task performance?
3. How can future AR technologies replace the passive conveyance of task information with an intelligent and proactive teaming with the worker?

Overall, the analysis of the literature shows an important number of challenges that AR systems must overcome to meet industrial requirements and favor their adoption.

### 3.4 Conclusion

The potential of AR training and guidance has been evaluated in numerous research works and domains, in real world industrial use cases and laboratory settings as well. The findings reported in the literature regarding the benefit of using AR over traditional training methods are positive, even though some are contradictory. It has been consistently demonstrated that AR-based training leads reduce errors and allow a better understanding of the procedural assembly tasks, while decreasing the perceived mental workload. Literature indicates as well that AR generally improves assembly times. However, this evaluation metric seems to be highly dependent on the considered use cases. In addition, we note that most of the proposed AR training systems do not consider content authoring challenges. Most of the time, studies and experiments are conducted to demonstrate the benefits of AR training, while the authoring efforts to create the corresponding AR work instructions, including the required expertise, time and resources are almost never evaluated simultaneously with the training proposal.

We observe as well that most research papers, including recent industry-focused ones, either compare AR training systems with traditional methods such as paper-based and mentor-based instructions (Werrlich et al., 2019), or focus on comparing the AR instructions conveyance by using different hardware devices (projection-based AR, hand-held devices, or HMDs). Only few studies (Jasche et al., 2021) compare different visualizations approaches by using the same AR hardware, which makes it difficult to identify the optimal information conveyance and

visualization techniques of manual assembly. Because of its significance to this work, we try to address this research gap in a dedicated state of the art chapter (Chapter 4).

Another issue of crucial important concerns the ergonomic and design aspects of the AR user interface. (S. K. Ong et al., 2008) suggest that an efficient and suitable user interface that can be conveniently used to interact with the augmented manufacturing environment should be provided. (Nee et al., 2012) discuss the importance of the design phase of an AR application, such as the development of highly interactive and user-friendly interfaces and providing valuable in-sight to make AR an interesting tool in the manufacturing and engineering field. Swan and Gabbard (Swan & Gabbard, 2005) argue that there is a strong need to further develop AR interfaces and systems based on a user-centric perspective.

Regarding the AR authoring, numerous methodologies and tools have been proposed in the literature in the last decade. However, their adoption in industrial sectors is extremely low. We observe that the AR authoring tools proposed before the release of modern AR HMDs like Hololens, are limited by hardware capabilities. For this reason, the proposed hardware configurations dedicated to immersive AR authoring are complex, hard to replicate and evaluate. Other authoring techniques require the use of PC software and specific hardware (e.g., from sensors to AR devices), together with technical expertise (and often programming skills), for creating relevant AR work instructions. We observe that many of the recently proposed AR authoring workflows require data preparation and training (i.e., machine learning), hardware setup/calibration and multiple authoring stages. The complexity and the expertise required to develop and operate such systems potentially explain the lack of comparative evaluations between these tools in the literature.

Further, we observe that manual authoring proposals not only require specific expertise as well but is usually a very time-consuming task (Mohr et al., 2015). Automatic authoring generally requires user input (e.g., system calibration, post-processing, etc.) and a set of prerequisites including existing data like printed documents and CAD models, controlled environments, and dedicated data processing tools. Finally, fully automatic authoring, supposed to be the easiest and require the less effort, is not effective, neither suitable in view of the complex tasks (Fernández del Amo et al., 2018). In addition, the lack of field experiments and in-depth user feedback, questions the usability and effectiveness of these systems in industrial setups.

Finally, we observe that the AR authoring approaches proposed in the literature generally do not address the concerns identified during our case study, including the most relevant ones formulated in the literature (e.g., simple, reliable, efficient, adaptable, non-technical, etc.). However, we note that challenges like content authoring, information representation and visualization, user interface and interactions are becoming more prominent in recent AR surveys, potentially suggesting that state-of-the-art AR devices have reached a certain maturity level that will lead the AR research more towards content- and human-related topics, rather than technological. We believe therefore that more importance should be put on defining guidelines and elaborating AR applications that target specific use cases instead of general solutions that generally fail in real-world use cases.

# Chapter 4. AN ANALYSIS ON INFORMATION REPRESENTATION AND CONVEYANCE IN AR AND HCI CONSIDERATIONS

---

*First-person view video is potentially the most appropriate approach for capturing manual assembly expertise by shop floor experts and convey it to novice workers.*

---

---

This chapter addresses an open issue of research, which concerns the optimal way of conveying manual work instructions via AR. We propose an in-depth analysis regarding AR information conveyance for guidance and training purposes, with a particular focus on procedural assembly operations. The first part of the chapter is dedicated to the analysis of digital assets with respect to their corresponding ability and strength to represent information. Here, we take into consideration both cognitive implications and authoring costs. The second part of the chapter assesses the suitability of the most relevant visual information conveyance media in AR, for the considered assembly context. Finally, we investigate human-computer interactions (HCI) paradigms, aiming to identify optimal ways of addressing technological and human challenges in the design of an AR training system dedicated to the shop floor environment.

We present this state-of-the-art analysis as a separate chapter because it deals with multiple individual topics of major interest to training and information representation in general, not necessarily related to AR. Therefore, we expect that the research methodology adopted in this chapter, including most of the findings, can be useful and eventually applied to other works dealing with information representation and knowledge transfer by using other means than AR.

---

---

## 4.1 Introduction

Work instructions in AR can be presented in different ways, from descriptive form like text or audio-based, to depictive form like pictures, video, 2D/3D animations, or combination of them. The type of technology used to convey the assembly information differs between instructions and impacts the active cognitive processes of workers. (Moghaddam et al., 2021) show that the effectiveness of different modes of AR information conveyance and their impact on task performance in real world manufacturing use cases remains an open research topic. (Gattullo, Evangelista, et al., 2020) claim as well that more studies are needed to identify optimal ways for conveying AR instructions in industrial sectors. It has been observed in a recent study that only few research works compare different AR visualization types by using the same AR hardware device (Jasche et al., 2021). (W. Li et al., 2019) conclude that future work should examine the appropriateness of AR guidance for more complex assembly tasks and provide a robust, hands-free interaction. Similarly, (Jasche et al., 2021) claim that regardless of the chosen hardware and its advantages in specific applications (e.g., HMDs for hands-free applications, HHDs for cost-effective and easy-to-use cases, or projector technologies for static and group settings), a crucial factor is the actual visualization and presentation of AR content in relation to the complex tasks and knowledge-intensive work.

An important aspect that needs be considered when designing a conveyance information method is the cognitive style (also referred to as learning style or learning preference) of the end user, which can be either visualizer or verbalizer (Koć-Januchta et al., 2017). (Moghaddam et al., 2021) claim that “*one-size-fits-all*” information conveyance of assembly expertise via AR should be replaced with an intelligent system that dynamically adapts the delivery of information based on user experience and preferences. Previous research highlights the necessity of such mechanisms that align AR instructions with the learner attention and cognitive processes to optimize the learning experience (Bujak et al., 2013), (Zydney & Warner, 2016), (Ibáñez & Delgado-Kloos, 2018). When creating work instructions in general, arranging the information in a way that suits human cognitive processes, including perception, through vision and hearing, memory, and attention, is important (Rasmussen, 1983). The cognitive load is either determined by the nature of the material being learned and the expertise of the learner (intrinsic cognitive load) or by the instructional design, organization, and presentation of information (extraneous and germane cognitive load) (Holsanova et al., 2009). This theory assumes as well that information should be structured to eliminate any avoidable load on the working memory to enhance learning (Kalyuga et al., 1998). The worker can perceive differences in cognitive load (intrinsic, germane, and extraneous load) affecting therefore the assembly performance (Mattsson et al., 2016), depending on how the work instructions were created (R. C. Clark et al., 2006). As identified by (Mattsson et al., 2016), during the creation of the work instructions, the designers should follow guidelines to support shop floor workers during their assembly training from a cognitive perspective as well. Their evaluation results indicate that, when considering cognitive aspects of the target audience, the authors of the instructions focus on pictures to a higher extent, while globally simplifying the presentation.

To address assembly task information representation in AR in a way that addresses optimally the considered manufacturing process, we acknowledge that further analysis must be conducted. As the literature did not provide sufficient information regarding this aspect in the AR domain, we have broadened our analysis to other research fields particularly concerned with topics like information conveyance, visual modalities, learning and memory.

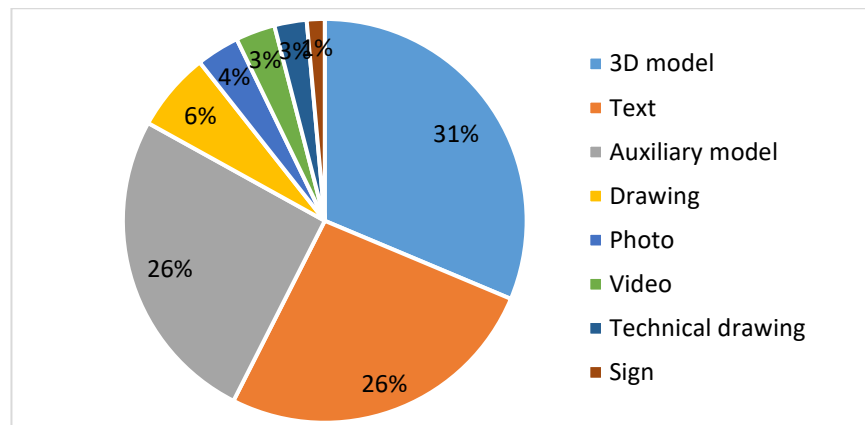
## 4.2 Digital assets

In this section, we present an analysis and classification of various multimedia data types, in terms of their ability to capture and convey information, with a focus on procedural assembly operations. We address media content effectiveness and usability by considering the relations between information conveyance methods and aspects like cognitive load, working memory and transient information effect. As identified in a similar recent study (Gattullo, Evangelista, et al., 2020), exploring the effectiveness of different modes of information conveyance through AR, and their impact on the workers' ability to complete manufacturing tasks faster and with fewer errors, is a fundamental question that should be addressed when designing and prototyping user-centered AR applications for industrial sectors.

### 4.2.1 General classification

Multimedia data consists of alphanumeric, graphics, image, animation, video, and audio objects (Grosky, 2003). Digital assets used to convey information in AR include text, audio, static (time-independent) 2D/3D and dynamic (time-dependent) 2D/3D (Palmarini et al., 2018) content. The visual assets are classified as text, sign/symbol, image/picture, video, drawing, CAD, and animations (W. Li et al., 2019). Traditionally, assembly and disassembly procedures are described by textual descriptions, often combined with a series of images for illustrating motions (Mijksenaar & Westendorp, 1999). As complex manipulations are difficult to represent with images, static illustrations are often complemented with instructional videos (Mohr et al., 2015).

A recent systematic review (Gattullo, Evangelista, et al., 2020) on information conveyance in industrial AR applications proposes a classification for visual assets according to *what* is displayed (e.g., text, sign, photo, video), *how* it conveys information (e.g., frame of reference, color coding, animation), and *why* it is used (e.g., operating, checking, locating). Authors show that product CAD 3D models are used in 31% of the cases, followed by text and auxiliary models (i.e., generally 2D and 3D annotations such as circles or arrows) with 26% each, while the rest of the assets is used in less than 6% of the cases (Figure 4-1).



**Figure 4-1. Visual assets used in industrial augmented reality (Gattullo, Evangelista, et al., 2020).**

The study also demonstrates that the number of visual assets used in maintenance is much higher (49%) than in assembly (37%) or training (14%). For maintenance, the most used visual asset is text (30%), for assembly the product model (38%) and for training the auxiliary model (33%). The study reveals as well that the frame of reference for product models is most of the time (92% of cases) world-fixed (WF) - i.e., the AR graphics are rendered such that they are perceived at specific locations in the real world. On the other hand, for drawings, photos, and videos it is screen-fixed (SF) - i.e., the AR graphics are rendered at a fixed location on the display, not spatially anchored to the real world. Color coding is used in only 14% of the cases, while animations, mostly found in product and auxiliary models, are scarcely used as well. Furthermore, they observe that identifying real world positions or objects is mainly indicated by using spatially registered, static, product models or highlighting the corresponding space region by using auxiliary models. The authors remark that operating tasks are most often (27% of cases) indicated by WF animated product models. They recommend this method by claiming that animations provide a powerful preview of the steps to accomplish for a given task.

The authors also discuss the need of using different visual assets for representing assembly operations depending on their difficulty. However, they do not provide guidelines or a methodology for choosing the most adapted visual asset based on the assembly type and complexity. It is suggested that *designers should evaluate if the use of a product model would provide more information than an auxiliary model, considering that the formers require a higher authoring effort [...] and precise alignment with real products*. This statement represents a relevant concern identified during our case study as well.

The study also indicates that the most used visual assets for conveying information via AR are product models, followed by text and then auxiliary models. The video material seems to be used much less frequently. The authors however observe that the effectiveness of product models in IAR still needs to be proved. This remark is also stated in (Radkowski et al., 2015), where authors claim that users need more time to understand complex 3D models. Let us also note a parallel literature review (Palmarini et al., 2018), which identifies that from the visualization perspective, the most common method for conveying information in AR is based on dynamic 2D/3D content (40%), followed by static 2D/3D content (26%) and text (26%).



## 4.2.2 Media guidance strength analysis and cognitive considerations

### 4.2.2.1 Literature review

In general, AR content with high guidance strength consume more time and increase information redundancy but provide more details (Siew et al., 2019), (Gattullo et al., 2019). The so-called *redundancy effect* occurs under conditions in which multiple sources of information, each intelligible individually, provide similar information but in a different form. The redundancy effect causes an increased and unnecessary working memory load, since the learners must split their attention between multiple sources of information (Kalyuga et al., 1998).

In a research work close to our objectives, (Geng et al., 2020) propose to classify media types by their corresponding *guidance strength* (Table 4-1).

**Table 4-1. Media types and their guidance strength.**

Type	Guidance Strength	Strength Level	Suitable Scenes
Text	3	Expert	Simple action or non-visual
Audio	5		Not noisy environment, simple action
Picture	6	Intermediate	Single-step action
Static annotation	7		Position, direction, or object
Static 3D model	7		Augmented position, direction, or object
Video	8	Novice	Complex or multi-step action
Dynamic annotation	9		Emphasized position, direction, or object
Dynamic 3D model	9		Action with complex path or multi-step

By considering the reported guidance strength assessment of each media type, we observe that the most adapted media content for conveying assembly expertise to novice workers include video, dynamic annotation, and 3D animations. To fully describe the assembly operations and ensure an effective guidance, notably for tasks requiring complex manipulation, a combination of multiple media types is more common. The study however does not discuss guidelines and strategies for representing the assembly expertise by using such media assets, neither provides information regarding how the guidance strength values are obtained and how they relate to the learning performance.

(Gattullo et al., 2019) provide a comparison of visual elements for AR technical documentation, summarized in Table 4-2. The values presented here have been obtained following a subjective analysis on a specific use-case. We argue that simple 2D graphics and symbols do not have a high level of intuitiveness in an assembly environment, except perhaps for simple and common operations. Our claim is based on the findings of our field studies discussed in Section 2.3, indicating that even the existing paper instructions which are generally comprised of textual description, reference numbers and images with annotations, are not sufficient for conveying the assembly information to novice workers. We observe that icons are only used together with a short textual description for describing security instructions, but never to convey assembly information. This hypothesis is supported as well by the user evaluation conducted in the considered study, as the users remarked *the crucial importance of knowing the meaning of the*

*symbols*, which requires a training session prior to the actual evaluation. Oddly enough, the intuitiveness value of this media type is reported as medium-high. Regarding the availability of the considered media assets, we argue that a general classification cannot be provided, considering the numerous variables that define manufacturing environments and their informatic systems.

**Table 4-2. Visual assets suitability classification for information representation in AR.**

Media type	Intuitiveness	Availability	Authoring difficulty	Updating difficulty	Is standard	Eligibility for AR deployment
Simple 2D graphics	High	Always	Low	Low	No	High
Icons and symbols	Medium	Always, limited	Low	Low	No	High
Multimedia elements	High	Not always	Medium	High	No	Low (occlusion)
2D technical drawings and illustrations	Medium	Not always	High	Low	Yes (ISO EN 128-20)	Low (occlusion)
3D models	High	Not always	High	Low	No	Medium (world registration)

Regarding the eligibility of the media types for conveying information via AR, this issue remains questionable because the values reported in Table 4-2 are not obtained from a field experiment or an objective evaluation. We observe as well that the media type classification is not in line with other studies, e.g., (Geng et al., 2020), (Gattullo, Evangelista, et al., 2020), as images and videos are treated under the same category, while 2D graphics, icons, symbols and 2D technical drawings/illustrations are analyzed separately (Table 4-2).

A summary of another relevant study regarding optimal visualization methods to display technical information in AR, reported in (Gattullo, Scurati, et al., 2020), is presented in Table 4-3.

**Table 4-3. Visual assets characteristics regarding optimal ways of information representation in AR. Red dot = not appropriate, green dot = appropriate, blank cells = depends on multiple factors.**

	Occlusion	Registration	Communication time	Authoring	Associated information	Computational cost
Text	●	●			●	●
Sign	●	●	●	●	●	●
Photograph	●	●		●	●	
Video	●	●		●	●	●
Drawing	●	●	●		●	
Technical drawing	●	●	●	●	●	●
Product model	●	●	●	●	●	●
Auxiliary model	●		●	●	●	●

Following a heuristic evaluation, the visual assets presented are characterized from multiple perspectives, including occlusion, registration, communication time, authoring difficulty, associated information, and computational cost. A red dot suggests that the asset is not

appropriate for the considered concern, a green dot indicates the opposite, while a blank cell suggests that the choice depends on multiple factors. The authors point out that signs, auxiliary models, and text are the only visual assets that present solely one critical issue. Therefore, they are potentially the most suitable for communicating information via AR.

Our informal experiments conducted during the case study reveal some contradictory findings. Regarding occlusion, we found that only visual assets superimposed on the corresponding real-world object, generally product models, represents a concern. Regarding the communication time, except video demonstrations and animations, none seem to be effective for conveying complex assembly information. The authoring difficulty for text and multimedia assets like photos and videos is low, while indeed the authoring of 3D content is expensive. Finally, it is unclear why video was considered inappropriate from a computational cost perspective, especially for state-of-the-art AR devices.

In a general manner, we find these studies extremely valuable and relevant to our research as they identify a set of variables that should be considered when using visual assets for conveying information via AR, including intuitiveness, availability, authoring and updating difficulty, occlusion, registration, communication time, computational cost and more generally, eligibility for AR.

#### 4.2.2.2 *Text versus visuals: the picture superiority effect*

*“A picture is worth a thousand words”* is a well-known expression, whose origin goes back and paraphrases a thought that Henrik Ibsen expressed in the 19<sup>th</sup> century: *“A thousand words leave not the same deep impression as does a single deed”*. The numerous research studies reported in the literature (W. Li et al., 2019), (Korhonen & Wallenius, 2020), (Defeyter et al., 2009), (Whitehouse et al., 2006), demonstrate that indeed pictures and images are much more likely to be remembered than words. This phenomenon is known as the *picture superiority effect*, established by (Paivio & Csapo, 1973).

A recent study shows that people remember 3 days later 10% of what they hear, 20% of what they read, 65% when a picture is added and 80% of what they see and do (Korhonen & Wallenius, 2020). The same study claims as well that visuals are processed 60.000 times faster than text, that 90% of the information transmitted to the brain is visual and that 65% of the population is represented by visual learners - i.e., learners that prefer to receive information via visual information. Research shows as well that informative images are better remembered than decorative ones (Harp & Mayer, 1997) and concrete words are better remembered than abstract ones (Reed, 2013). Furthermore, as we store visual and verbal memories separately, the best recall is when we can access one or the other (Kosslyn, 2007). This suggests that words and pictures should be used together to facilitate learning. (Reed, 2013) pointed out that *our cognitive abilities to comprehend, remember, reason, solve problems, and make decisions depend on a rich combination of words and images*, while (Seppänen, 2006) claimed that in the unconscious mind, the verbal and the visual are inextricably linked, therefore it makes little sense not to rely on both.

#### 4.2.2.3 *Animated vs. static content*

In a general manner, the visual content can be categorized in two families: animated (e.g., video, 3D animation) and static (e.g., still images, text, annotations).

Some studies show that video and animations have a performance advantage over static illustrations and text annotations (Carroll & Wiebe, 2004). However, this is not always the case, potentially because of human perceptual and cognitive limitations in the processing of a changing visual situation (Tversky et al., 2002). The benefit of animated over static content however is generally that *animations present information not available in its static version, in particular the details of the micro-steps between larger steps* (Tversky et al., 2002). Even though it seems reasonable to assume that animated content facilitates learning when compared to other types of instructional media-textual explanations, with or without static visualizations (Souza & Dyson, 2008), (Spannagel et al., 2008), (Harrison, 1995), research shows that what learners prefer is not necessarily what leads to best learning results (Despotakis et al., 2007). The benefits of animated tutorials are higher during the initial acquisition or the first learning phase of a specific procedure. Numerous studies demonstrated that subjects learning with animated tutorials have a faster and a more accurate immediate performance (Spannagel et al., 2008), (Harrison, 1995), (Palmiter & Elkerton, 1993). Similarly, (Watson et al., 2008) demonstrate as well in an assembly task experiment the immediate benefit of using animated instructions over text and diagrams, notably during the initial learning or the first assembly cycle. However, given the additional cost and expertise needed to create and deliver animated tutorials (not video though), it is worthwhile asking whether static illustrations with written procedural text can provide good or even better results (Novaković et al., 2013).

(Palmiter & Elkerton, 1993) found that learners who studied animated tutorials acquire procedures in a significantly lower amount of time and with fewer errors than learners using textual instruction. However, they have more difficulty in recalling the procedure or performing a similar one after one week. This outcome, also known as the *animation deficit*, is due to the learners mimicking animated demonstrations by passively memorizing the procedure steps without fully understanding the tasks. Another explanation for the *animation deficit* is that watching an animation creates an illusion of understanding (Despotakis et al., 2007), (Betrancourt, 2012), which results in a superficial processing of the content (Ertelt et al., 2006), (Lowe, 2003).

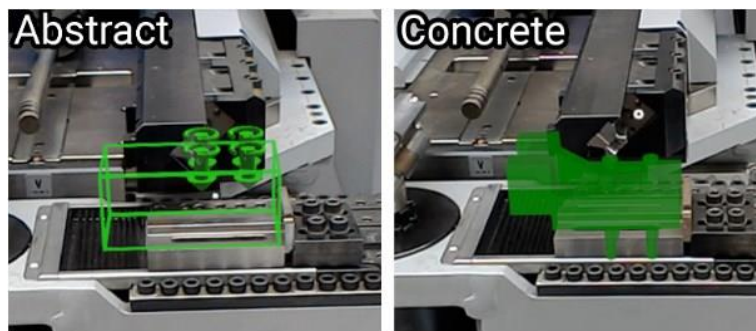
Tversky and Morrison show that stopping, starting, and replaying animated content can allow re-inspection and focusing on specific parts and actions. Zooming, alternative perspectives, and control of speed are additional features that potentially facilitate perception and comprehension (Tversky et al., 2002). Such functions may partially address the *animation deficit*, as the learner is persuaded to interact more with the animated content and eventually be more focused on processing the information. However, the study claims that animations of events may be ineffective because they are often too complex or too fast to be accurately perceived. It seems however that animated tutorials provide better results and are preferred by participants during the initial learning phase, while static illustrated content is better adapted for recalling procedures which have been already acquired (Novaković et al., 2013). For learners with higher

previous knowledge or spatial abilities, dynamic visualizations are less effective, potentially inducing extraneous cognitive load if redundant/unnecessary information is presented (Sweller et al., 2011). Similarly, (Ayres & Paas, 2007b), (Ayres & Paas, 2007a) argued that instructional animations may be ineffective because they frequently generate high levels of extraneous load because of the transitivity inherent in animations. As noted in (Marcus et al., 2013), instructional design that fails to account for the limitations of the working memory (WM) can negatively impact learning. Effective instructional design increases germane load and decreases extraneous cognitive load.

The literature indicates however that the information provided in animations makes them potentially more adapted than static content for conveying complex/difficult assembly operations, particularly for novice workers.

#### 4.2.2.4 *Abstract versus concrete*

Visual content can be also encountered in two different forms, abstract or concrete. Abstract content includes contours, lines, shapes, and colors while the concrete one is defined by CAD or product models which represent the corresponding real-world object as realistically as possible. A recent study conducted by (Jasche et al., 2021) compared abstract and concrete visualizations (Figure 4-2) of assembly instructions in AR, as well as complementing the two by video.



**Figure 4-2. Abstract versus concrete AR visualizations.**

Experimental results reported show clear advantages of using concrete rather than abstract visualizations but indicates that abstract visualizations coupled with videos leads to similar user performance as with concrete visualizations. This claim is supported by other literature findings from other domains, indicating that concrete words are better remembered than abstract ones (Reed, 2013). Other studies however contradict this theory. (Wiedenmaier et al., 2009) claimed that good manuals often use line drawings instead of photos, to reduce complexity. In the same way, assembly tasks can often be simplified by more abstract objects. (Haller, 2004) concluded that 3D models look nice for AR developers, however the actual end users often ask for abstract but familiar schemes or 2D drawings. These findings are not in line with the ones reported in (Blattgerste et al., 2018) and (Radkowski et al., 2015). It is important however to note the advantages of abstract content from an authoring perspective, as creating simple 3D models such as arrows, boxes and lines does not require modelling expertise. In addition, such simple elements can be used as generic visual cues in different use-cases.

#### 4.2.2.5 Video

Research shows that the use of videos is a practical teaching method with low production costs, wide reach, high information quality and variety of use options. In addition, the use of videos seems to be equal or more effective for acquiring advanced skills compared to traditional face-to-face instructions (Bright et al., 2015). Video tutorials are becoming increasingly prevalent as a mean of helping users perform and learn new tasks (Plaisant & Shneiderman, 2005). They have already been proposed for desktop applications (Pongnumkul et al., 2011) but also as AR tutorials that overlay 2D videos in the AR scene, communicating human motion (Kerbl et al., 2015), general task workflows related to real-world objects (Petersen & Stricker, 2012), (Damen et al., 2014) and assembly instructions (Goto et al., 2010). (Goto et al., 2010) overlay 2D videos over the task workspace using homographies, while (Petersen & Stricker, 2012) overlay automatically segmented input video directly onto the relevant real-world objects. To address occlusion, (Damen et al., 2014) are placing segmented 2D videos next to the instruction area. As discussed in (Yamaguchi et al., 2020), these approaches demonstrate the general applicability of registered videos to visualize instructions in AR. However, their effectiveness is not extensively evaluated in field experiments. Furthermore, the authors claim that the considered approaches are not applicable to complex assembly structures.

However, in contrast to static content, videos provide time-varying visual and audio information that can help users learn and understand dynamic procedures (Pongnumkul et al., 2011). The popularity of video tutorials exploded thanks to smartphones and video sharing sites, which makes it relatively easy for anyone to capture instructional or demonstrational videos and share them over the Internet. This information conveyance technique is common and well understood by most people nowadays. The video quality (visual and demonstrational) is a characteristic that impacts the understanding and assimilation of the information presented in the video. Its reliability is directly affected by both the capture and transmission systems, as well as by the author and by the environment being captured (e.g., lightning conditions, task complexity). As identified in (Tversky et al., 2002), effective graphics should conform not only to the congruence principle, but also to the apprehension principle, i.e., the structure and content of the external representation should be readily and accurately perceived and comprehended. Moreover, during the consumption of a video tutorial, users typically employ what we refer to as a *work along* strategy, where they try to perform the procedure while simultaneously playing the relevant portions of the video. While this approach allows the user to see the details of each step, it also rises several concerns. Firstly, the user has to manage the video playback and the task at the same time. Secondly, since the user typically works at a different pace from the video demonstration, he often has to switch focus between the video and the task. Finally, if the user misses or forgets a step, finding the relevant portion of the video can be difficult and time consuming. (Pongnumkul et al., 2011) analyze such limitations and propose a *Pause-and-Play* video tutorial system, which adapts and controls the video playback based on the working pace of the user. However, the proposed video playback system seems to be unsuited for deployment in an industrial context, mainly because of its complexity.

The effectiveness of the video demonstrations compared to other types of instructional material is still an open topic of research. Some evaluations of video instruction, e.g., (Harrison, 1995), (Hategekimana et al., 2008), (Palmiter & Elkerton, 1991) find little benefit of video tutorials over traditional text tutorials, even though videos provide explicit demonstration of the tasks. More recently, (Grossman & Fitzmaurice, 2010) demonstrate that short (i.e., 10-25 seconds) contextual videos that provide *a bare minimum demonstration of how to use a specific tool* are more effective in helping users accomplish tasks and retain what they learned than traditional text-based tutorials. In contrast, longer task-oriented tutorials (2-3 minute), seem to be better described with text tutorials as the video pace is not always in phase with the user's capabilities. Thus, synchronization between user actions and video demonstration often requires attention switching and video pause/play control (Grabler et al., 2009). The literature agrees however on the need of providing segmented videos that emphasize individual steps (Harrison, 1995). Existing interactive tutorial or learning systems which offer compelling alternatives to static, text-based tutorials typically require significant instrumentation of the application source code, incurring therefore heavy cost for supporting new applications (Pongnumkul et al., 2011).

The video-based tutorial technique proposed in (Langlotz et al., 2012) introduces an *in-situ* video recording/replaying approach, offering a live view of the real environment. The proposed method requires a fixed camera during the video recording, a static background, offline video processing and a GPS-equipped smartphone. The usability and applicability of the method have been evaluated by authoring a skateboard tutor. However, the system seems to be designed only for specific outdoor demonstrations.

The effectiveness of conveying assembly expertise by using *in-situ* videos is demonstrated in a recent study (Smith et al., 2020). The experimental results reported demonstrate that videos lead to significantly fewer errors than world-registered annotations, offer a slight improvement over CAD models, and a significantly faster assembly time than both annotations and CAD models. The authors conclude that video represents a very effective way to convey procedural instructions for assembly tasks. Similarly, (Jasche et al., 2021) demonstrate the effectiveness of FPV demonstration videos in an assembly field experiment, as well as a reduced cognitive load and better performance. Participants considered videos very useful and practical, giving them a feeling of safety and strengthening their confidence during the assembly execution. In addition, the study highlights the benefits of the video material as a complement to abstract visual assets, leading to a much better overall comprehension of the assembly procedure and to an error rate reduced by two. Finally, (G. A. Lee & Hoff, 2020) demonstrate the benefits of enhanced FPV task instruction videos with AR cues (i.e., arrows) for conveying task instructions. We retain the authoring advantages and the usability of such a technique, which potentially addresses some of the industrial requirements identified in our case study.

#### 4.2.2.6 Audio

Audio content can also be used to convey assembly information and potentially complement the visual guidance. Recently, (Koumaditis et al., 2020) claim that verbal assistance does not seem to lead to a significant difference in training effectiveness. In addition, using audio guidance in AR, particularly for industrial usage, raises a set of concerns. The safety rules of the

manufacturing factory may prohibit the use of headsets or other means of listening to audio instructions. In addition, noisy manufacturing environments can make the listening to audio instructions unfeasible. This issue is identified in (Syberfeldt et al., 2017), where authors recommend considering alternative user interaction or rigorous noise reduction strategies. Finally, from the UX perspective, providing a playback interface for listening to audio instructions is not in line with user-centered design principles (Section 4.4) and may have a negative impact on the user acceptance.

From a more generic cognitive load perspective, (Leahy & Sweller, 2016) demonstrate the negative impact imposed by lengthy spoken text, as all speech is transient. Thus, as the length of spoken text increases, remembering and processing the information becomes more difficult. The authors claim that the complexity of verbal information is critical and likely to overwhelm the working memory due to the transitory information effect.

Similarly, (Singh et al., 2012) show that written text outperforms spoken text for both learning and transfer tasks, mainly because the transient nature of the spoken text. We expect therefore that any conveyance technique where the information disappears or changes over time (e.g., 3D animation, video) may potentially be concerned by the transient information effect. This issue is further detailed in the following section.

#### **4.2.3 Transient information effect. Segmentation and playback control.**

The literature shows that techniques like stopping and segmenting, compensate for transient information, by reducing the amount of information that the learner must deal with at one time. This informational reduction can diminish the negative effects of transience in the case of animated instructional presentations (Marcus et al., 2013). More generally, animations that are stopped either by the user (Hasler et al., 2007) or automatically by system (Schwan & Riempp, 2004), lead to better learning outcomes than continuous presentations. Thus, providing some form of control (e.g., play/pause, sequence selection, play speed) can partially solve the transient information effect concern (Hasler et al., 2007), (Schwan & Riempp, 2004). However, this type of interactivity may be problematic, as novice workers do not necessarily know where or when to apply this mode of control. As pointed out in (Biard et al., 2018), novice learners make very little use of pauses to segment a training video, perhaps because they cannot identify the key steps of the assembly procedure. This assumption is supported in (Wouters et al., 2007). Authors claim that without prior knowledge of the underlying structure of the information being presented, novice students cannot know when to pause a video to enhance their learning. Providing novice learners with a segmented video instead makes it easier to identify key actions in the procedure. The evaluation conducted by (Biard et al., 2018) supports this hypothesis, highlighting the positive effects of video segmentation on novice learners' performances. (Bright et al., 2015) demonstrate the benefits and effectiveness of this approach as well, in a clinical psychomotor use case.

(Spanjers et al., 2011) show as well that segmented, animated content is more efficient in conveying information than continuous animated content for novice students. Segmentation successfully reduces the high cognitive load imposed by animations and leads to more efficient



learning. Dividing the animation into meaningful chunks makes it easier to grasp the underlying structure of the process or procedure being described. However, the superiority of segmentation disappears at higher levels of prior knowledge.

Numerous other studies identify as well that an effective method to address the transient information effect is to divide the animation into smaller segments (Mayer & Chandler, 2001), (Spanjers et al., 2011), (Wong et al., 2012). By considering these findings, we expect that short video sequences demonstrating unique assembly actions and for which the playback is easily controlled by the user is likely to be the most adequate technique for conveying complex manual assembly instructions, independently on the transmission medium (e.g., 2D screen, AR).

#### 4.2.4 Cueing in hand-manipulated operations

(Marcus et al., 2013) claim that learning motor skills from instructional animations may tap into our innate ability to learn by observing. The study further demonstrates that animated instructions are superior to an equivalent static presentation in learning hand manipulation tasks (e.g., tie nautical knots) and that showing hands during the demonstration poorly impacts the overall performances. However, the authors note that there are potentially more benefits if hands are included as the instructional efficiency is improved, particularly for more complex tasks where the cognitive load may be higher. (Mayer & Fiorella, 2014) support this theory by claiming that the presence of hands may produce *signaling* or *cueing*, an effective technique to reduce cognitive load and to improve learning from visualizations. The signaling principle states that people learn more deeply from a multimedia message where added cues highlight the organization of the essential material. These findings suggest that in addition to segmentation and playback control, observing movement and potentially hands may improve understanding of the process and training performance. Guiding user's attention during the interaction with the AR system by using visual cues, represents an aspect that should be considered in the design of an AR training proposal.

#### 4.2.5 Verbalizers versus visualizers and the *multimedia effect*

The verbalizer-visualizer dimension refers to the cognitive style of learning which originally derives from the dual-coding theory introduced in (Paivio, 1986). According to this theory, incoming information is processed and mentally represented in two ways: verbally (i.e., verbalizers) and visually (i.e., visualizers). In the case of verbalizer-visualizer differences, the aptitude-treatment interaction (ATI) hypothesis predicts that visualizers will achieve the best learning performance when they receive visual rather than verbal methods of instruction, whereas verbalizers will perform best when they receive verbal rather than visual methods of instruction (Massa & Mayer, 2006).

The literature is however somehow reluctant regarding the effectiveness of providing instructions designed to accommodate individual differences (Massa & Mayer, 2006). In a recent study, (Koč-Januchta et al., 2017) support the hypothesis that individual differences in visual-verbal cognitive styles do exist and can be observed in the eye movements. The study demonstrates that the cognitive style has a considerable influence on the learning behavior and therefore on the learning outcome. In addition, the difference in the spatial layout of the

learning material (i.e., text, images, or schemes) has a significant effect on the readers' behavior (Holsanova et al., 2009). Numerous studies show however that a combination of text and pictures support learning and deepens understanding and problem-solving processes (J. M. Clark & Paivio, 1991), (Mayer & Fiorella, 2014), (Wittrock, 2010), independently on the cognitive style of individuals (Massa & Mayer, 2006).

The theory according to which people learn better from words and pictures than from words alone was first introduced by (Mayer, 2001) and is known as the *multimedia effect*. However, simply combining text and pictures does not seem to always improve learning, since such a complex process is dependent on various aspects including the form of visualization, the type of learning task, the number of referential connections between text and pictures, and the personal characteristics of the learner (Mayer & Fiorella, 2014), (Schnotz & Bannert, 2003). Thus, the learning performance differs with respect to individual differences like prior knowledge (Kalyuga, 2007), spatial ability (Hegarty, 2012), (Höffler & Leutner, 2011) or cognitive style (Höffler et al., 2010). Even though visualizers and verbalizers process multimedia information differently, it seems that the best outcome is generally obtained when information from pictures and text is integrated (Koć-Januchta et al., 2017).

This hypothesis is supported in (D. Li et al., 2018), where three sets of manual assembly instructions (combinations of text and pictures), designed by following a set of Design Principles for Information Presentation (DFIP) (Mattsson et al., 2018), are analyzed. The evaluation results indicate that the perceived workload is higher for text-only instructions and that pictures decrease cognitive load and alleviates workers frustration during the assembly. From a quality of information perspective, workers prefer text and picture-based over text-only instructions.

In addition, the spatial contiguity principle states that *people learn more deeply from a multimedia message when corresponding words and pictures are presented near rather than far from each other on the page or screen* (Mayer & Fiorella, 2014). The literature shows that a shorter physical distance between text and the corresponding picture facilitates the integration of information from the two sources (Holmqvist & Andersson, 2015), (Holsanova et al., 2009). We note that even though these studies have been conducted on learning materials like newspapers, brochures, textbooks and scientific articles, the findings can provide practical guidelines for information conveyance in AR.

#### 4.2.6 User and environment-adapted guidance

(Kalyuga et al., 1998) provide strong evidence that the efficiency of an instructional design is in part attributable to the expertise of the learner, with trainees at different levels of expertise gaining optimal benefits from different formats. In line with this observation, (Palmarini et al., 2018) state that, in function of their experience, workers need different ways of guidance. For novice workers, detailed assembly descriptions such as video or 3D animations is more suitable during training as performing and learning the assembly operations correctly is more critical than speed and high operation efficiency. On the other hand, for experts, a less time-consuming description such as text or static annotation is more appropriate (Geng et al., 2020).

(Radkowski et al., 2015) show that beyond the worker level factor, the assembly environment can also affect the choice of the visual content used for conveying the information via AR. For example, audio is not adapted to noisy work environments; a physical location is better indicated by a static annotation with an arrow rather than text or video; a complex action is better described by an animation rather than a static image or text (Geng et al., 2020).

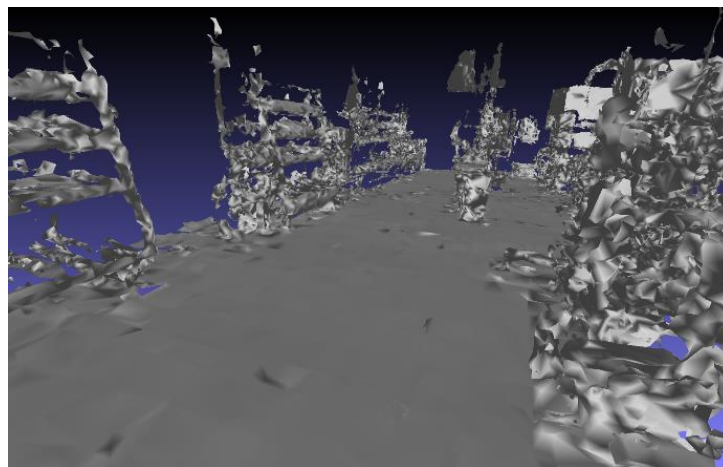
Concerning the assessment of manual assembly difficulty to provide assembly-adapted guidance in AR, related works (e.g., (Falck et al., 2016)) indicate that such procedures are cumbersome, unlikely to be adopted in industrial sectors. These findings suggest that the capturing of the assembly expertise should be as complete and comprehensible as possible, independently on the assembly complexity or environment, to ensure an effective guidance during the initial stage of training independently on the user profile and experience. Concerning the assembly time, as any form of guidance will slow down the assembly process and therefore affect the nominal time of the workstation, the conveyance of the AR instructions during training should target simplicity and speed.

#### **4.2.7 Authoring cost evaluation: a study on AR HMD-based content creation.**

To evaluate the authoring cost of different media files in AR and remain consistent with the authoring requirements discussed previously, in this section we estimate the difficulty and measure the time required for capturing different multimedia types by using Hololens 2, the most advanced state-of-the-art AR HMD device. We use an AR headset device for estimating the cost of media capture in the idea that the proposed authoring method will not require the use of other devices (e.g., smartphone, digital camera) or informatic systems (e.g., desktop PC) during the authoring procedure. The analysis proposed in this section supposedly holds for other AR devices with similar functionalities as Hololens 2, offering alphanumeric input, mixed reality capture (i.e., photo and video), voice commands, 6DoF, hand and eye tracking.

##### *4.2.7.1 3D capture and modeling of real-world environment*

Hololens 2 can capture a 3D mesh representation of the real-world environment. However, the resulted mesh is not sufficiently detailed to represent correctly and completely the assembly components and environment (Figure 4-3).



**Figure 4-3. A 3D capture of the assembly line L2 at elm.leblanc, obtained with Hololens 2.**

To make it usable, a post-production process is needed, which generally requires advanced, expensive, and laborious refinement procedures. In addition, the quality of the reconstructions is highly affected by the lightning conditions and environmental particularities. To overcome such limitations, a wide range of techniques and commercial high-precision 3D scanning systems exist (Daneshmand et al., 2018). Some of them are potentially adapted for capturing non-textured assembly components in uncontrolled lighting conditions. However, the related production costs make them poorly fitted to industrial AR training applications.

In addition, the utility and effectiveness of CAD data in AR is today questioned by authoring and manipulation costs (i.e., expertise and time). Most augmented reality techniques that make use of dense 3D CAD models implicitly assume the existence of these models (Kahn, 2013), while the authoring effort is not evaluated or even considered. Such systems are most of the time evaluated in controlled environments and their reliability/effectiveness in real world industrial environments is not proven.

For feasibility and acceptability reasons, we aim therefore to elaborate an AR training system that does not depend on CAD data. The most significant concerns we are trying to address by not using CAD data are that: (1) the overall authoring process becomes more expensive and dependent on existing assets or specific expertise and that (2) the authoring would be potentially decoupled, therefore demanding multiple experts. However, to ensure extensibility and take advantage of future technological advancements, the proposed AR training approach should support the usage of CAD data.

#### 4.2.7.2 3D animation creation

Animations (and particularly world-registered 3D animations) can provide an effective way of conveying manual task instructions. However, the difficulty of authoring such content is a well-known challenge and remains an open topic for research. The authoring systems identified in the literature are complex from a hardware perspective and generally require a set of prerequisites (e.g., CAD model with individual movable parts) and expertise that make them unsuited for the use-case considered in our work.

The instinctual interactions provided in HoloLens 2 (*Mixed Reality Instinctual Interactions* / *Microsoft Docs*, 2020) allow one to easily translate, rotate and scale CAD objects. A rather simple way of generating 3D animations would be by tracking the motion of these graphical elements, as they are manipulated by the author within the AR space. However, our informal evaluations in the factory suggest that this authoring technique is unsuited for shop floor workers due to time constraints, lack of AR expertise and intuitive user interfaces. The concerns regarding CAD data and current AR technical challenges (e.g., reliable 3D pose estimation and tracking) represent limitations that question the use of this data type for authoring AR step-by-step instructions. Similarly, as for CAD models, our proposed approach will not rely on but support the usage of 3D animations.

#### 4.2.7.3 Text insertion

Text input techniques and systems in AR/VR is a widely discussed topic. Such techniques are mainly divided as hands-free (X. Lu et al., 2020), controller-based (Boletsis & Kongsvik, 2019), (Y.

Lee & Kim, 2017), (Whitmire et al., 2017), physical or touchscreen (Knierim et al., 2018), (Otte et al., 2019) and hand-gesture (Nooruddin et al., 2020), (Dudley et al., 2018). It is hard to assess objectively which of these methods is the most adapted to manual assembly scenarios. Our case study findings strongly suggest that it is preferable to avoid equipping shop floor workers with additional hardware, in addition to the AR device, reason for which controller-based and other hardware-dependent techniques are not suitable. Informal evaluations at elm.leblanc regarding the usage of the virtual keyboard natively provided by the HoloLens 2 indicated that such technique is rather familiar and intuitive, thanks to the natural gesture interaction. Moreover, the dictation or speech-to-text technique that is also available in HoloLens 2, allows the insertion of text in a much faster and easier way than virtual keyboard typing.

To objectively evaluate the text insertion speed, we have considered two different techniques: the traditional way (keystrokes) and dictation. For the first approach, we used the virtual keyboard to write each of the instructions of one of the existing paper instructions. Our measurement reports a speed of 12.4 words per minute, a value which is lower but similar to those identified in other AR/VR text insertion studies (Boletsis & Kongsvik, 2019), (L. H. Lee et al., 2019). With an average of 10 words per instruction, this led to 48 seconds for a single instruction. The average number of hand gestures (corresponding to the number of characters per instruction) required for inserting a single instruction is 55.

Secondly, we used dictation with the same input data. Our measurement reported an accuracy of 97% at a speed of 100 words per minute. We note that, in our case, the background noise of the factory did not affect dictation. Dictation requires entering and exiting the dictation mode by using touch interaction, which requires on average 2 seconds.

#### 4.2.7.4 Photo and video/audio capture

We discuss the capturing of first-person view (FPV) photos and instructional (demonstration) videos jointly. In the following, we use the term “video” for any combination between audio and video: video-only, video + audio, audio-only. We do not deal with video and audio separately because the capturing process is the same from an authoring perspective. Capturing a FPV photo or video in AR by using a HoloLens 2 device is a straightforward process that can be triggered by the author either by interacting with the UI (e.g., pushing a button), by using a voice command or by using hardware buttons. The main difference between photo and video capture is that a photo requires a single trigger from the author, while the video requires one for starting the recording and one for stopping it.

To objectively assess the required time for capturing a photo we have used the vocal command “photo”, already implemented by our prototype. We followed the same principle as before, by taking a photo for each assembly operation found in the considered paper instructions. The reported average time of taking a photo was approximately 5 seconds. Time timespan includes saying aloud the vocal command and its interpretation by the system, 3 seconds for the countdown and the actual photo capture and displaying in the AR scene. We note that traditional interaction techniques (e.g., buttons) can be used for triggering the capture of a first-person view (FPV) picture, in addition to voice. Such technique or a combination between voice

and touch interaction could potentially be more adapted for other assembly environments. We note that the reported time does not include neither the manipulation of the photo, nor its transmission to a server for storage purposes.

To measure video recording, a similar voice command technique was used as for taking a photo, with two vocal commands (*start recording* and *stop recording*). The recording time is dependent on the author of the AR instructions and the assembly operation, therefore, the only relevant approximation that we could make was the time required for the starting and stopping the video recording. To provide an approximation, we recorded 20 videos, 9 seconds long each, by using the considered prototype. Our measurements reported that capturing a video requires approximately 6 seconds for triggering and stopping the recording, including the 3-second countdown time. We chose a 9-second video capture because this represents the average time spent by a shop floor worker for performing one assembly instruction. By considering this information, we could approximate that capturing a video demonstration for a single assembly operation requires a total of 15 seconds on average.

#### 4.2.8 Conclusions

Let us present a summary of the measured time and difficulty required to capture each of the three media types, per AR work instruction. We note that these values have been extracted by analyzing one of the existing paper instructions of the factory, which consists of 23 instructions, 232 words and 1260 characters. The nominal assembly cycle is 3 minutes and 30 seconds. We do not provide average values by considering all the workstations in the factory as the assembly cycles of each workstation is around the same value. Therefore, we expect our approximation to be valid for any workstation of the factory.

**Table 4-4. Reported time and difficulty estimations for capturing low-cost assets (i.e., text, FPV photo and FPV video) by using Hololens 2.**

Modality / trigger	Text (10 words)		Photo	Video
	Touch	Dictation	Voice command	
Difficulty (number of interactions/touches)	55	2	1	2
Time (in seconds)	48	8 = 6 (speech) + 2 (touch)	5	15

We classify, by difficulty, the capturing of low-cost assets as follows (the values are provided per instruction):

1. **Photo:** photo capture is the least difficult and the fastest, as it requires a single trigger (i.e., a voice command) and approximately 5 seconds,
2. **Video:** video capture is second, as it requires 2 triggers and 15 seconds,
3. **Text:** text insertion is considered the most difficult among the three, as it requires 55 touches and 48 seconds. By using dictation, text insertion however becomes much easier, requiring 2 touches and speaking out loud the assembly instructions. This technique requires therefore 2 triggers, voice input, and 8 seconds on average.

Using short and concise sentences for briefly describing the assembly operation and short videos which demonstrate strictly the assembly operation is recommended. In addition to HCI

considerations, this recommendation brings other potential benefits. Firstly, the author of the AR work instructions is encouraged to capture only the essential information, decreasing therefore the authoring complexity and time. Secondly, the captured information can be used as is, without requiring offline processing or refining steps. Finally, we expect that short and concise AR instructions lead to a lower cognitive load, less ambiguity, reduced error rate, and potentially to faster assembly times.

Let us also note that the characteristics of the assembly environment and operations (e.g., light conditions, visual obstruction of the assembly, difficulty of recording the interest zone of the assembly operation with a head mounted device, etc.) can impact the perceived complexity of capturing low-cost assets. For example, capturing an instructional FPV video in an inconvenient location may be considered more difficult than inserting a textual description. As a multitude of variables can play an important role in the cost and difficulty of authoring effective low-cost assets via HMD AR devices, a generic objective assessment cannot be derived. Our authoring approach tries to address this concern by being flexible and allowing the description of assembly tasks by using multiple media types.

### 4.3 Visual information conveyance mediums in AR (hardware devices)

Visual information conveyance in AR can be achieved by using multiple hardware technologies, including Head Mounted Display (HMD), Handheld Display (HHD), desktop PC and projector. A recent survey (Gattullo, Evangelista, et al., 2020) shows that HMDs and HHDs are both used in 27% of the cases, desktop PC in 30%, whereas projection and SAR concern solely 5% of them. We note that information in AR can be delivered by using audio and haptic means as well. However, such methods are not adapted for the considered research context.

#### 4.3.1 Handheld display (HHD)

HHD AR applications are generally run on a smartphone or tablet. There are several important differences between using a smartphone as AR interface compared to a desktop interface, including limited input options, screen resolution and graphics support. Even though high-resolution cameras, touch screen and gyroscope have already been embedded in recent HHD AR devices (Nee et al., 2012), to become mainstream, AR interfaces should provide interaction techniques that allow people to interact with AR content in a much more intuitive way (Billinghurst et al., 2009). HHD devices provide great mobility, are cost-effective and easy to use compared to HMD devices. However, these advantages come with significant implications from a usability perspective. Thus, they do not address the hands-free requirement, the interaction techniques remain limited and not adapted for AR environments, while the small display contribute as well to a rather poor AR user experience.

#### 4.3.2 Head mounted display (HMD)

Two types of HMD can be identified: *see-through* (or optical) HMD (e.g., HoloLens 2 - Figure 4-4) and *video display* HMD. See-through HMD is based on semi-transparent mirrors, allowing the user to see the real-world environment and, at the same time reflecting computer-generated information into the user's eyes. In video display HMD the real-world view is captured with video

cameras, and the computer-generated images are electronically combined with the video representation of the real world (Rolland et al., 1995).



**Figure 4-4. Microsoft HoloLens 2, a state-of-the-art AR see-through HMD.**

A great disadvantage of video display HMDs is the higher latency compared to the see-through HMD. We note that this technological limitation can have serious implications on the suitability of such a system in human-operated manufacturing environments, as a delayed visualization of the real-world view might represent a safety concern in many assembly use-cases. This concern is solved by optical HMDs. However, a potential disadvantage of this technology is the limited FoV, a well-known technological concern with a negative effect on the user experience. The physiological and psychological impact of wearing HMD AR devices represents another concern that can have a significant impact in the adoption of such a technology, particularly in industrial sectors. The main advantages of HMDs are portability, user experience because of spatially registered visual content (Palmarini et al., 2018), and more importantly for the considered research use-case, they address the hands-free requirement, flexibility and do not require setup or maintenance. In addition, systems using Head-Mounted Displays (HMD) mimic how people access spatial information naturally (Zhao et al., 2020).

#### **4.3.3 Desktop/monitor-based**

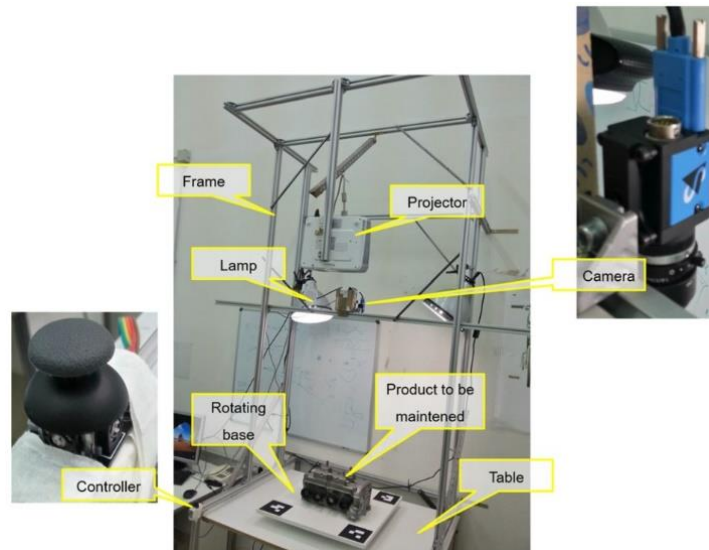
Another way of conveying AR information is a low-immersion technique through 2D monitors or screens (Gonzalez et al., 2019), (Tang et al., 2003), (Bosch et al., 2017). This method has been intensively used to demonstrate the benefits of AR compared to paper instructions because it does not require complex setup, can be easily deployed/tested and is low cost. However, this technique does not meet one of the most important AR principles, which is the user immersion. As noted in (Billinghurst et al., 2009), in an AR experience an intimate relationship between 3D virtual models and corresponding physical objects is expected. The poor expression of the spatial relationship between the virtual content and the real-world environment in a 2D medium does not take full advantage of AR, potentially making it hard for novices to understand complex assembly operations and setups. Unlike immersive AR, where users can move naturally and navigate through the content, in desktop-based AR controlling movement, orientation and interaction is provided through abstract navigation interfaces (e.g., keyboard, mouse, or joystick) (Zhao et al., 2020), leading to a user experience that is less intuitive.

#### **4.3.4 Spatial augmented reality (SAR)**

A hands-free AR information conveyance alternative is represented by SAR (Spatial Augmented Reality), where the combination of virtual and real objects is produced by projecting virtual images onto real objects using projection mapping. Hence, SAR does not require the use of



display monitors, HHDs or HMDs, eliminating as well the need of visual multimedia assets. Numerous studies assess the potential of this AR information conveyance technique (Doshi et al., 2017), (Uva et al., 2018), (Funk, Mayer, et al., 2015), (Bosch et al., 2017), and claim that it is the most well-fit for industrial constraints (Mengoni et al., 2018).



**Figure 4-5. A prototype of a SAR system proposed in (Uva et al., 2018).**

By considering the findings of our industrial case study however, SAR and other AR-projection techniques do not seem to be adapted to the considered manufacturing environment, mainly because of the initial effort required to install and calibrate such systems, their lack of mobility and adaptability, and projector's hardware limitations. The initial setup of SAR systems is generally extremely complex and composed of frames/supports/tripods, depth cameras, projectors, PC workstations, monitors, joysticks, lighting systems and markers (Mengoni et al., 2018), (Uva et al., 2018).

#### 4.3.5 Screen-based video see-through

Another alternative was proposed by (Fiorentino et al., 2014) with a screen-based video see-through Display (SBVD) technique. This approach consists in an augmented visualization on a large screen and a combination of multiple, fixed and mobile cameras. Authors claim that this technique provides a less immersive augmentation (i.e., not co-located with the user's line of sight), but without the drawbacks of current HMDs. We note however that the hardware setup is complex and rigid while the potential benefits of SBVD systems are questioned by the features proposed in the state-of-the-art HMD devices like Hololens 2.

### 4.4 Human-computer interaction (HCI) considerations

*Most sophisticated machines are worthless unless they can be used properly by men* (Karray et al., 2008). Designing an AR system dedicated to humans should rigorously consider the interfacing between the system and its end-users. The field of study in charge of this relation, so-called HCI (*human-computer interaction*), emerged in the late 1970s, about the same time when the world's first head-mounted display was proposed by (Sutherland, 1968). However, the

literature did not yet manage to provide standardized methods and guidelines to support designing effective and optimized AR systems and experiences. As identified in (Agati et al., 2020), there is an evident gap in design guidelines and methodologies in AR. We observe that HCI research for AR started to emerge in recent years. However, the findings are generally domain-specific and not well-established. The consensus of the AR community is that more research and evaluation need to be conducted, particularly in real-world use cases, to better identify and eventually formalize design and implementation guidelines for industrial AR.

Let us further investigate this concern by analyzing HCI paradigms within the framework of the considered use case.

#### 4.4.1 Usability and usefulness

A fundamental objective of human-computer interaction research is to make systems more usable, more useful, and to provide users with experiences fitting their specific background knowledge and objectives (Fischer, 2001). Usability refers to the ease of use, user-friendliness, and accessibility of the system while usefulness denotes whether a system adequately supports the tasks for which the system was intended (Karray et al., 2008). In other words, usability of a system with a certain functionality is the range and degree by which the system can be used efficiently and adequately to accomplish certain goals for certain users (Karray et al., 2008). However, both criteria are interdependent, since a system is hardly useful, if it is not usable (Landauer, 1996). The actual effectiveness of a system is achieved when there is a proper balance between the functionality and usability of a system (Nielsen, 1993). We note that original HCI approaches focused on making the systems more usable, reducing therefore their complexity and implicitly their power of expression, to accommodate novices or users without prior experience. High-functionality applications (HFA) on the other hand are represented by complex systems for professional use and prioritize usefulness rather than usability, making them more powerful but more difficult to use. As noted in (Shneiderman et al., 2018), the value of functionality is visible only when it becomes possible to be efficiently utilized by the user, suggesting that the focus should be put on prioritizing the design of a usable AR system dedicated to novices. Another consideration supporting that complex, high-functionality systems may not be adapted, is the high employee turnover rate, which is the case in various industrial sectors.

Therefore, considering our specific research context from a HCI perspective, the AR authoring system should be easy to learn and use, intuitive and effective, safe, and tolerant to errors. From a functional viewpoint, the authoring should ensure correctness and completeness and be as well practical and fast, while not requiring technical intervention.

Regarding the AR training procedure, the challenge is to display the *right* information at the *right* time and in the *right* way. The AR training should be adapted to people without AR experience and not require a learning phase. It should guide the user at any time in the least intrusive manner and without introducing safety concerns. The information conveyance via AR should optimize comprehension and learning while considering human and environmental factors.

Usability and usefulness are reflected as well in the user acceptance, identified in the literature as the most important success factors of an AR system in industry (Masood & Egger, 2019), (Masood & Egger, 2020). Recently, (Lorenz et al., 2018) suggested that workstation experts are the most suited to create the AR instructions, suggesting therefore that such a tool must be designed for shop floor workers rather than for AR experts or engineers. The usage of the system should be straightforward and avoid ambiguity. A complex system that provides non-essential functions and choices can make the user question his decisions, lead to lack of confidence, increased mental workload and time spent, and eventually impacting the acceptance rate.

Finally, it seems that generic assumptions about users may be adequate in designing the AR training system specifically for novices (Fischer, 2001). In other words, adopting HCI principles predominately adapted to inexperienced workers may not only be the best approach for the considered context, but could eventually simplify the design and implementation of the overall AR system while making it more predictable and reliable in different working environments. We anticipate that user-adapted interfaces that accommodate optimally different human traits like cognitive characteristics and learning preferences will remain an open research topic until more usable AR systems are evaluated in real world use-cases.

#### 4.4.2 User interface and interaction techniques

(Shneiderman et al., 2018) defined *eight golden rules of interface design*, which include: (1) strive for consistency, (2) seek universal usability, (3) offer informative feedback, (4) design dialogs to yield closure, (5) prevent errors, (6) permit easy reversal of actions, (7) keep users in control and finally (8) reduce short-term memory load. The authors claim that these principles can be applied in most interactive systems and provide a comprehensive summary of general UI design guidelines provided in the literature. Considering these recommendations, independently on the interface type (e.g., 2D vs. 3D), the question is how they can be optimally applied to AR interface design, given the major visualization and interaction differences between classical 2D interfaces (e.g., screen) and 3D. These generic principles must be interpreted, refined, and extended for AR, by taking into consideration the considered research context, the industrial environment, and the profiles of the potential end users.

When it comes to user interfaces and interaction in AR, users struggle to understand and perform actions in the 3D space, as the physical world with which they are familiar with contains characteristics that cannot be yet accurately represented in computer simulation (Steinberg, 2012). (Karray et al., 2008) claim that it is economically and technologically crucial to make HCI designs that provide easier, more pleasant and satisfying experience for the users. To achieve this goal the interfaces are getting more natural to use, a relevant example being represented by the interfaces of note-taking tools which evolved from typewriters to keyboards and finally to touch screens that can even recognize handwriting and change it to text. Further, the importance of adaptive HCI in which the interface incorporates at least some kind of intelligence is considered as an important factor in designing new generation of interfaces. The most common examples include speech (Oviatt et al., 2000), a natural language interaction technique, eye tracking (Gavrila, 1999) and gaze tracking (Sibert & Jacob, 2000) which are implicit

interaction techniques. Based on the input methods, devices and interaction techniques, the UIs in AR can be classified into multiple categories: information browsers, 3D interaction, tangible user interfaces (TUI), natural user interfaces (NUI), and multimodal user interfaces (Aliprantis et al., 2019), (Muhammad Nizam et al., 2018).

NUIs are intuitive, flexible, and fluid, while interacting with them requires little or no training at all (Steinberg, 2012). With the emergence of NUIs, HCI aims in evolving human-computer interactions as natural as interactions between humans by incorporating gestures, speech and implicit interactions like eye tracking and gaze (Aliprantis et al., 2019). Similarly, multimodal interfaces provide better user experience when compared to unimodal interfaces, improving the efficiency and accuracy of the interactions (Irawati et al., 2006), (Turk, 2014), while reducing learning time and error rates (Lv et al., 2014), (Heidemann et al., 2004). Multimodal interfaces in AR environments are considered as a solution to enhance interaction between physical and virtual entities (Muhammad Nizam et al., 2018), leading to better immersion and therefore to improved AR user experiences.

Our informal experiments performed with shop floor workers at elm.leblanc suggested that a simplistic user interface is optimal, considering their profile and the organization of the considered manufacturing environment. To address this requirement, we analyzed information-presentation methods (i.e., registration, media types, semi-transparent effect, and rotation) proposed as guidelines for AR assembly task support (Tainaka, 2020) and explored information access and peripheral awareness methods discussed in a study related to information access methods for HMD AR (F. Lu et al., 2020). To address our context optimally, we figured that combining and adjusting these guidelines (Tainaka, 2020) and techniques (G. A. Lee & Hoff, 2020) is necessary for designing a hands-on AR training system. A relevant example is represented by the timing, location and displaying modality of the augmentation media. When, where and how the virtual elements which describe the assembly information are displayed and how the user interacts with them is therefore essential. In addition, the user interface should be designed in a way that minimizes the potential negative impact of the AR functions which are not precise or reliable. One such example is the real-time spatial registration of 3D CAD models on a moving real-world object. A faulty or inaccurate registration leads to a poor user experience, potentially errors and even injury. It is therefore more appropriate to elaborate a system that is a reliable and predictable even if this has consequences on the system functionality.

## 4.5 Conclusions and guidelines

The numerous research works reported in the literature suggest considering jointly assembly difficulty, work experience and cognitive characteristics in designing optimal ways of describing assembly expertise in AR. The analysis conducted in this chapter in relation to the findings of Chapter 2 persuaded us to believe that such approach is unfeasible because of the numerous and diverse assembly operations, manufacturing processes and environments, as well as different cognitive characteristics and experience levels of workers, difficult to estimate and classify. We acknowledged that a realistic objective for this thesis was to propose use case

specific guidelines, in a similar way as the one proposed in (Tainaka, 2020), that addresses the three topics mentioned above.

Furthermore, the design of the AR training system should consider HCI paradigms. These suggest that system usability and usefulness are far more important than functionality and performance, particularly when considering the adoption of an emerging technology like AR in industrial sectors. Reducing system complexity and implicitly its power of expression to accommodate novices seems to be the most advantageous approach for the considered manufacturing context. We claim that a human-centered design approach should be adopted for designing an easy to use and intuitive user interface adapted to the shop floor environment. Finally, the set of functionalities proposed should be rather limited, but reliable and predictable.

### **Assembly difficulty**

Proposing a generic, complexity-based assembly representation for manual assembly operations is unfeasible. In our work, we aim at providing authoring guidelines and content description flexibility (i.e., text and photo or video demonstration), allowing the author of the AR work instructions (i.e., the shop floor expert), to decide on the way an assembly operation is described, to optimally accommodate novice workers, based on the assembly complexity but also on environmental characteristics such as light conditions, hidden parts, specific manipulation, and tools usage.

### **Work experience**

As identified during the considered case study at elm.leblanc, it is justifiable to assume that any worker is novice or unexperienced, as long as he operates an assembly workstation for the first time. The assembly representation in AR should therefore be as complete and explicit as possible, while avoiding redundancy. To accommodate knowledgeable workers or experts, the information conveyance should allow skipping certain material and offer a fast (potentially partial) reading of the AR work instructions.

### **Cognitive characteristics**

By considering the high number of cognitive characteristics (e.g. sustained and selected attention, long-term and working memory, logic and reasoning, auditory visual processing, processing speed) and other variables that affect user comprehension, learning and responding processes in general, we acknowledge that finding the most adapted AR information representation method adapted to accommodate cognitive differences and learning preferences is an open research topic that can only be partially covered in this thesis. In our opinion, most assembly operations require a visual representation (e.g., photo, video demonstration, animations) in addition to text descriptions and visual cues (e.g., arrow, trajectory line, etc.).

### **Assumptions**

Our findings suggest that first-person view video demonstration has the highest guidance strength among all other multimedia types and represents the best compromise, when

considering authoring costs, for conveying complex work instructions in similar manufacturing context. Moreover, unlike other multimedia contents, video demonstration provides an easy, yet flexible method for capturing manufacturing processes that have not been anticipated during the design of the system. Our assumption is therefore that FPV video demonstrations complemented with additional digital content (e.g., text, pre-defined graphical elements) is the most adapted approach for capturing and conveying any type of assembly information via AR.

### **Guidelines**

We have identified and collected guidelines and best practices for video-based tutorials, useful in designing the proposed AR training system. (Pongnumkul et al., 2011) defined four design goals that a video-based tutorial system should address, including (1) automate play and pause, (2) easy access to important frames, (3) visual highlights, and finally (4) low transaction cost. Further, we have identified another set of guidelines for recorded demonstrations proposed by (Plaisant & Shneiderman, 2005) and potentially adapted to recording instructional videos for manual assembly procedures. They include the following recommendations: (1) provide procedural instruction rather than conceptual information, (2) keep segments short, (3) ensure the tasks are clear and simple, (4) coordinate demonstrations with textual documentation, (5) use spoken narration, (6) use highlighting to guide attention, (7) ensure user control and finally, (8) strive for universal usability.

Finally, we have considered the general guidelines proposed by (Mattsson et al., 2016) to help assembly instruction designers to effectively support workers: (1) Support active cognitive processes - not too much information and focus on the most important and consider worker experience. (2) Support workers' mental models - how a person perceives a situation affects his behavior. (3) Support cognitive abilities and limitations - memory and attention is limited, thus take away redundant information. (4) Support individual preferences - humans may require different information. (5) Support perception through correct placement of information/pictures.

# Chapter 5. ATOFIS, THE PROPOSED AR CONTENT AUTHORING SYSTEM

---

*Low-cost, formalized, in-situ, immersive, on-the-fly authoring, adapted to the shop floor environment and workers without technical expertise.*

---

---

In this chapter we present the proposed AR authoring approach, so-called ATOFIS (**A**uthoring **T**ool **O**n-the-Fly **I**n-Situ). We start with a summary of the most relevant challenges and requirements that the proposed authoring approach aims to address. Then, we detail the adopted principles, particularly the 2W1H (What, Where and How) one and its three visual elements (i.e., text, location arrow and photo/video). To investigate generalization capabilities, we discuss two authoring extensions as well, adapted to assembly environments with specificities that are different from the ones of the considered assembly use case at elm.leblanc (e.g., seated workstation). Furthermore, we illustrate the user interface and present the authoring procedure of an AR instruction. Finally, we discuss the most relevant visualization and interaction techniques employed by the proposed authoring approach.

Part of the work presented in this chapter has been published in (Lavric et al., 2022a).

---

---

Before describing our AR authoring proposal, let us summarize the challenges, requirements and expectations identified during the considered industrial case study at elm.leblanc, which had a significant impact on the proposed authoring approach.

## 5.1 Challenges and requirements

We note that the following sections present challenges and requirements for the AR authoring system as well as for the information conveyance system. Some of these challenges and requirements are common between the two, however their justification is not always the same, as the training is dedicated to a different category of end-users than the authoring (i.e., novice workers *versus* shop floor experts) while the objective of the two functions is completely different as well (i.e., creation vs. conveyance of AR work instructions).

### 5.1.1 Technological challenges and expert knowledge

The literature shows that the AR technology has not been widely adopted in complex industrial operations (Bottani & Vignali, 2019). High cost and the uncomfortable wearing experience, susceptible tracking robustness/registration accuracy, unsuitable user experience, bad design, poor integration with enterprise systems, lack of flexible and easy to operate AR work instruction authoring systems represent some of the major concerns that restrained the large scale industrial adoption of AR (Palmarini et al., 2018), (Bottani & Vignali, 2019), (Nee et al., 2012), (X. W. S. K. Ong & Nee, 2016), (Fite-Georgel, 2011), (Fraga-Lamas et al., 2018). Current design and authoring of AR work instructions depend on specifically customized code development claimed (Nee et al., 2012), emphasizing the importance of the design phase of an AR application, such as the development of highly interactive and user-friendly interfaces and providing valuable insight to make AR an interesting tool in manufacturing and engineering fields. Today, most AR authoring tools are generally operated by highly qualified developers rather than by the end users. This situation leads to high costs and low adaptability (Palmarini et al., 2018).

### 5.1.2 User acceptance and shop floor reluctance

From the shop floor perspective, the goal is to obtain better training while relieving the work duties of people in charge of training. However, authoring AR work instructions comes as a new responsibility for shop floor managers or experts, which represent the potential and eventually most adapted end users of the considered AR system (Lorenz et al., 2018). This concern becomes even more important when dealing with the reluctance of shop floor personnel to change, learn and assimilate new technologies within the shop floor context (van Marrewijk, 2018), (Courpasson et al., 2012), (Lines et al., 2015). Therefore, to address user acceptance, the authoring process should ideally not require technical expertise, and even less the expertise of multiple individuals. In addition, it should be easy to learn and use, reliable and able to successfully address different assembly environments.

Regarding training, the system should be designed such that novice workers would prefer it over traditional training methods. To address user reluctance to change, the training system should allow one to perform the training easily and effectively, independently on his AR or assembly



experience, cognitive qualities, preferences, and learning style. Even though user acceptance requirements for training are less strict than for authoring (as novice workers do not have the same power of decision as experienced workers or shop floor managers), training feedback still represents a key aspect in the overall acceptance of the considered AR training system.

### 5.1.3 System performance and usability

Performance and usability represent two critical indicators in supporting domain experts in authoring AR applications without technical knowledge (Gonzalez et al., 2019). *To be a success on the shop floor, the [AR] system must prove its usability and effectiveness right away*, affirms the CIO of the factory. We note that to provide an objective assessment, content authoring and training should be evaluated together. The authoring system should allow a shop floor expert to capture his assembly expertise fast and efficiently. The information conveyance should be as explicit as possible to support an accurate execution of the assembly operations by the novice worker. At the same time, the information conveyance should stimulate a fast assembly process, while not compromising product quality. Human supervision during training should not be necessary or at least minimized, to relief the shop floor experts in charge of the training.

### 5.1.4 Safety, psychological and physiological concerns

The AR system must not introduce potential safety risks during the authoring and training procedures. Until reliable and comprehensive field experiments regarding the psychological and physiological impact that AR can have on humans are conducted, designing an AR system should delicately consider these implications. The most adapted AR devices, visualization methods and interaction techniques should be identified and employed in the system implementation, to optimally address human concerns in relation to application usages. For instance, minimizing the interaction time with the virtual elements to reduce eye stress, providing a user interface to limit head movements and hand gestures, or providing a fast user experience to shorten as much as possible the AR session to reduce physical and mental effort and discomfort are approaches that partially address human psychological and physiological concerns.

### 5.1.5 Efficiency and viability

From an organizational perspective, the main challenge is to provide better training without compromising product quality. More specifically, to be considered a success, the proposed AR training should be reliable, more flexible, and less costly than the traditional one, while producing the same or better results and respecting organizational norms and requirements.

To ensure industrial viability, the proposed AR training system must also be safe, effective, and adaptable to new scenarios without the need of expert knowledge or technical setup. The system should be mobile, easy to manipulate and ready to use on-demand, independently on the assembly location or context. A technical setup should ideally not be necessary, even in new assembly use cases, except perhaps initial (i.e., one-time) system calibrations or procedures which do not require expert support. The maintenance of the solution, if necessary, should as well be performed by the shop floor experts.

The system should allow easy integration of technological features that are not yet available or did not reach a maturity level for industrial usage, even though this would potentially imply the need of specific technical expertise. However, system extensions should be carefully considered to not compromise usability and effectiveness. The authoring process should be independent of external services, especially those that involve recurrent costs.

### 5.1.6 KPI and costs

A highly influential aspect in the adoption of novel technological tools in the industry, the KPIs cannot be measured unless the considered tool is deployed and used in a real-world use case for months, potentially years. In addition, to evaluate the performance of assembly systems powered by emerging technologies, standard KPIs need to be adapted to include both traditional metrics and specific indicators (Miqueo et al., 2020). To partially overcome these challenges, the solution must be able to demonstrate potential benefits to the stakeholders very fast. To increase the success rate, the initial and running costs should be minimized, as the budgets for innovation in manufacturing factories are generally limited.

## 5.2 Methodology and system implementation

In this section we present the proposed AR content authoring system, dedicated to creating AR step-by-step work instructions for manual assembly operations. We aim to address the concerns presented in the previous section and provide a human-centered AR authoring tool adapted to shop floor experts, while considering the large variety of people, products, processes and working conditions, as recommended in (X. W. S. K. Ong & Nee, 2016) and (Fite-Georgel, 2011).

We note that, because of the holistic nature of the overall AR training system, particularly when considering organizational challenges (e.g., shop floor acceptance), the design of the proposed AR authoring approach has deep implications on the design of the AR training proposal. As observed in (Fernández del Amo et al., 2018), AR knowledge transfer is affected by authoring through its ability to create content. The knowledge transfer is validated by the correctness of the authored augmented content.

In our work, we aim at providing a deeper exploration of an AR training system as a single unit, since its main two functionalities, authoring and training, are interdependent and complementary at the same time. We have observed that the literature often analyzes and evaluates the two as separate entities. In our opinion, this represents a research gap that must be addressed to redefine needs, expectations, and challenges of an AR training system as a whole.

### 5.2.1 Concept and principles

Our objective is to provide a reliable system implementation of the AR authoring proposal that can be objectively and effectively evaluated in real world conditions, namely in manual assembly context. The proposed authoring tool considers critical factors discussed in Chapter 4, including authoring effort and user acceptance, cognitive considerations and HCI paradigms, system flexibility and viability. We aimed at providing a formalized yet flexible AR authoring that can

deal with the complexity and variety of manual assembly operations from different manufacturing environments and sectors.

The most relevant principles that we have adopted and that serve as basis for the proposed authoring approach include:

1. Generic and consistent, independently on the manufacturing context.
2. No installation setup or prerequisites required.
3. No technical expertise required.
4. Formalized and easy to learn authoring procedure.
5. Adapted to the shop floor, fast, intuitive, and easy to use.

#### 5.2.1.1 *Assembly process description formalization*

As discussed in Section 2.2.4.3.2, the literature does not yet provide a standardized way of describing assembly expertise in AR. However, this represents an essential issue in the industrial context, where the work instructions represent a directive document of the enterprise that must be designed and audited in a standardized manner before it can be released on the production site (D. Li et al., 2018), (Haug, 2015).

However, the literature reveals a set of difficulties when it comes to standardization. One is that too many rules might limit human actors do the right thing especially in states of abnormal operation where they would need strong, but flexible guidance (Grote & Weichbrodt, 2007). Rules specifying the exact operations to execute can have a detrimental effect on action because they do not allow the user performing the action to develop an underlying plan of their own. Instead, he must face a certain atomization of the actions involved and focus on micro-difficulties. In non-routine situations, reliance on standards may turn into an over-reliance (Grote, 2004), potentially persuading the creator of the AR work instructions to follow the rules more than using his abilities and knowledge to describe the assembly operations clearly and effectively. We address these concerns by trying to identify the best compromise between formalization and flexibility, as further described.

##### 5.2.1.1.1 *Assembly instructions chunking*

An analysis of the assembly instructions composition, as defined in existing paper instructions, reveals that each assembly instruction is generally composed of multiple actions or tasks. For example, the instruction “*Screw the crosspiece with 4 screws using the gray screwdriver*” requires the worker to perform additional actions that are not explicitly specified in the text description. The missing information includes the picking of the screws and the screwdriver and their location, the four screwing locations, and finally the tool manipulation. We note that generally, in paper instructions the location of the assembly components is not specified or described, neither how the assembly operation should be performed. Such details may be needed to successfully execute the assembly operation.

To address assembly description completeness (i.e., missing vital assembly details), and eventually avoid ambiguity during the creation of the AR work instructions, we propose a technique that requires the author to describe one and only one assembly action per AR work instruction. In other words, each AR work instruction should describe one assembly operation,

composed by a single verb (i.e., assembly action) and a single subject (i.e., assembly component). By applying this technique, the assembly operation example presented above will be chunked into three discrete actions:

- action 1: “**Grab** four screws”,
- action 2: “**Grab** the gray screwdriver”,
- actions 3 to 6: “**Screw** the 1<sup>st</sup>/2<sup>nd</sup>/3<sup>rd</sup>/4<sup>th</sup> screw”.

Advantages of a similar chunking technique have been demonstrated in (Tainaka, 2020). (Gattullo et al., 2019) propose dividing long instructions into atomic actions as well, a procedure which represents the first stage for converting traditional paper instructions into AR-compliant documentation. Benefits of segmenting a dynamic presentation (e.g., video demonstration) into cognitively manageable chunks have been demonstrated also in (Carroll & Wiebe, 2004), within the framework of spatial procedural instructions consisting of folding an origami figure.

#### 5.2.1.1.2 The 2W1H principle

As it can be observed in the above-mentioned assembly example, each of the six decomposed actions specifies **what** should be done, in a specific order, without providing details about **where** the components are stored or installed, neither **how** the assembly operations should be executed. By performing the same analysis on all the existing paper instructions of the considered assembly line, we have noted that each assembly operation, independently of its type and complexity, can be described by the three variables **what**, **where** and **how**, to ensure a complete description. **What** represents a brief description of what should be done, similar to the ones already provided in the paper instructions. **Where** indicates the spatial location of the considered assembly operation. Finally, **how** describes how the assembly operation is performed.

We note that by using these variables for describing an assembly operation in the given order, *what-where-how*, we try replicating the oral human-to-human explanation of manual tasks. This approach is encountered in written and digital tutorials as well. During our case study experiments, we have observed this natural, generic explanation process adopted by the shop floor experts while explaining and demonstrating assembly instructions to novice workers. Firstly, the instructor “announces” the trainee **what** the assembly operation is about, e.g., “*Next, we are going to grab the crosspiece*”. Secondly, the instructor indicates **where** the assembly operation takes place by moving or reaching towards the location and pointing with his hand or finger, e.g., “*...from here*” or just “*...here*”. Finally, the instructor demonstrates **how** the operation is executed by performing and eventually explaining it at the same time, e.g., “*...you do it like this...*”. In the following, we will refer to this principle as the *2W1H* (2-Where and 1-How) principle.

By adopting this principle, it becomes possible to set up a formalized authoring procedure, that is consistent, straightforward, and independent of the assembly operations involved (e.g., grab, place, insert, push, etc.). We naturally expect that the authors of AR instructions provide accurate and effective descriptions of the assembly operations during the authoring procedure.

### 5.2.1.2 Visual representation of assembly information in AR

As noted in a recent study (Jasche et al., 2021) regarding different types of AR visualizations for assembly instructions, when dealing with complex operations, the error rate is often more important than the completion time. This claim is supported as well by our case study, which indicates that product quality is the top priority and should not be compromised to the detriment of faster assembly times. As pointed out in Chapter 4 - Section 4.2.2.4, concrete visualization is better adapted rather than abstract representations, particularly for conveying complex assembly information to novice workers. Furthermore, the visual representation of the assembly should be as complete as possible, avoiding however redundancy. Finally, our in-depth analysis regarding information representation and conveyance discussed in Chapter 4, together with the authoring concerns and the considered shop floor context, suggests that the optimal way to capture and represent assembly information in AR is by using a combination of multiple low-cost visual assets, including text, photos, video, and predefined auxiliary models.

We note that unlike CAD data, such low-cost assets can be captured by state-of-the-art AR devices, *in-situ*, during the AR authoring procedure itself, as demonstrated in the preliminary evaluations conducted during the case study. As suggested in (Jasche et al., 2021), videos should always be considered as an essential part of AR-based instructions as they lower the perceived complexity and implicitly the mental effort during the execution of assembly tasks, eventually leading to an error rate reduction as well. In addition, user feedback suggested that in addition to video, participants prefer concrete presentations in general and textual information as a complement. Finally, it was noted that the utility of video demonstrations is diminished with the assembly complexity, suggesting that abstract content, potentially complemented by photos of the final assembly could be sufficient for representing simple or obvious assembly operations.

Proposing an AR training system which does not rely on CAD data for conveying assembly expertise might be intriguing. However, let us recall that CAD content creation requires skills such as 3D modelling, computer graphics, animations, and programming. Shop floor experts often have little or no knowledge in creating AR content (Gattullo et al., 2019), which means that appropriate AR authoring tools are required to allow a quick and easy way to create content (Jasche et al., 2021). Another concern regarding the usage of CAD data for authoring AR work instructions is represented by the time and effort required to place them in the AR scene. Existing work deals with this concern by proposing automated registration of CAD models in the real-world environment (Stock et al., 2005), (Knöpfle et al., 2005). However, the tracking accuracy is generally not sufficient to make such technique reliable and scalable (Scurati et al., 2018). The overall research conducted throughout this thesis strongly suggests that in similar industrial sectors, the usage of CAD data for AR authoring systems is not yet convenient. The major concerns related to the usage of CAD data, identified during our case study, include:

- availability and preparation (for AR usage),
- positioning during the authoring: tedious and time consuming, particularly for small assembly pieces like screws and clips,
- registration accuracy: the precision is not sufficient for complex assembly environments, e.g., two holes at 1 cm from each other,

- registration for objects in motion: state-of-the-art proposals for real-time world registration and 3D tracking for objects in motion do not produce acceptable results, particularly for industrial usage,
- occlusion, user distraction and eyestrain.

Another potential advantage of not depending on CAD models is minimizing the negative effect on the UX, during both authoring and training phases, in the case of an imprecise spatial registration with their corresponding real-world objects. An imprecise registration of other graphical media elements like photos and video, which do not require a superposition on the real-world objects, will not affect the UX as much as CAD models would, notably in the case of small objects and mobile components.

#### 5.2.1.3 *In-situ authoring*

As pointed out by (Lorenz et al., 2018), shop floor experts are the most suited to create the AR work instructions, instead of AR or IT experts. Our case study findings support this hypothesis, claiming that any industrial tool should be designed for the people that are supposed to benefit from it, because they are the potential end-users. To address such requirements, we propose therefore an authoring process that is conducted by shop floor experts, in the shop floor environment. By adopting an *in-situ* approach, we aim to provide the optimal authoring context during the creation of the AR instructions, which is the workstation itself and its components. This context provides the raw data of the assembly process, which the AR work instructions are supposed to convey to novice workers during training.

#### 5.2.1.4 *Immersive authoring (WYSIWYG)*

Shop floor workers generally have no technical knowledge, and even less skills in manipulating AR authoring tools. (Gonzalez et al., 2019) claimed that one of the biggest disadvantages of AR is the lack of general awareness of the authoring progress. Better visualization techniques need to be explored to show users their current authoring state, for AR to have a real definitive advantage over desktop. To address this concern, we propose an immersive authoring of the AR work instructions, similarly to the well-known *What You See Is What You Get* (WYSIWYG) principle, or his AR variant proposed by (G. A. Lee et al., 2005), so-called *What You eXperience Is What You Get* (WYXIWYG). Immersive authoring is the most prevalent form of visual authoring (Foley D. et al., 2004), its benefits being leveraged by the interaction usability and ease of use (G. A. Lee & Kim, 2009). Quite a few research works have explored the use of immersive authoring tools (Rumiński & Walczak, 2013), (Hodaie et al., 2019), (G. A. Lee et al., 2005), (Zauner et al., 2003), in which virtual content is created *in-situ*. Results from prior studies show that interactively placing and manipulating the virtual objects is easier and faster for users to carry out authoring tasks compared to a non-immersive environment (G. A. Lee & Kim, 2009). Other literature findings suggest as well moving from current desktop environments to web-based environments, or, in the considered context, to AR-based environments (Fischer, 2001). By adopting an immersive AR authoring approach, we aim therefore to minimize the authoring time while making it more efficient and intuitive for people without AR experience. Furthermore, the author visualizes and validates the AR work instructions during the authoring procedure itself.

### 5.2.1.5 AR content registration

Content registration is a core function of numerous AR systems but remains an open issue of research. The three main types of information registration methods for HMD-based AR are head, object, and environment (Tainaka, 2020). Marker or fiducial-based techniques represent the most utilized (57%) registration procedures in industrial applications. Other techniques - i.e., 2D/3D recognition, sensor-based, location-based and marker less - do not comply with industrial robustness and precision requirements and are generally limited to test environments (de Souza Cardoso et al., 2020). Consequently, to address the WYSIWYG paradigm, we employ head (head-gaze technique) and environment (marker-based technique) registration methods for AR content registration during authoring.

To ensure a precise world registration of the captured data during authoring, the proposed approach depends on fiducial markers, which represents today the most reliable AR registration technique. We note that fiducial-based registration requires static assembly environment, a condition that is generally met in the considered manufacturing use-case. It is desirable and expected however that future technological advancements replace the need of fiducial markers by providing intelligent and precise real-time world registration that relies on natural features, to address any kinds of assembly, including dynamic and mobile.

### 5.2.1.6 On-the-fly authoring

To further optimize the authoring process, we propose the creation of the AR work instructions on-the-fly (i.e., at the same time with the assembly process). We observe that *in-situ*, WYSIWYG authoring does not only support the adoption of such approach, but potentially makes the creation of the AR work instructions more intuitive and concrete, as the shop floor expert is required to capture his expert knowledge in AR while performing the assembly activity itself. Furthermore, on-the-fly authoring requires the shop floor expert to perform the workstation assembly cycle only once, while capturing the information successively, for each assembly operation, by applying the chunking and 2W1H principles discussed before.

### 5.2.1.7 AR HMD device (Hololens 2)

From a hardware perspective, wireless, HMD AR devices seem to answer best the manufacturing environment challenges. Handheld devices (i.e., smartphones, tablets) do not fulfil the hands-free requirement, while SAR systems are definitely not adapted to our use case. To support on-the-fly, *in-situ* WYSIWYG authoring, the use of an optical head-mounted display AR device represents the optimal hardware choice. In addition to allowing the authoring of the AR work instruction to visualize the real-world assembly environment in real time, state-of-the-art HMD devices like Hololens 2 provide support for natural interaction and world registration, two of the most important features that an AR system is expected to provide for a rich user experience.

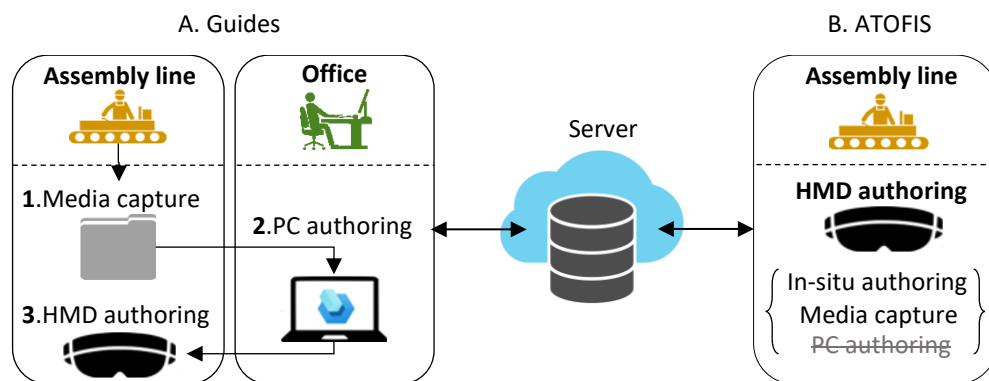
It must be mentioned that today's AR HMD devices rise numerous concerns including resolution, field of view, weight, short battery life, ergonomics, poor optical system, and costs (S. K. Ong et al., 2008), (Lindberg et al., 2015), (S. J. Henderson & Feiner, 2009). As noted in the literature, until a pair of AR glasses can blend as well as any other pair of glasses with a user's style, their popularity and trust will not take off (Martinetti et al., 2019). On the other hand, compared to

other AR devices and information conveyance techniques, AR HMDs provides portability and flexibility, and most importantly, a natural and immersive user experience.

To minimize the potential negative effect of some of these concerns, our authoring proposal aims to be as fast as possible, without compromising the user safety and the quality of the authored AR instructions. Field studies which consider psychological and physiological considerations must be conducted to provide objective evidence regarding the limitations of state-of-the-art AR HMD devices and their implications on humans, especially in the long term. Our experiments show that, in general, devices like Hololens 2 can be utilized for at least one hour without participants making any negative remarks about weight or discomfort.

#### 5.2.1.8 Independent, one-step authoring process

Finally, the proposed authoring approach does not depend on existing resources, external services, the usage of Desktop applications, preparation, or post-processing steps, as in the case of popular commercial AR authoring tools like Vuforia and Guides. The proposed authoring process is performed in a single authoring stage, requires the use of a single hardware device (Hololens 2 or other HMD AR device that provides the same functions), a unique QR code and an Internet connection. The authored AR work instructions are immediately ready to be used for training, once the authoring finished. Figure 5-1 summarizes the authoring workflows of Guides (A) and ATOFIS (B), highlighting the main differences between the two approaches.



**Figure 5-1. Authoring workflows of Guides and ATOFIS. A) Guides, a 3-step authoring composed of media capture, PC authoring and HMD authoring; B) ATOFIS, a 1-step in-situ, HMD-only authoring.**

### 5.2.2 Authoring procedure and guidelines

In this section, we describe the procedure and guidelines that a shop floor expert is required to follow during the authoring of the AR work instructions, from a content perspective. The authoring workflow and the way the author interacts with the user interface to capture the considered media contents is explored in Section 5.2.5.

The author of the AR work instructions must follow the formalization imposed by the 2W1H principle, for describing each assembly operation in AR. Thus, the visual description of any assembly operation is composed of three elements:

- a text instruction, briefly describing **what** the assembly operation is about,
- an arrow pointing to the **where** the physical location of the operation is,



- a FPV photo illustrating the final assembly result or a FPV video capture demonstrating **how** the assembly operation is executed (optional).

#### 5.2.2.1 *What – textual description*

The text instruction should describe the assembly operation in a clear and concise manner. Even though the length of the text description should be reduced as much as possible, capturing important and subtle assembly details should be prioritized and therefore not omitted. The training interface does not use visual cues to highlight specific parts of the text description. Therefore, the author should keep in mind that every piece of information that he provides during the text insertion will eventually have the same level of importance from the perspective of the trainee. We do not provide visual cues like bolding and underlining during text insertion simply because such a functionality would make the overall authoring procedure more complex and slower, while the potential benefits are not guaranteed. The author should also consider that the textual description provided at this stage represents the first and potentially less relevant information regarding the considered operation, as it only suggests what should be done without specifying details.

#### 5.2.2.2 *Where – auxiliary data (i.e., 2D arrow)*

*Where* represents the physical location of the assembly operation. The author must provide this information for every AR instruction as all assembly operations take place at a certain location in the real world. The author is expected to mark the assembly location in the physical world by placing and orienting the location arrow to where he would have pointed his finger at during a classical training procedure. The arrow should not interfere visually with the assembly location. In addition, to avoid occlusion, the location arrow should be placed at least a few centimeters away from the actual assembly location but oriented as precisely as possible towards it. It is important to note that the main objective of *where* is to guide the trainee towards the assembly location and its associated media content, but not necessarily to indicate a very precise position (which might be impossible to do). For example, indicating a screwing location on an assembly component with multiple wholes which are extremely close to each other (e.g., 0.5 cm away) will not only be nearly impossible to indicate during the authoring, but an imprecise registration (e.g., 0.26 cm, which today is considered as excellent) would erroneously indicate the assembly location during training. The author should be aware of this limitation, and, in the case of a highly sensitive assembly environment, the location arrow should be used only to serve as approximate real-world indication. This location approximation will be further clarified by a demonstration video, as discussed in the following section.

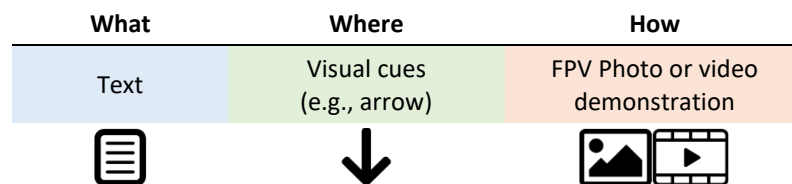
#### 5.2.2.3 *How – assembly photo or video demonstration*

Finally, the author is expected to demonstrate the assembly operation by either taking a photo of the final assembly or by capturing a video demonstration while performing the assembly. It is up to the author to decide, based on the assembly difficulty and his training experience, which of the two media types would be more appropriate for a novice worker to understand and perform the assembly correctly. It is recommended that a video demonstration is captured in case of doubt, rather than providing insufficient information with a photo that could lead to comprehension issues, assembly errors and eventually need of human intervention during the

training procedure. In the case of a photo capture, the author should indicate clearly the expected final assembly result, eventually highlighting it by using the most adapted method (e.g., hand, finger, specific tool). In addition, the quality of the captured media should carefully be checked to ensure good focus and sharpness.

During video capturing, the author can also explain the assembly orally. However, the assembly demonstration should not rely on audio instructions as listening to them during training may not be possible (in the case of noisy industrial environments). It is recommended that the assembly demonstration is performed slowly, while the hands are always in the FoV of the camera, to address the cueing principle discussed in Section 4.2.4. For complex assembly operations or environments, the author should carefully indicate and highlight details which are error prone during the video recording.

Figure 5-2 summarizes the relation between the three elements of the 2W1H principle and the associated visual assets used to represent each of them in AR.



**Figure 5-2. Low-cost visual assets used to capture and represent the assembly information via AR in our proposed approach.**

### 5.2.3 Authoring extensions

The proposed authoring approach was elaborated to specifically address technological limitations and industrial requirements identified during the considered case study. We acknowledge however that other manufacturing sectors can present different needs and objectives. We claim that the proposed authoring can be extended to address different use cases. We exemplify this by presenting two functions that have been implemented and analyzed within the same context of this work. The first, related to the authoring of a digital twin of the assembly environment, has been proposed and evaluated by Bégout P. in his thesis. The usage of such function as a complement to the proposed AR authoring system is discussed in Section 5.2.3.1. The second extension, presented in Section 5.2.3.2, is related to the way the author decides to create the AR work instructions for successive similar or identical assembly operations. This content-based extension was exemplified in the authoring of the AR instructions used in the experiment described in Section 7.2.2.

#### 5.2.3.1 Usage of 3D CAD content and workstation modeling

The proposed AR training approach has been designed without being dependent on 3D content for reasons that have been discussed in detail in Section 5.2.1.2. It must be noted that this decision was mainly influenced by the impossibility of creating 3D CAD models (i.e., product models of assembly components) in the same manner as the low-cost visual assets (e.g., text, image, and video), that is during the authoring procedure itself by only using an AR HMD device.

However, the potential of 3D content, especially 3D animated content is non-negligible. For this reason, the proposed AR training approach should be in a certain sense 3D-ready, by considering capturing and conveying assembly expertise by using CAD data and 3D animations, as long as the main authoring principles and guidelines are respected.

We identify and propose therefore a modality of exploiting 3D content that partially addresses our requirements and extends the authoring proposal without impacting the overall design approach. This extension implies the existence of a preparation stage that must be conducted off-line, which consists in *modeling* the assembly workstation. More precisely, this preparation stage requires the author to create a 3D virtual representation of the assembly workstation, by placing the corresponding model products of the workstation (i.e., the 3D model of an assembly component) to their corresponding world location in the storage area for picking, but also in their final assembly position. During the authoring of the AR instructions, the spatially registered 3D content is presented. The author has the possibility to indicate the position of a component or assembly by simply selecting the corresponding 3D model, instead of placing the location arrow as defined in the default authoring methodology proposition. We note that 3D animated content can be also used as a replacement for the location arrow and, at the same time, as a complement to the photo illustration or the video demonstration of the assembly.

The preparation stage can be conducted in the same manner as the proposed authoring procedure of the AR instructions. Once a workstation identified (by QR code or other means), the modeling application displays the available 3D content associated to it. The author is required to spatially register the virtual elements at their corresponding workstation location. This process is performed *in-situ* and in a WYSIWYG manner, in a very similar manner with the positioning of the location arrows, by using natural hand interaction. Such functionality would potentially simplify the authoring of the AR instructions, as the author would be required to only select the corresponding spatially registered 3D content instead of indicating the assembly location by spatially registering the location arrow (i.e., step 2 of the proposed authoring procedure). We note that the 3D media content preparation and linking with the assembly workstation is a mandatory prerequisite of the workstation modelling stage. Such an authoring extension - i.e., WARD (Workstation Augmented Reality Digital Twin) (Bégout et al., 2020), was not only elaborated and proposed within the same industrial conditions at elm.leblanc but was as well validated in the considered industrial use case. We note that the AR authoring of digital twins proposed by Bégout P. in his thesis can be used as a complement but is not a requirement for ATOFIS. We acknowledge however the importance of such authoring function, that would not only leverage the potential of 3D CAD data in the proposed AR training system but, in the case of changes of the physical assembly layout, the system would be able to identify and reflect these changes during training in real-time. In other words, such an authoring extension as WARD, would ensure the correctness of both the CAD models of the AR instructions corresponding to the physical assembly components as well as their spatial registration, in the case the workstation was subject to change since the authoring of the AR instructions. The challenges of this approach are represented by the usage of 3D CAD data, the live update of the

corresponding digital twin of the assembly and the precise registration with the real-world environment.

### 5.2.3.2 Aggregation of successive identical assembly operations

The proposed AR authoring approach can also be extended from a usage perspective, by aggregating successive identical assembly operations under a single AR work instruction. This may be helpful for speeding up the authoring process and most likely the training as well. Thus, we have noticed that simple assembly operations like screwing or riveting can be aggregated under a single AR instruction, as long as they are performed consecutively, require the same assembly components/tools and their assembly locations are very close to each other (ideally in the same field of view). Such example, encountered during the field experiments conducted at elm.leblanc, requires the worker to grab four identical screws, the screwdriver and screw an assembly part at four different locations. By strictly following the 2W1H principle, authoring this assembly sequence in AR would result into 6 instructions, each describing the following assembly actions: 1. Grab 4 screws, 2. Grab the screwdriver, 3. Screw the 1<sup>st</sup> screw, 4. Screw the 2<sup>nd</sup> screw, 5. Screw the 3<sup>rd</sup> screw and respectively, 6. Screw the 4<sup>th</sup> screw. By slightly adjusting the 2W1H principle, which states that each AR instruction should describe a single assembly operation, and consider that the screwing of the four screws represents a single operation, we can therefore combine this information and author it as a single AR instruction, as follows:

- **(what)** Text instruction: “*Screw the four screws at the indicated locations*”,
- **(where)** The author spatially registers four location arrows instead of one,
- **(how)** The author illustrates the assembly by either providing a photo with the final assembly, by capturing a video demonstrating the entire assembly or just the first screw (with an indication that the same should be performed for the remaining screws).

We note that such approach is adapted only if the 3 conditions are fulfilled: (1) same and (2) consecutive assembly operations performed at different but (3) relatively close locations.

### 5.2.4 System implementation

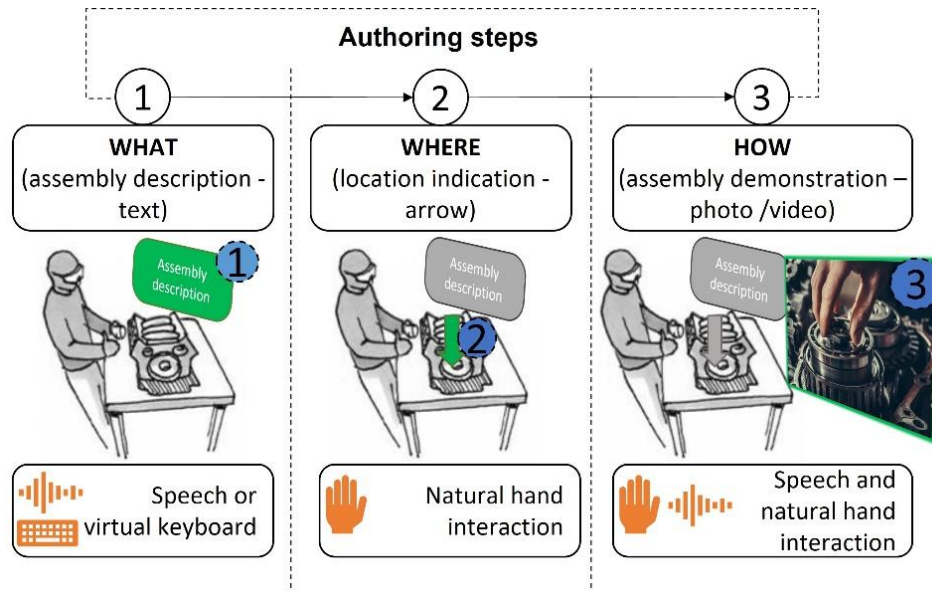
The implementation of the proposed AR authoring methodology is developed for Microsoft® HoloLens 2, by using Unity 3D (*Unity 2019.4.10*, 2019) and MRTK (*Microsoft MRTK v2.4.0*, 2020).

A backend service for data storage purposes has been implemented as well, for storing the authored AR instructions remotely rather than locally on the AR device, for concerns regarding data loss, security, and local storage limitations. The backend implementation is not part of the authoring of the AR work instructions itself and therefore will not present it in detail. We simply note that no technical challenges concerning data storage (i.e., server setup, DB, API) have been encountered during the development. Finally, in addition to the application being installed on the HoloLens 2, the authoring requires a unique QR code, used for linking the set of AR work instructions to its corresponding assembly workstation, and an Internet connection.

### 5.2.5 Authoring an AR work instruction

Further, let us describe how the content authoring is performed for a single AR work instruction. We note that the same process applies for creating any other AR step-by-step work instruction.

Figure 5-3 illustrates the authoring workflow overview, highlighting the three steps corresponding to the variables of the 2W1H principle, and the actions performed by the author at each of them, i.e., (1) text insertion, (2) location indication and (3) assembly demonstration. The interaction techniques (i.e., voice and natural interaction) employed by the author for completing each step are provided as well.



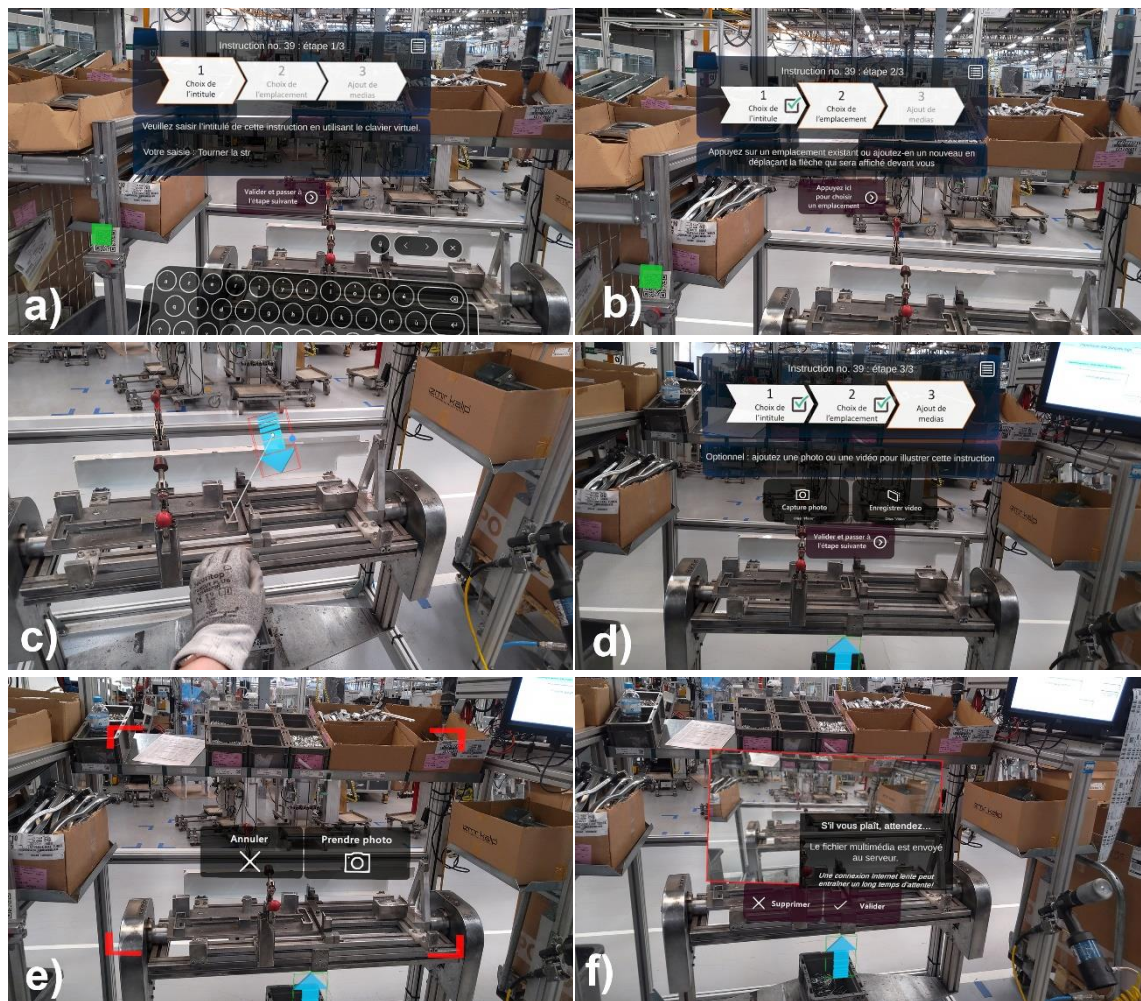
**Figure 5-3. ATOFIS authoring workflow overview for an AR instruction: (1) what: assembly description - text, (2) where: location indication - arrow and (3) how: assembly demonstration - photo or video.**

Figure 5-4 illustrates the HoloLens 2 implementation of the application interface of the considered authoring workflow presented in Figure 5-3.

At any time during the authoring procedure the author uses hand gestures and voice commands to interact with a 2D panel (Figure 5-4 a), b), d)), displayed in front of him by using head registration technique. The authoring panel has multiple functions. Firstly, it displays the current assembly instruction number and the authoring step within the current AR work instruction. Secondly, it allows the author to access the AR functions for text insertion, assembly location indication and FPV photo/video capture. Finally, it makes it possible to validate the captured data, advance to the next step and create a new AR work instruction. The proposed implementation provides in-app functions like visualizing, selecting, and editing existing AR work instructions.

Let us further describe how the author creates a step-by-step AR work instruction.

**What:** At this stage, along with the 2D authoring panel, a virtual keyboard is displayed in front of the user (Figure 5-4.a). The author uses the keyboard to insert a text for briefly describing the current assembly task, by using one of the two modalities: (1) natural hand gesture technique, which require touching the virtual keystrokes or (2) dictation by using voice. This function is activated by the user by clicking the microphone button, which is part of the virtual keyboard. The user goes to the next authoring step by clicking the validation button (part of the 2D panel).



**Figure 5-4.** AR authoring example. a) Step 1. Insert text instruction by using the virtual keyboard or dictation; b) Step 1 validated, step 2 active; c) Step 2. Positioning of the location arrow by using far interaction technique; d) Step 2 validated, step 3 active; e) Step 3. Photo-capture view; f) Step 3. Photo taken, the author validates or removes it.

**Where:** The author is required to place the virtual arrow for indicating the physical location of the assembly operation (Figure 5-4.c). The arrow is displayed in front of the user by using head registration technique, as a static object. The author uses his hand to grab, place, scale and rotate it. For manipulating the virtual arrow, the author can use the close interaction technique (*Mixed Reality Direct Manipulation with Hands | Microsoft Docs, 2020*) by touching the virtual representation of the arrow with his fingers as it were a real object. The second method is the far interaction technique (*Mixed Reality Instinctual Interactions | Microsoft Docs, 2020*) where the user can manipulate the virtual arrow from distance as illustrated in Figure 5-4.c. Finally, once the arrow placed at the desired real-world location, the author validates it by clicking a “validate” button displayed under the arrow. We note that at this stage, the only visible UI element is the arrow together with the corresponding validation button. We have adopted this approach to address the HCI principles discussed in Section 4.4. The interface is simple, straightforward and does not provide multiple authoring functions at the same time to avoid ambiguity and error. In addition, by only presenting the virtual elements with which the author is supposed to interact at a certain point, we avoid UI clutter and potential safety concerns.

**How:** The author decides if capturing a FPV photo (Figure 5-4.d)) or demonstration video is required. The author captures one of the two by using the multimodal user interface, either by interacting with the corresponding buttons of the authoring panel or by voice command: “photo”, “video” and “stop video”. We note that capturing the video demonstration from a first-person view is not only necessary because of the hands-free requirement but also because following video tutorials may be challenging when the user’s current viewpoint is not aligned with the one in the video (Breedveld, 1997). To address this concern, our authoring approach requires the user to (1) capture first-person view (FPV) instructional videos and then (2) spatially register these videos with the real-world environment so that, during training, the video element viewpoint is aligned with the user’s current view of the assembly component. The author spatially registers the captured media at a convenient location in the real world by using hand interaction techniques (*Mixed Reality Direct Manipulation with Hands | Microsoft Docs, 2020*) & (*Mixed Reality Instinctual Interactions | Microsoft Docs, 2020*). The author is required to position the media preferably in the same field of view with the assembly location, to limit the head movement during the training and potentially decrease the assembly time and the effort required by the trainee while following the AR instructions. We note that during training, photo and video elements automatically rotate towards the user; therefore, during authoring, the author should not spend time rotating these towards specific real-world locations, trying to guess the position and orientation of the worker during the AR training experience. We highlight the fact that a fixed orientation of visual illustrations could have a negative impact on the training experience, in cases where the height of the author is considerably different than the one of the trainees.

### 5.2.6 Interaction techniques

Let us summarize the interaction techniques considered for creating formalized AR work instructions *in-situ*, in a WYSIWYG manner.

**Speech and touch.** Speech and touch are interaction modalities that complement each other and allow the author to choose the preferred or the most familiar one. For example, the author can provide the text description by either using dictation or by interacting with the virtual keyboard by finger touch. Multimodal interaction is often required as well to address different assembly characteristics, like noisy environment or unreachable (by hand) location.

**Head gaze.** This implicit interaction technique allows anchoring persistent visual elements in front of the author (e.g., the authoring panel) or display temporary visual elements in the field of view of the user (e.g., photo and video captures). By using this technique, the position of the graphical elements with which the author interacts at different authoring stages, are dynamically displayed in front of the author. This technique simplifies the access to information, while providing contextualized data and speeding up the authoring procedure.

**Eye gaze.** The eye gaze interaction technique is used to control the video playback: the video plays as long as the user looks at it. This technique addresses the hands-free UI requirement, allowing at the same time removing the traditional video playback interface and avoiding therefore cluttering the UI and simplifying interaction.

Table 5-1 presents a summary of the interaction techniques retained. In addition, Table 5-2 presents the corresponding 2W1H-based authoring workflow.

**Table 5-1. ATOFIS: summary of the authoring interaction techniques, user input and results.**

Interaction technique	Author input	Result
Speech	Dictation or voice commands: “photo”, “video”, “stop video”	Input assembly text description FPV photo & video capture
Touch	Virtual elements (e.g., authoring panel, virtual keyboard, buttons)	Interact with authoring panel, type text description, trigger photo and video capture, validate steps.
Hands object manipulation	Instinctual interaction	Scale, rotate and move virtual elements
Head gaze	Implicit interaction	Dynamic positioning of virtual elements
Eye gaze	Implicit interaction	Video playback

**Table 5-2. ATOFIS: summary of the 2W1H-based authoring workflow.**

2W1H	Media type	Author task	UI elements	Interaction technique
What	Text	Briefly describes the assembly operation	Authoring panel, virtual keyboard, and buttons	Speech / touch
Where	2D arrow	Indicates the assembly location	Authoring panel, 2D arrow and validation button	Head gaze / hands object manipulation
How	Photo / video demonstration	Illustrates or demonstrates the assembly	Authoring panel, photo, or video capture screen, captured media, validation buttons	All interaction techniques

### 5.2.7 Key functional and design aspects

**Content formalization by interface.** The proposed authoring user interface requires the author to describe the current assembly operation in a formalized manner by following the 2W1H principle. The authoring panel displays the number of the current AR instruction and a graphical representation to indicate the active authoring stage. By design, the interface is hardcoded to not allow the user to go to the next authoring step unless the current one is completed. For example, the authoring interface takes the author automatically to the last authoring step of an AR instruction once a location arrow is validated. We ensure therefore the correctness of the authored content at least from the content formalization viewpoint.

**Dynamic and multimodal interface.** The authoring interface adopts a set of design and behavior guidelines for providing an easy to follow, guided procedure by using multimodal interaction while avoiding UI clutter. For example, to address multimodality interaction requirements, the text insertion can be added either by using a virtual keyboard or by dictation, while photos and videos can be captured either by using vocal commands or by interacting with visual elements (e.g., buttons). It is up to the author to choose the preferred interaction method or the most adapted considering the assembly environment (e.g., no vocal command in extremely noisy environment). Another behavior example is represented by the authoring panel which is hidden during the placement of the location arrows to avoid obstruction of the real world during this



process. The location arrows can be manipulated by using both close and far interaction, a choice of the author based on his preferences or depending on environmental constraints (e.g., placing an arrow at a location which cannot be reached by hand requires far interaction). The author must validate the position of the location arrow to get to the next authoring stage. Once validated, the arrow interaction is disabled to prevent involuntary manipulation during the assembly demonstration. This behavior aims to address a well-known AR concern which is represented by unintentional interaction with the virtual elements.

**Information contextualization by using *in-situ* authoring.** Further, we note that the contextualization of the media elements is one of the main differences between our approach and state-of-the-art AR authoring tools like Guides. The HCI considerations analyzed in Section 4.4 suggest that the author should decide where the augmentation media is presented during training, rather than teaching and allowing trainees to interact and change the position of the virtual elements. This is the reason why, during authoring, the author is required to position the captured photo or video next to the assembly location, preferably in the same field of view but without occluding or covering the assembly environment. During training, these elements are displayed exactly at the same locations, which brings us to another key aspect adopted by our proposed authoring, which is represented by the way that the AR instructions are created, in a WYSIWYG manner. This approach allows the author to create the AR work instructions *in-situ*, during the assembly process, to visualize and validate his creation right away, during the authoring itself. The authored content does not require any post-processing phase, making each AR work instruction available as soon as its authoring procedure has finished.

## Chapter 6. THE PROPOSED AR TRAINING SYSTEM

---

*An intuitive AR training approach that relies on familiar visual assets and implicit interaction techniques, to support novice workers effortlessly follow the AR work instructions and perform the manual assembly in the least intrusive manner.*

---

---

This chapter details the AR training proposal, with underlying concepts and information representation principles. We notably analyze how the proposed approach addresses industrial challenges and expectations of the considered assembly context, while dealing at the same time with human and environmental factors. Then, we describe the associated user interface and usage of the system during training, while analyzing how the proposed visualization and interaction techniques address human-centered design paradigms. Finally, a summary of the key functional and design aspects of the proposed AR training approach is presented.

Part of the work presented in this chapter has been published in (Lavric et al., 2022a).

---

---

The elaboration of an AR training system should prioritize human-centered paradigms, for it to become effective and potentially replace traditional training procedures. Even though novice workers do not generally have power of decision at organizational level, providing them with a complex and difficult application would most probably be detrimental to acceptance, because of the difficulty to use the system effectively in a short time span. It is important to note that in assembly environments like the one considered at elm.leblanc, learning the operation sequence of certain workstations generally requires only a few hours. It would then not make sense to provide a training alternative to the traditional one unless the corresponding learning curve is extremely fast.

The industrial partners also indicated that it would be ideal for the system to support the training of novice workers in real-time. For this reason, they recommend that the considered AR training proposal should be capable of assisting novice workers in executing the assembly as good as experienced workers, immediately. In addition to training time and efficiency, worker safety and product quality are priority factors that need to be considered. Let us also note that, from an organizational perspective, technical concerns are of little interest if the system fulfills the learning expectations in the considered industrial setup.

Under the challenging perspective of providing real-time training, our proposal aims to overcome major limitations of traditional training methods employed in similar industrial sectors. Thus, it is available on-demand, portable and easily reusable, does not require human supervision, and the assembly expertise is described and conveyed in a formalized manner.

## 6.1 Information representation of assembly operations in AR

The visual representation of an assembly operation presented during the AR training is dependent on the content authoring process described in Section 5.2.2, which adopts the 2W1H principle. Consequently, any assembly operation will be represented by the three visual elements retained:

- (*what*) A text instruction, briefly describing the assembly operation,
- (*where*) A location arrow, pointing to the real-world assembly location,
- (*how*) A FPV photo or video, illustrating the assembly result or demonstrating the considered assembly operation (optional).

During the training phase, any combination of these elements can be used to convey the assembly information. As the authoring is performed in an immersive WYSIWYG manner, it is expected that the corresponding visual elements are displayed in a similar manner as they were presented to the shop floor expert during the authoring procedure. As the proposed information conveyance method was specifically designed to accommodate novice workers, all the media content captured during authoring is used to convey the assembly information during training, as described in the following section.

## 6.2 Information conveyance: principles and techniques

The information conveyance and the authoring are coupled by the 2W1H and WYSIWYG principles to allow shop floor experts not only to create the AR instructions *in-situ*, in a formalized and immersive manner during the assembly procedure itself, but also to visualize the

AR instructions as they will be presented during training. By adopting this strategy, we provide an implicit validation of the AR instructions, literally part of the authoring procedure itself.

We note that the proposed information conveyance technique aims to address in addition to requirements such as information correctness and completeness, human considerations like cognitive load. As identified by (Abraham & Annunziata, 2017), AR can reduce cognitive load by supplementing human labor with the right information at the right moment and in the right mode, without disturbing the user's focus. This requirement is supported by (Sahu et al., 2021), which claim that seamlessly displaying AR procedural information to the user is essential to minimize the workers' cognitive load. In addition, to provide an appropriate AR experience, the system should not display redundant augmentation media during the assembly procedure (G. Lee et al., 2019).

The information conveyance approach retained in our work is adapted to the manual assembly context of the considered case study, and presents the following characteristics:

- the organization of the assembly is static,
- the worker is standing and required to move during the assembly procedure.

In the following, let us exemplify the functioning of the proposed approach by using the generic assembly example provided in the case study. A comparison between classical human-to-human (H2H) explanations and the proposed AR information conveyance technique is also provided.

### 6.2.1 What

**H2H:** First, the instructor "announces" (by voice) the novice worker **what** the assembly operation is about, e.g.: "*Next, we are going to grab the crosspiece*". The worker may ask for repetition if the audio information is not heard or understood.

**AR:** The transient nature of spoken text suggests that conveying this information in AR via audio is not adapted (Section 4.2.2.6). Providing an UI where the worker can re-listen the audio instruction would break the HCI principles discussed Section 4.4. Presenting this information in a non-transitory and semi-invasive manner is the most appropriate approach, particularly adapted for novice workers. Conveying therefore this information via text displayed in front of the worker as the very first instruction of an assembly operation seems to replicate as closely as possible the classical information conveyance, while addressing the transitory information effect and the signaling principle (Mayer & Fiorella, 2014) as well, to avoid potential overlooking.

### 6.2.2 Where

**H2H:** The instructor indicates **where** the assembly operation takes place by moving and/or looking towards the corresponding assembly location, eventually pointing with his hand or finger, e.g.: "*...from here*".

**AR:** Human cues for indicating real world locations within reach, including body gestures and line of sight, are familiar and intuitive. Replicating these cues in AR is not obvious, unless using a realistic virtual avatar that imitates human gestures, which represents an unreasonable approach for the considered research context. (G. Lee et al., 2019) demonstrated the benefit of using AR tips to improve spatial understanding, ranging from procedural step-by-step

instructions to simple navigation inside a building. For localizing assembly locations in AR, it is common to place virtual icons or text labels on the physical object to be located (Nassani et al., 2015). If the augmentation is outside the field of view, it is helpful to provide interface cues to direct user's attention to the location of the assembly component (G. A. Lee & Hoff, 2020). Multiple techniques for localizing out-of-view objects in AR are proposed in the literature, including the attention funnel (Biocca et al., 2006) or the virtual tunnel (Hanson et al., 2017), 3D arrows and compass-like 2D arrows (Schwerdtfeger et al., 2011), or more complex systems like the EyeSee360 technique proposed in (Gruenefeld et al., 2017). Among these, the most appropriate approach, consistent with the considered authoring constraints and HCI principles, is the usage of predefined arrows (2D or 3D). Our analysis indicates that arrow-based guidance techniques for localizing out-of-view objects in AR are adapted to the considered scenario because: (1) arrows are familiar visual cues, potentially easy to understand and follow by people without AR expertise; (2) visually, arrows are less intrusive and easier to integrate with other AR scene elements; (3) arrow implementation is straightforward.

Therefore, we propose to indicate the assembly location during training with the help of two arrows. The first is called *directional arrow* and directs the user towards the assembly location, if this is not in the field of view. The second one is the location arrow and points to the real-world assembly. As discussed in Section 5.2.2.2, the location arrow is placed by the author during the creation of the AR instruction and may represent an approximate location of the assembly. Only one of the two arrows will be visible at a given moment, the first being head-registered whereas the second is world-registered. We note that the directional arrow is added automatically at the application level, based on the position and orientation of the user and the position of the location arrow.

### 6.2.3 How

**H2H:** The instructor demonstrates **how** the operation is executed by performing and explaining the assembly process at the same time, e.g.: "...you do it like this...".

**AR:** By considering the media guidance strength analysis conducted in Chapter 4, as well as cognitive considerations and authoring cost, we argue that an FPV photo illustrating the final assembly or an FPV video demonstration (the choice between the two depending on the complexity of the assembly operation), represents the best compromise of capturing and conveying complex assembly expertise via AR. Our experiments strongly suggest that short and spatially registered (i.e., contextualized next to the assembly location) FPV demonstration videos represent the most adapted information conveyance technique for most manual assembly operations.

## 6.3 Addressing different assembly environments

Depending on the assembly context, including workstation organization and manner to perform the assembly procedure (seated, standing, or both), the information conveyance can be adapted to address such specificities, needs and user preferences.

### 6.3.1 Seated assembly environment

Seated assembly in a fixed workstation environment represents one such scenario for which the proposed information conveyance technique may require slight changes. An informal field evaluation of the proposed approach, conducted in an assembly workstation where the worker is seated (Figure 6-1), suggests that the display of the text description can be optimized.



*Figure 6-1. Informal field experiment conducted in a different manufacturing factory. The worker is required to seat during training (i.e., the execution of the assembly procedure).*

As the worker does not change his spatial position during the assembly procedure, except moving his hands and head, displaying the text description by using head registration is no longer well-suited, because this would be too intrusive and potentially occlude the real-world assembly. Even though the participants have been able to successfully follow the set of AR instructions specifically authored for the considered workstation, the user feedback indicated that it would be more suitable for the text description to be world registered, potentially as part of the visual element of the assembly illustration. For example, the text can be displayed either on the top or at the bottom of the corresponding photo or video demonstration. However, this design technique may lead to information omission and thus to assembly errors. Larger scale comparative field evaluations are still required for analyzing how the information conveyance approach should be adapted and optimized for such specific assembly environments.

### 6.3.2 Dynamic and mobile workstation environment

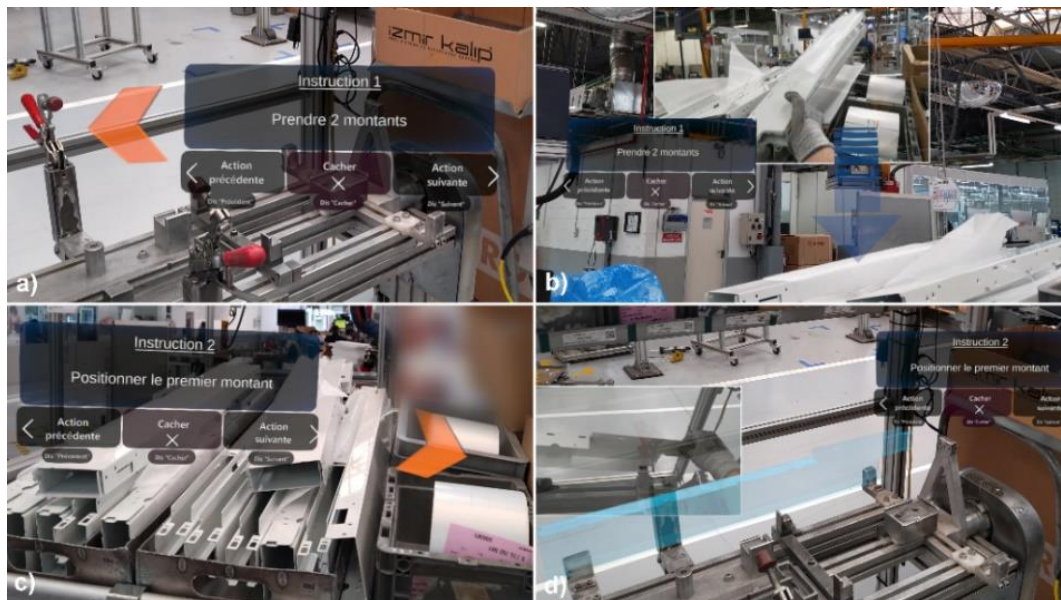
In the case of a manufacturing environment where the organization of the workstation would make the world registration by QR code unreliable and thus the media contextualization impossible, the proposed information conveyance method requires some changes. In such a scenario, the authoring procedure still can be performed in a similar manner. However, the information conveyance technique would require head rather than world registration. The experimental results obtained (Section 7.2.2) show that an information conveyance technique like the one proposed by Guides is appropriate and satisfactory from most viewpoints (i.e., performance, user acceptance, mental workload). However, the main concern is related to the identification of the real-world assembly location without the help of the spatially registered visual cues (i.e., location arrow), which in this case would not be ensured. To overcome such a limitation and remain in line with the proposed approach, during the creation of the AR work

instructions, the author would perhaps need to capture video demonstrations for most assembly operations, by clearly indicating the assembly locations. This approach would underperform by far the one where a spatial registration is guaranteed. A system able to provide automatic registration of the virtual elements with their corresponding real-world objects or locations would not only address this concern ideally, but it would also replace the fiducial registration method on which our proposal and most AR systems rely today.

## 6.4 System implementation

The information conveyance approach presented in the previous section is implemented with the help of a HoloLens 2 application specifically developed to this purpose. Let us focus on how the information is presented and how the trainee (i.e., the shop floor worker) is supposed to interact with the various visual elements involved during the training procedure.

Figure 6-2 illustrates the user interface and the different visual elements of an AR work instruction conveyed during a training procedure. Note that Figure 6-2.d) illustrates the usage of a CAD model, a replacement of the location arrow. This technique is discussed in Section 5.2.3.1 and represents a potential extension of the proposed system.



**Figure 6-2.** Example of how the assembly information is conveyed via AR. **Instruction 1:** “Grab 2 uprights”. a) Text & indication arrow; b) Location arrow & FPV photo; **Instruction 2:** “Place the first upright”. c) Text & indication arrow; d) FPV video & CAD model.

**What:** Each task starts by displaying a text instruction panel (Figure 6-2. a)) in front of the trainee, between 0.6 to 0.7 meters away. The instruction panel follows the trainee’s head for 1 second (head registration) then it stops (environment registration). By using this technique, we try to ensure that the text instruction is not overlooked and at the same time, that the panel does not visually interfere for more than necessary. The *sticking time* of 1 second is adjusted for the considered assembly use case, based on the required movements during the assembly procedure. The trainee hides the panel by clicking a “hide” button or by using the voice command “hide”. Complementing the “hide” button with a voice command was necessary for

cases where the panel is rendered behind the physical environment, and thus unreachable to hand touch. The contrary is true for noisy assembly environments, where voice may not be suitable for interacting with the system (Jasche et al., 2021). Our use case validates the requirement of multimodal interfaces discussed in (Irawati et al., 2006).

**Where:** The next step consists in identifying the assembly location, pointed to by a spatially registered arrow (Figure 6-2. b)). If the assembly location is not in the FoV of the trainee, a screen-registered directional arrow (Figure 6-2. a) and c)) will guide him towards it (Section 6.2.2).

**How:** Once the assembly location identified, the trainee should observe in addition to the location arrow, the corresponding FPV photo of the final assembly or the video demonstration (Figure 6-2.d)). Capturing these elements is optional during authoring, and generally they are not provided for simple assembly operations. If any of them exists however, they should be visible (at least partially) in the same field of view as the corresponding assembly location. The trainee interacts then with these elements and executes the corresponding assembly operation.

The trainee visualizes the next/previous instruction by clicking the “*next*”/“*previous*” button or by using the corresponding voice command. The “*help me*” voice command brings the instruction panel in front of him. We note that unlike (G. A. Lee & Hoff, 2020), our proposal makes use of text and photos, in addition to video and indication arrows. We use predefined arrows as AR visual cues, spatially registered during the authoring by using instinctual interaction (*Mixed Reality Instinctual Interactions | Microsoft Docs*, 2020). In our approach, the FPV instructional video is presented during training exactly as captured in the authoring procedure.

## 6.5 Interaction techniques

The interaction techniques employed by the training interface is conceived to allow novice workers follow the AR instructions in the least intrusive manner, while aiming to stimulate the user focus on the assembly rather than the application usage. Within this context, we have adopted multiple natural interaction techniques, including eye gaze, for controlling the video playback and head gaze, for automatically rotating the videos/photos towards the trainee.

**Speech and touch.** The action of iterating between the AR instructions can be achieved by either a touch interaction with the text panel (finger touch) or by using a vocal command. We complement interaction techniques for executing the same action to address different assembly environment characteristics for which unimodal interaction may not be adapted. An example would be using voice in noisy environments or hand touch where movement is restricted. In addition, by providing multimodal interaction, we aim to address user preferences as well.

**Head gaze.** We use head gaze, an implicit interaction technique, to contextualize the displaying of the text description and to orientate the assembly illustration (i.e., photo or video) based on the trainee’s position. We propose this technique mainly to facilitate the access to information during training, without requiring the trainee to move to certain physical locations for improving his viewpoint. This feature is useful as well for cases where the author of the AR instructions and the trainee have a significant height difference, in which case a good viewpoint may not even



be possible to achieve. By employing such techniques, we aim to address human-centered design principles, while providing hands-free interaction and avoiding UI clutter. We encourage the trainee as well to focus on the assembly process and not on the application usage.

Regarding the text description, this is a punctual information that should not be omitted by the trainee, as it marks the beginning of a new AR instruction, and it summarizes the objective of the current assembly operation. At this stage, the position and orientation of the trainee cannot be known. For this reason, by using head registration for displaying the text panel in front of the trainee eventually forces him to notify and read the instruction. Displaying this information otherwise, would potentially require intentional searching (if out of the FoV of the trainee) and focus on the text description, leading to longer training time and an increase in mental/physical effort as well. The second utility of the head gaze is the automatic orientation of the contextualized captured media, always facing the trainee. This functionality ensures a direct, frontal vision on the corresponding media, independently on trainee's position and orientation.

Finally, let us note that the quality of the video can have an important impact on the comprehension of the assembly operations described in the video demonstration. The lighting conditions of the considered factory and the quality of the videos recorded by HoloLens with default camera parameters have proven to be sufficient during our experiments.

**Eye gaze.** We use eye gaze to control the video playback during training for the following reasons. Firstly, this implicit, natural video playback technique allows the trainee to start and stop the video without requiring specific, deliberate action (e.g., button touch or voice command). Secondly, depending on his skills, the trainee can synchronize his assembly work and, at the same time, align the viewpoint of the real-world assembly components with the ones presented in the instructional video for a better understanding, by effortlessly watching short, consecutive chunks of the instructional video. Finally, this playback technique potentially optimizes the assembly time as, ideally, the trainee watches the video only once, while performing the assembly at the same time. This approach is consistent with the recommendations suggested in (Biard et al., 2018), who demonstrate that users perform better when videos are automatically paused between steps. By using eye tracking, we allow the trainee to pause the video whenever he feels the need of switching his attention from the video to the assembly itself.

In addition to providing an effortless video playback interface, this technique allows the trainee to perform (two-hand) discrete assembly tasks while glancing at the video demonstration, encouraging him to perform the assembly work during the video demonstration itself. By using different interaction techniques for specific media contents as exemplified above, we seek to provide a multimodal interface, faster and easier access to information while addressing cognitive and physical aspects and speed up the overall training procedure.

A summary of the interaction techniques retained is presented in Table 6-1. In addition, Table 6-2 summarizes the UI and the adopted information conveyance technique. We use here the Frame of Reference (FoR) notation for referring to the two registration methods considered: Screen-Fixed (SF) and World-Fixed (WF) (Gabbard et al., 2014).

**Table 6-1. A summary of the interaction techniques used during training.**

Interaction technique	User input	Output
<b>Speech</b>	<i>next, previous</i>	Goes to the next / previous instruction
	<i>help me, hide</i>	Shows / hides the text instruction (panel)
<b>Touch</b>	Touch virtual elements (buttons)	Hides the text instruction by touching the <i>hide</i> button with the fingertip; goes to the next / previous by touching the corresponding button
<b>Head gaze</b>	Implicit interaction	Controls the position and orientation of the text panel and the orientation of the media elements (photo or demonstration video)
<b>Eye gaze</b>	Implicit interaction	Controls the video playback

**Table 6-2. Training UI and information conveyance.**

2W1H	Media type	FoR	Information	User action
<b>What</b>	Text instruction	SF / WF	Briefly describes the assembly operation	Reads text, then hides or ignores the panel
<b>Where</b>	Indication arrow	SF	Guides the user toward the assembly location	Turns the head towards the indicated direction
	Location arrow	WF	Indicates the assembly location	Identifies the location
<b>How</b>	Photo / video	WF	Illustrates the assembly action	Performs the assembly

## 6.6 Key functional and design aspects

Let us now summarize the key functional and design aspects of the proposed training approach.

**Information formalization.** In line with the authoring, the assembly information is presented in a 2W1H-based formalized manner, independently on the assembly type or complexity. We aimed to address the need of assembly standardization in AR, providing at the same time a simple to follow yet consistent information conveyance method that potentially supports a fast comprehension and learning of the training system by non-technical people.

**Information contextualization.** Both head and world registration are used for contextualizing the visual elements that convey the assembly information during training. The various visual elements involved are optimally displayed in the FoV of the worker, not only to facilitate the access to information in AR, but also to reduce mental workload/physical stress, while supporting a faster assembly as well.

**FPV video demonstration as central element of the proposed training approach.** Perhaps surprisingly, the key element of our proposed AR training approach is not represented by 3D content. Our proposed information conveyance method consists of three visual elements, none of them being a 3D content. The importance of these elements depends mainly on the assembly type and complexity. For example, a brief text description complemented by a location arrow is sufficient for describing a simple picking operation, while the importance of the text diminishes

drastically in the favor of a video demonstration for more complex assembly operations. As remarked during our case study, most assembly operations are not obvious, even for experienced workers. For this reason, the video demonstration, complemented by the location arrow, becomes the key elements of our proposal.

Finally, let us summarize how the guidelines and design goals identified in Section 4.4 are addressed in the proposed AR information conveyance approach:

1. Automate play/pause when possible: use eye tracking for natural video playback control. No authoring effort required.
2. Easy access to important frames: the creator of the video demonstration should consider the authoring recommendations by capturing short video demonstration sequences, per assembly task. This removes the need of navigating the video.
3. Visual highlights: the author oversees providing quality instructions, highlighting subtle assembly details during the assembly demonstration. This requirement is not dealt with at the application level; therefore, it is up to the author to follow this recommendation.
4. Low transaction cost (switching back and forth between the video and the target application – the assembly in our context): we address this by contextualizing the visual elements during the authoring in such a way that the virtual content is ideally placed in the same field of view with the real-world assembly location.
5. Provide procedural instruction rather than conceptual information: the video demonstration of an assembly task is, by definition, a procedural instruction.
6. Keep video segments short: the author is recommended to demonstrate a single assembly operation per AR instruction; therefore, the video length is expected to be limited to a single assembly task.
7. Ensure the tasks are clear and simple: addressing this recommendation is up to the author of the AR instructions. We enforce this suggestion, to a certain extent, by proposing the 2W1H authoring principle together with the recommendation of describing one assembly operation per AR instruction.
8. Coordinate demonstrations with textual documentation: we address this recommendation by conveying the assembly information in a manner that follows the 2W1H principle (i.e., text description then assembly illustration).
9. Use spoken narration: the author can choose to verbally describe the assembly operation during the recording of the video demonstration.
10. Strive for universal usability: the authored AR content, including the video demonstrations do not require post-processing and they can be reused as they are by other information systems.

# Chapter 7. EVALUATIONS

---

---

---

This chapter describes the extensive evaluation of both authoring and training approaches proposed. The first part of the chapter describes two field experiments, denoted by T-FE1 and T-FE2, conducted for evaluating the training part of the system. T-FE1 is a preliminary field experiment that aims to assess the proposed training system from multiple perspectives including usability, error rate and completion time. One of the main objectives of this study is to determine whether low-cost visual assets are sufficient for visually describing complex assembly operations in an industrial setup and for novice workers. The second field experiment (T-FE2) is an extension of the first and has as main objective to assess the utility of using CAD data for conveying assembly expertise, as a complement to the proposed low-cost approach. A comparative evaluation is finally conducted to objectively evaluate the potential benefits of CAD-complemented AR instructions versus low-cost-only AR instructions.

The second part of the Chapter concerns the evaluation of the AR authoring methodology proposed. The first field study, denoted by AT-FS, has been carried out at elm.leblanc and proposes a comparative evaluation of the proposed AR authoring system versus Guides, from a time perspective. The objective here is to provide an estimation of the time required to authoring the AR instructions of an assembly workstation. The second experiment, denoted by AT-LE, has been conducted in a laboratory setting because of organizational concerns. It proposes a complete evaluation of the proposed AR training system, with both authoring and training procedures, in comparison with Guides.

T-FE1 has been published in (Lavric et al., 2021b). T-FE2 has been published in (Lavric et al., 2021a). AT-FS has been published in (Lavric et al., 2022a). Finally, AT-LE has been published in (Lavric et al., 2022b).

---

---

## 7.1 Training experiments

The two field experiments have been carried out in the elm.leblanc boiler-manufacturing factory. The first field experiment is a preliminary one and evaluates the usability and effectiveness of low-cost visual assets for conveying manual assembly instructions. The second experiment, an extension of the first, has as main objectives to (1) assess the potential benefits of authoring CAD-based AR work instructions and to (2) validate the hypothesis that the proposed low-cost-based training approach is potentially the most adapted technique for conveying assembly information in similar industrial use cases.

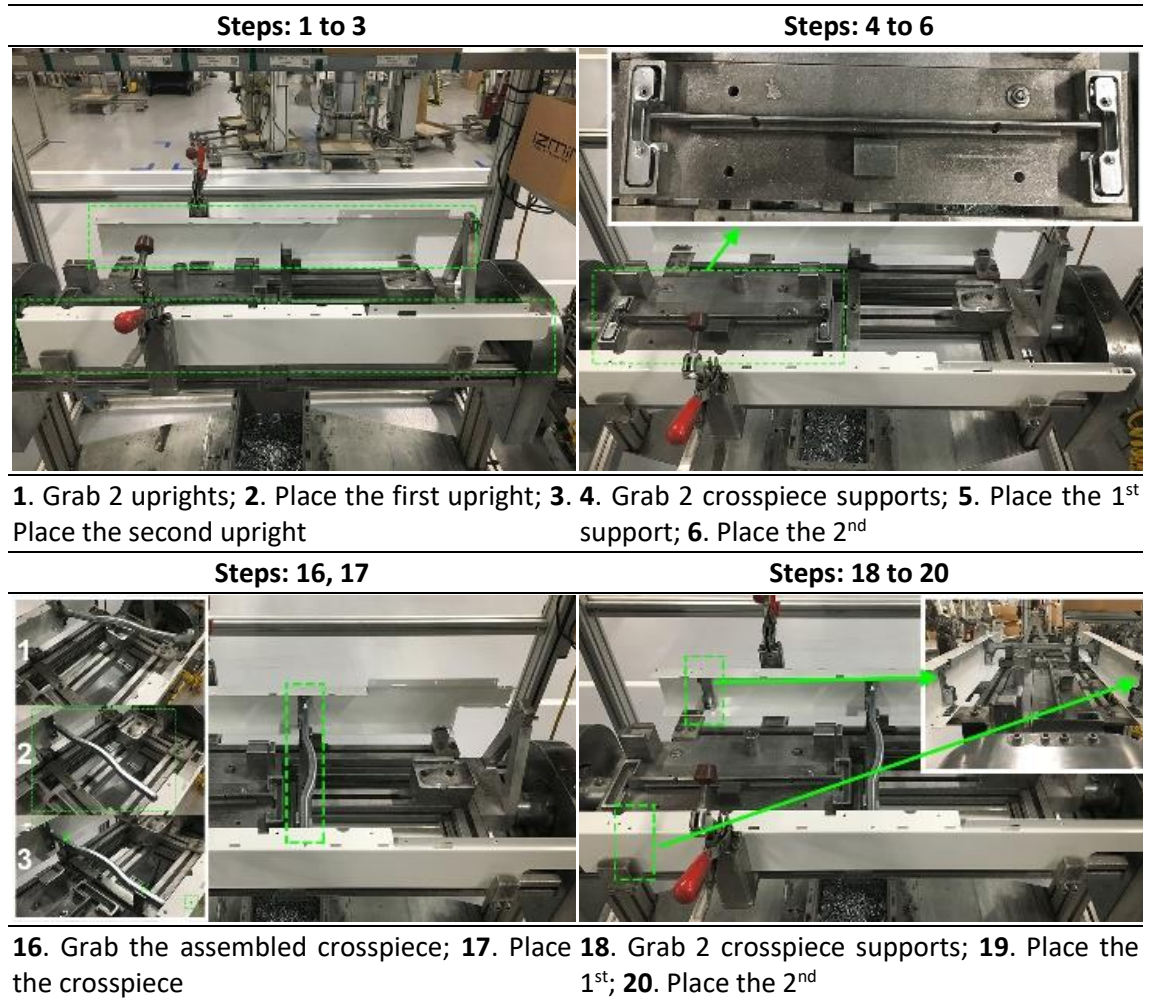
Both experiments concerned the assembly of a boiler frame and have been conducted under the same assembly setup. More precisely, they concern the first workstation, denoted by W201 (Chapter 2, Figure 2-2), of the assembly line L2 at elm.leblanc. They consist in performing one or multiple (within approximately 30-minutes period) assembly cycles (comprised of 38 assembly operations), guided with step-by-step work instructions in AR. We chose workstation W201 because its assembly tasks are relatively safe for novice workers, quite diversified and because the assembly itself is performed on an adjustable (i.e., rotating up-down and left-right) structure, which increases the complexity of the use case. The workstation on which we have performed the experiment has not been adjusted in any way for the specific purpose of the evaluation. Figure 7-1 illustrates the final assembled product used during the considered training experiments, while Figure 7-2 shows the most relevant assembly operations performed.



**Figure 7-1.** T-FE1 & T-FE2: the final product assembly obtained by participants during the two training field experiments.

The assembly operations, denoted by ATx, are divided into four types:

- 14 x AT1 – picking (assembly components and tools)
- 8 x AT2 – installing / placement (assembly components)
- 12 x AT3 – screwing & riveting (screws and rivets)
- 4 x AT4 – manipulating (assembly structure and tools)



**Figure 7-2. T-FE1 & T-FE2: the most relevant operations of the assembly procedure used during the training field experiments.**

We used ATOFIS to author two different sets of AR instructions. The first set, evaluated in T-FE1, was based solely on low-cost visual assets. The second instruction set was identical with the first, except that CAD models replaced the location arrows in assembly instructions of type AT2. We note that every AT2 instruction has a FPV demonstration video associated to it in the first instruction set, complemented with a CAD model in the second. There is no potential benefit in complementing with CAD models the other assembly types (AT1, AT3 and AT4), reason for which these instructions are the same in the two instruction sets.

### 7.1.1 Low-cost instructions set training (field experiment 1 - T-FE1)

12 participants, aged between 20 and 60, with different profiles (students, engineers, and researchers) have participated to this experiment. 10 of them have no or solely basic assembly experience, while 2 are proficient in assembly tasks. Similarly, 10 of them have none or basic AR knowledge while 2 have good or advanced AR knowledge. Among them, 7 participants have already used an AR headset before.

#### 7.1.1.1 Evaluation procedure

For each of the participants, the evaluation procedure has been conducted as follows (with some very minor adjustments based on the user profile):

- 15-20 minutes discussion to present the goal of the application and perform the eye calibration in Hololens 2 (required for a better AR experience but also because statistics are collected by using the eye tracking function of Hololens 2). The users learn how to interact with virtual buttons.
- 30 minutes in front of the workstation, where the subjects performed at least one training cycle, up to 4, depending on their skills. Each subject was limited to 30 minutes because of the workstation availability.
- 5-10 minutes discussion next to the workstation to debrief the experience and collect first remarks.
- filling in the SUS (*System Usability Score*) questionnaire for evaluating the usability of the application and a specific one designed by ourselves for sketching user profiles and collecting their feedback.
- an open discussion to analyze the experience in detail and ask for clarifications if needed, concerning the answers given in the three questionnaires.

In total, a complete user evaluation lasted from one hour to one hour and 40 minutes.



**Figure 7-3. T-FE1: a participant executing an assembly during the training field experiment.**

The shop floor training experiment has been conducted as further described. Before starting the actual assembly, the participant was asked to get ready by putting security shoes on, gloves and a bouffant cap. Next, the examiner handed the Hololens 2 and asked the participant to start the training application. Before visualizing the assembly instructions, the participant was required by the application to authenticate himself and select the workstation by scanning the unique QR code associated to it. This procedure is not part of the training evaluation, therefore not measured. Once the first AR instruction displayed, the examiner explained for the last time the application usage. During the training procedure, the examiner was always in the proximity of the participant to ensure his safety, to take notes or to stop him if an error is committed. For each assembly error, the participant was briefly interrupted, and the examiner performed the assembly while briefly describing it. The participant continued with the next assembly instruction soon after.

### 7.1.1.2 Evaluation metrics

We have adopted the two evaluation methods identified in (X. W. S. K. Ong & Nee, 2016) in their AR assembly research survey, which concern effectiveness and usability. The effectiveness of the proposed training method is assessed by measuring the error rate, the assembly completion time (**ACT**) and the instruction reading time (**IRT**). ACT represents the time spent for completing an assembly task, while IRT represents the time spent on reading the low-cost visual assets.

The first evaluation was a quantitative one and measured:

- The number of errors committed by the participants: measured by the evaluator,
- The completion time for each assembly operation: measured by the application,
- The time spent on reading the AR instructions: measured by the application by using eye tracking.

We assess the training performance by measuring the number of errors and the time required by the participants to complete the assembly procedure over 3 cycles. We consider that comparing the assembly time of novice workers with the nominal one (of experienced workers) is irrelevant.

The second evaluation metric is a qualitative one and measures the usability of the proposed training method by using the SUS questionnaire (*System Usability Scale (SUS) | Usability.Gov, n.d.*). In addition, to identify correlations between performance, visual assets, assembly task complexity and user profile, we have designed our own specific questionnaire.

### 7.1.1.3 Experimental results and interpretation

Each participant was asked to perform up to 4 assembly cycles, within a total time budget of 30 minutes. An assembly cycle represents the completion of a set of operations corresponding to the considered assembly workstation.

In total, the participants executed 29 assembly cycles during all training procedures. The reported evaluation results are described further.

#### 7.1.1.3.1 Error rate

Table 7-1 presents the error rates committed by the participants during the 4 assembly cycles.

**Table 7-1. T-FE1: error rates reported during the AR training.**

Assembly cycle number	1	2	3 & 4
Participants performing the n <sup>th</sup> cycle	12	9	8 (6 + 2)
Participants committing errors (%)	75%	33%	0%
Number of errors per cycle	12	3	0
Error rate (%)	2.63%	0.87%	0%

Our measurement reported that 75% (9 out of 12) of the participants committed one or two errors during the first assembly cycle. However, we observe that the average error rate of 2.63% is very low, a value which potentially suggests a lack of user attentiveness rather than an issue regarding the AR information conveyance method. The convergence of the error rate to zero and the progress of the ACT over the course of three assembly cycles demonstrate the usability



and effectiveness of the proposed training method. We note that all the reported errors except one consisted in a wrong orientation of the assembly component, supporting thus the effectiveness of the proposed AR training method, notably for tasks of type AT1, AT2 and AT4.

The post-evaluation feedback confirmed that some testers did not pay attention to some details explained in the instructional videos. The high number of subjects committing errors during the first assembly cycle indicates that a better technique is needed to highlight details in photos/videos. The participants agreed that a remark next to the text description may be sufficient to draw the attention over important assembly details. The convergence of the error rate to zero after 3 cycles is a potential indicator of the effectiveness of the training method.

#### 7.1.1.3.2 Assembly completion time

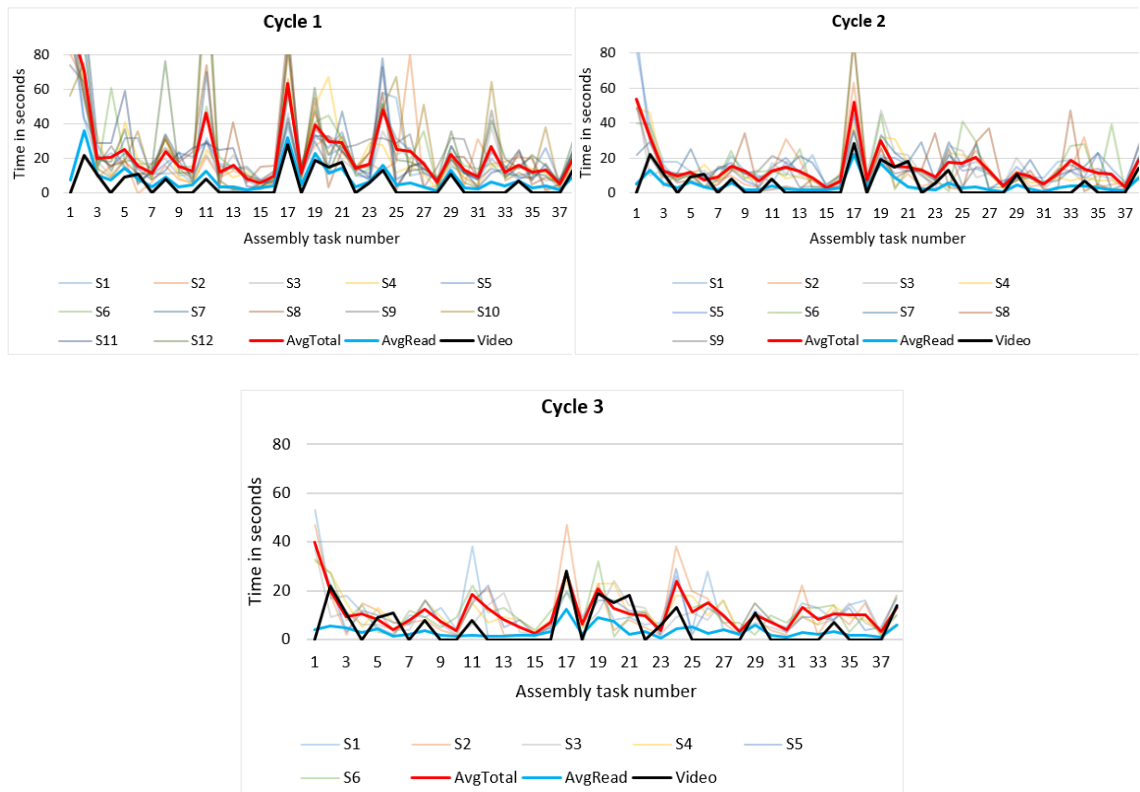
Table 7-2 presents assembly and instruction reading time measurements. The IRT is measured by using eye tracking and represents the total time spent by the participant to read each AR instruction. It includes the reading time of the text description, summed up with the time the participant focuses on the photo or the video demonstration.

**Table 7-2. T-FE1: assembly time measurements reported during the AR training.**

Assembly cycle number	1	2	3	4	1 to 4
Total assembly time (seconds)	899s	555s	421s	354s	
Progress from the previous cycle (%)		38%	24%	15%	67%
Reading time and percentage from the assembly time	315s 35%	169s 31%	125s 30%	82s 23%	34%

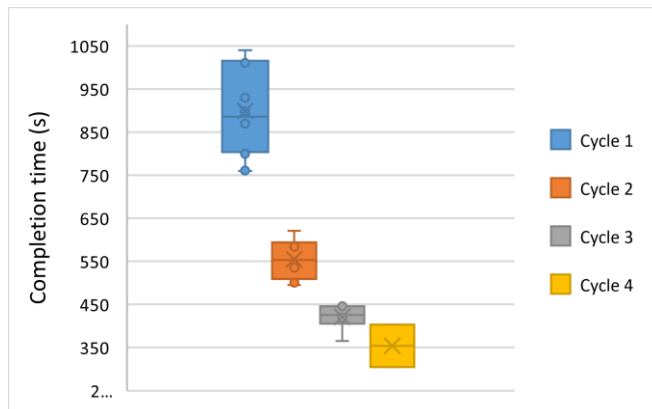
In a complementary manner, Figure 7-4 illustrates the time required by each participant to complete each operation over the first 3 assembly cycles. The 4<sup>th</sup> cycle is not illustrated as the number of participants which managed to execute it was very low (2). We note the average time spent by the participants to read the instructions, *AvgRead* (blue line), matching closely the video length, *Video* (black line), during the first cycle. We observe as well that *AvgRead* is flattening over the course of the 3 cycles. The peaks of *Video* over *AvgRead* in the third graphic indicates that participants stop watching the entire video at this point. This potentially indicates that after only 2 assembly cycles, the video can be replaced by photos or even removed. The flattening of *AvgTotal* (red line), indicates the learning progress and the familiarization with the assembly operations.

The results show that the participants get familiarized and recall the assembly instructions at a very fast pace: the time spent on reading the AR instructions decreases by 60% in the third assembly cycle, indicating that the utility of the AR instructions is rapidly diminishing. The worthiness of authoring AR instructions by using CAD data, animations and other expensive media that would make the authoring procedure more laborious is therefore already questioned. We also observe that the time spent on reading the instructions starts at 35% of the total assembly time and decreases by 34% (to ~23 %) in the last cycle, indicating a diminishing utility of the AR instructions due to assembly recall and familiarization.



**Figure 7-4. F-FE1: average assembly time (AvgTotal) and reading times per instruction (AvgRead), for a) Cycle 1, b) Cycle 2 and c) Cycle 3.**

Figure 7-5 is a *box and whisker chart* that illustrates the average assembly time per each of the four cycles. We note that the 4 data points (i.e., circles) within each box in the figure represents the data distribution into quartiles (i.e., the lowest 25% of the reported times, the next 25% - up to the median, the second highest 25% and finally the highest 25%). The lines which extend vertically (i.e., the whiskers) indicate variability outside the upper and lower quartiles, that is the outliers. Finally, the mean value of each series is indicated by the cross in each of the box. The time progress illustrated in the figure indicates that participants become familiarized with the assembly operations after one cycle. The variance and the gap between the cycles suggests that the participants are in an exploration mode during the first assembly cycle. This is a potential explanation of the lack of attentiveness of some participants, which led to assembly errors.



**Figure 7-5. T-FE1: box and whiskers chart illustrating the assembly times for each of the four cycles.**

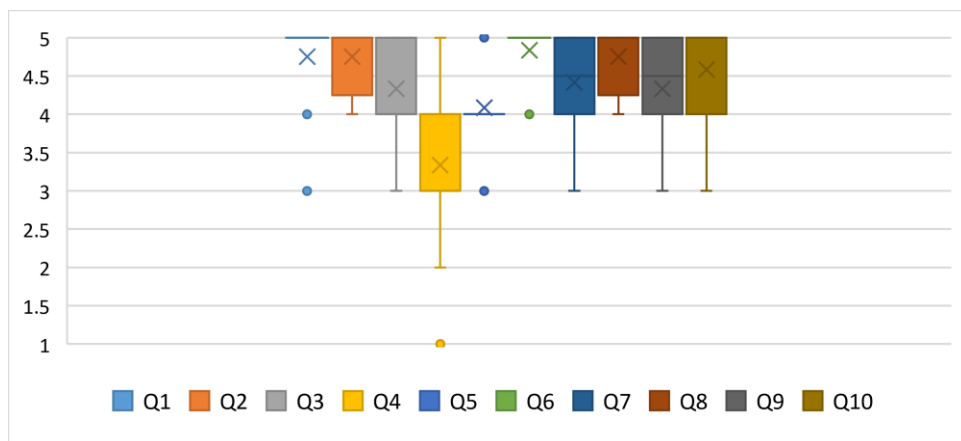
### 7.1.1.3.3 System Usability Evaluation

We used the System Usability Scale questionnaire to evaluate the overall usability of the proposed training method, by using a five-item Likert scale ranging from *strongly agree* (5) to *strongly disagree* (1). The 10 questions defined in the SUS are recalled in Table 7-3.

**Table 7-3. The System Usability Scale questionnaire.**

1.	I think that I would like to use this system frequently.
2.	I found the system unnecessarily complex.
3.	I thought the system was easy to use.
4.	I think that I would need the support of a technical person to be able to use this system.
5.	I found the various functions in this system were well integrated.
6.	I thought there was too much inconsistency in this system.
7.	I would imagine that most people would learn to use this system very quickly.
8.	I found the system very cumbersome to use.
9.	I felt very confident using the system.
10.	I needed to learn a lot of things before I could get going with this system.

Figure 7-6 is a box and whiskers chart which illustrates the raw SUS reported scores, with a highlight on the mean value and the outliers. The overall SUS result is 88.33. We note that the standard formula (Brooke, 2020) was used to compute this score, i.e., normalized on a scale from 0 (negative) to 100 (positive), with negative items using inverted scale. All the participants that ranked the question Q4 with a normalized SUS result inferior to 60 claimed that they might need human supervision at the beginning of the training procedure, further referred to as the workstation exploration cycle (WEC). This remark is supported by the reported error rate, which also suggests that human supervision during the first cycle might be required. However, the participants committing errors claimed that the instructions were clear and complete except that subtle assembly operations should be better highlighted. The same user remark was reported in a very similar study that compared different types of AR visualizations, where multiple participants using FPV demonstration video guidance suggested for additional highlights of important assembly details with marks or zooms (Jasche et al., 2021).



**Figure 7-6. T-FE1: box and whiskers chart illustrating the raw SUS scores reported by the participants.**

The findings of the present experiment suggest that spatially registered 2D visual assets together with a specific, human-centered set of interaction and visualization techniques provide an effective AR-based conveyance method for describing manual assembly operations in an industrial context. We note however that the reported error rate during the WEC suggests that a better technique for describing error-prone assembly operations is required, particularly for novice workers and during the WEC.

### 7.1.2 Comparative evaluation: low-cost *versus* CAD-complemented instruction set training (field experiment 2 - T-FE2)

The field experiment presented in this section has the following objectives:

- (O1) Extend and demonstrate the findings of T-FE2,
- (O2) Identify potential benefits of using CAD models for manual conveying assembly information by comparing the usability and effectiveness of two sets of AR instructions: low-cost *versus* CAD-complemented,
- (O3) Answer a research question suggesting that studies are needed to identify optimal ways to convey instructions in industrial sectors via AR (Gattullo, Evangelista, et al., 2020), (Gattullo, Evangelista, et al., 2020),
- (O4) Measure the perceived usability of the system and the mental workload of the participants.

Two instructions sets have been created and submitted for evaluation, a first one, denoted by LA (Low-cost Assistive-based instruction set) including solely low-cost visual assets (text, simple 2D graphics, photos, and videos), and a second one integrating complementary 3D CAD data and denoted by CAD (CAD-based instruction set). Let us recall that the CAD data concerns solely the AR instructions of the assembly set type 2 (AT2).

#### 7.1.2.1 Participants

We have involved 30 participants from the factory, 22 males and 8 females with a wide variety of age, education levels and professional positions, from line workers to engineers and managers. We have created two groups, denoted by **G-LA**, and **G-CAD**, each composed of 15 participants, for evaluating the two instruction sets. Five participants have assembly experience in each group. We created two subgroups for each group: **G-LA-N** = novice participants from G-LA and **G-LA-E** = experienced participants from G-LA. Similarly, for G-CAD: **G-CAD-N** and **G-CAD-E**. We have grouped the participants as such, to identify if assembly experience has a notable influence on the training performance. Table 7-4 summarizes this information.

**Table 7-4. T-FE2: participant division by instruction set (low-cost vs. CAD) and assembly experience (novice vs. experienced).**

Group	G-LA		G-CAD	
	G-LA-N	G-LA-E	G-CAD-N	G-CAD-E
Participants no.	10	5	10	5
Assembly exp.	No	Yes	No	Yes
Instruction set	1		2	

We assess the effectiveness and the usability of the proposed training system in the same manner as described in T-FE1, per instruction set and per subgroup, to address the main four objectives. In addition, we measure the mental workload by asking participants to complete the NASA-TLX questionnaire (Hart, 1980).

#### 7.1.2.2 Experimental results and interpretation

Table 7-5 presents the number of participants, per group, performing the  $n^{\text{th}}$  assembly cycle (78 in total). For each assembly cycle, we present the percentage of participants committing errors, the average error rate per instruction set, the total ACT (assembly completion time) and IRT (instruction reading time, expressed in terms of % of the ACT), and finally the average ACT of tasks of type AT2. We measure the IRT to identify differences and to estimate the utility of low-cost visual assets over multiple assembly cycles, in both instruction sets. A comparison between the two subsets of type AT2 is performed separately. Table 7-5 presents the results obtained for all instructions, to observe the impact of the CAD-based instructions over the whole set.

**Table 7-5. T-FE2: reported evaluation results – i.e., error rate, assembly completion times and instruction reading times, for each cycle of the two considered groups.**

Group	G-LA			G-CAD		
	Cycle no.	1	2	3	1	2
Participants no.	15	12	8	15	15	9
Participant error rate (%)	66%	25%	0%	66%	20%	0%
Total number of errors	13	3	0	18	3	0
Error rate per set (%)	2.2%	0.6%	0%	3.1%	0.5%	0%
Avg. ACT (s)	884s	538s	367s	838s	475s	336s
ACT progress ( $n^{\text{th}}-1$ )		39%	31%		43%	29%
Avg. IRT (%)	37%	29%	25%	31%	27%	19%
Avg. ACT of AT2 (s)	290s	165s	98s	268s	130s	74s

The global results presented in Table 7-5 show the following trends:

- in both cases, three cycles seem sufficient for achieving the training objectives, with a zero-error rate and reduced ACT and IRT factors,
- the LA approach leads to lower error rates when compared to CAD, particularly during the first assembly cycle,
- the best ACT and IRT values are provided by the CAD approach.

These results are in a certain sense surprising, since we would expect that 3D CAD data would offer enhanced visual guidance (and hence lead to lower error rate), at the cost of an additional time (both ACT and IRT) for understanding/handling such more complex information.

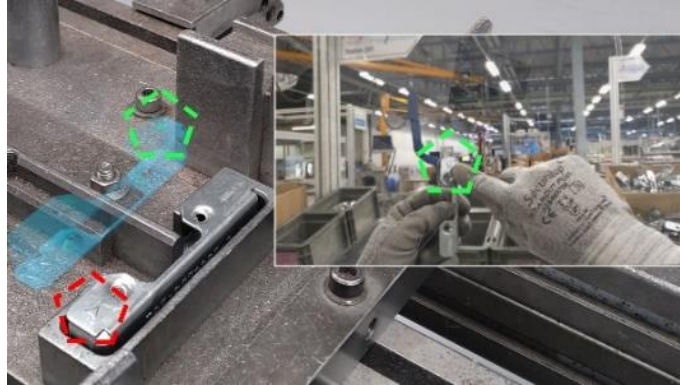
Let us analyze in more details the various indicators obtained.

##### 7.1.2.2.1 Error rate

During the first two cycles, all the errors have been committed on tasks of type AT2 and concern seven out of the eight instructions of type AT2. With a single exception, all errors consist in a wrong installation (erroneous orientation) of the assembly component. Some few riveting and screwing mistakes have been committed because of the lack of assembly experience. These

were detected and fixed without the intervention of the evaluator and therefore not accounted as errors.

Figure 7-7 illustrates an example of an error-prone assembly, where a symmetrical assembly part was incorrectly (i.e., upside-down) placed by the participant during training. The visual cue (i.e., arrow) that indicates the orientation of the component is highlighted on the CAD model and in the video demonstration by the green mark, and in the assembly executed by the participant by the red mark.



**Figure 7-7. T-FE2: Error-prone assembly example: participant places a symmetrical assembly component incorrectly during the training – i.e., upside-down.**

The green marks highlight the arrow, presented on the 3D model and in the video, which indicates the orientation of the component. However, some participants fail to notice this detail. It seems that subtle assembly details are prone to be overlooked, especially by participants without assembly experience, which commit more errors.

A comparison of the error rates of expert and novice groups is presented in Table 7-6. During the first assembly cycle, further referred to as the **Workstation Exploration Cycle (WEC)**, we observe that some participants are guessing the orientation of the component, as the assembly structure allows the placement of the component with an incorrect orientation. An informal evaluation of all assembly tasks of the line shows that subtle assembly operations are frequent. Therefore, a better visual modality is needed to highlight key assembly details, notably during the WEC. To address this concern, participants suggested to detail the text instructions, a proposal in agreement with the authoring constraints and with our design guidelines. Table 7-6 presents the average error rate committed per participant in each subgroup during the WEC.

**Table 7-6. T-FE2: reported error rates per group and subgroup during the WEC.**

Group	G-LA		G-CAD	
	G-LA-N	G-LA-E	G-CAD-N	G-CAD-E
<b>Total errors</b>	12		18	
<b>G-LA vs G-CAD</b>	+50%			
<b>Avg. errors per participant</b>	1	0.4	1.4	0.8
<b>Experienced vs. novice</b>	-60%		-42%	

Without surprise, participants with assembly experience commit fewer errors in both groups, by 63% and by 42% in the G-LA and G-CAD, respectively. However, a one-way analysis of variance (i.e., ANOVA) reveals no significant difference between the number of errors committed by novice and experienced participants [ $F(1,29)=3.11$ ,  $p=0.08$ ]. We note that the two groups do not have equal sample sizes, as the number of participants is different – i.e., 20 novices and 10 experienced. More unexpectedly, the G-CAD group commits more errors than G-LA (by 50%). Similarly, a one-way analysis of variance reveals no significant difference between the error rates between the two considered groups [ $F(1,29)=1.47$ ,  $p=0.23$ ]. The IRT measurement reveals that during AT2 instructions, G-CAD watched the instructional videos less than G-LA, relying therefore more on the CAD information. This potentially explains their higher error rate during the WEC. The reported measurements indicated that G-CAD used 29% less time for watching videos, leading to a 7% decrease in the ACT, but to an increase in the error rate by 50%.

The error rate convergence to zero after three assembly cycles support the hypothesis that both instruction sets are reliable for conveying manual assembly information in the considered use case. However, for assembly tasks of type AT2, the evaluation results suggest that human supervision may be necessary during the WEC. Our assumption, considering the evaluation results, is that a slightly better technique for highlighting subtle assembly details will lead to an error rate converging to zero, for both instruction sets, even during the WEC.

#### 7.1.2.2.2 Assembly Completion (ACT) and Instruction Reading (IRT) Times

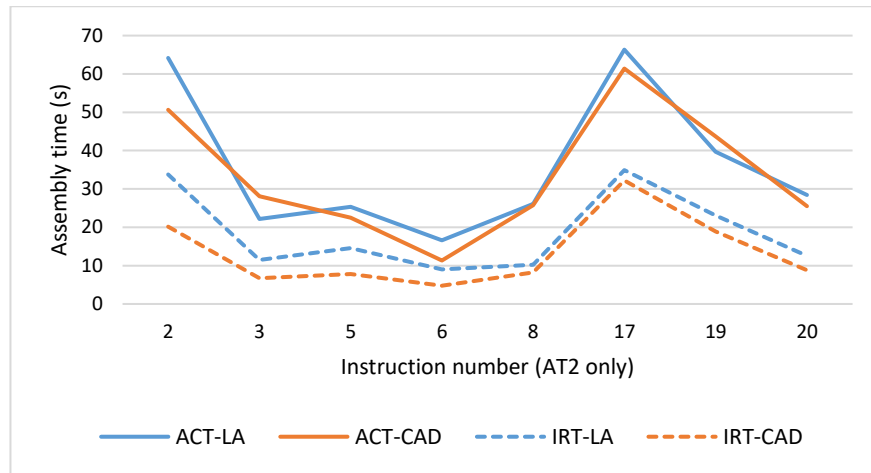
We discuss the ACT and the IRT of the WEC and of all cycles separately. For the WEC, we compare the ACT and the IRT of the two AT2 subsets, as the other instructions are the same in both sets.

**First assembly cycle (WEC).** Table 7-7 presents the reported ACT and IRT of the WEC, per group. We observe that G-CAD performed slightly faster (by 4.96%) than G-LA. However, a one-way analysis of variance reveals no significant difference between the two groups [ $F(1,29)=1.16$ ,  $p=0.28$ ]. Further, we found that G-CAD used 18% less time for reading the visual assets than G-LA. A one-way analysis of variance shows a significant difference between the IRT of the two groups [ $F(1,29)=5.55$ ,  $p=0.025$ ]. However, the overall ACT was decreased by only 5%.

**Table 7-7. T-FE2: average ATCs and IRTs reported per group, on the entire assembly set (ATx) and on the CAD-complemented instructions (AT2), during the WEC.**

Dimension (no. of tasks)	Group		Difference
	G-LA	G-CAD	
ACT for ATx (38)	884s	838s	-5%
IRT for ATx	328s	268s	-18%
ACT for AT2 (8)	295s	281s	-7%
IRT for AT2	152s	107s	-29%

A visual representation of the ACT and IRT of assembly tasks of type AT2 is presented in Figure 7-8. We observe a constant lower value of IRT-CAD in comparison to IRT-LA, potentially indicating that participants preferred CAD-based guidance to videos. G-CAD participants used 29% less time for watching videos, leading to a 7% decrease in the ACT. However, let us recall (Table 7-6) that CAD-complemented instructions led to an error rate increase of 50%.



**Figure 7-8. T-FE2: Reported ACT and IRT scores for CAD-complemented instructions (of type AT2) during the WEC.**

**All assembly cycles (3) for all participants.** Table 7-5 shows that the overall ACT decreases, on average for both groups by 41% in the second assembly cycle and by 30% in the third one. The mean (M) and the standard deviation (SD) between the ACT over the three assembly cycles presented in Table 7-8, support the WEC paradigm and indicate that the participants started familiarizing with the assembly operations at a rapid pace. These findings underpin the conclusion of T-FE1 and support the hypothesis which questions the worthiness of authoring of CAD-based AR instructions.

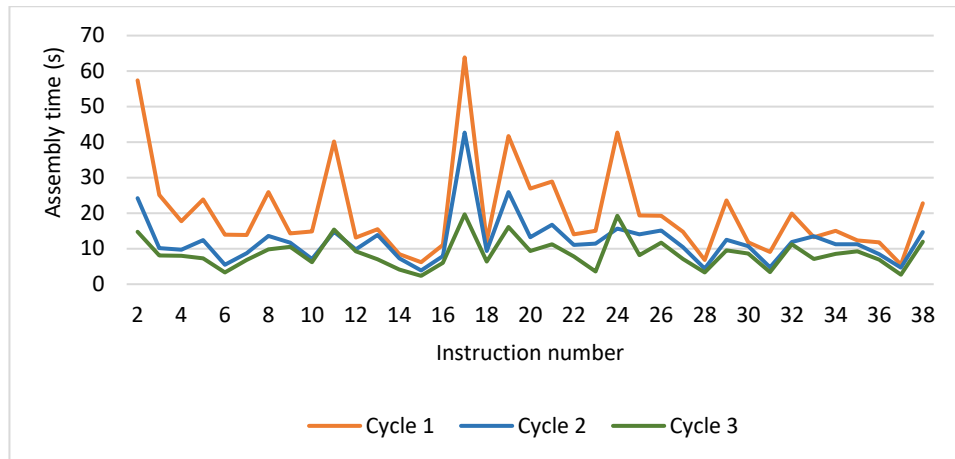
**Table 7-8. T-FE2: reported mean (M) and standard deviation (SD) of the ACT and IRT values over each of the three assembly cycles.**

Cycle number	1		2		3	
Global ACT/IRT	ACT	IRT	ACT	IRT	ACT	IRT
<b>M (seconds)</b>	22.67	7.85	13.34	3.80	9.26	2.07
<b>SD</b>	13.04	6.89	6.73	3.07	4.04	1.32

Figure 7-9 illustrates the average ACT reported by all participants, per each of the three assembly cycles. The ACT “flattening” over the three cycles supports the previous claim and demonstrates the learning progress of the participants. The overall value of the IRT decrease by 51% and by 45% in the 2<sup>nd</sup> and 3<sup>rd</sup> cycles indicates that participants became less dependent on the AR instructions rapidly (see Table 6).

By considering only the AT2 subset, we observe that CAD-based instructions led to a faster assembly progress: G-CAD required less time to perform the assembly operations of type AT2 compared to G-LA: -7%, -20% and respectively -22% over the three assembly cycles (Table 7-8).





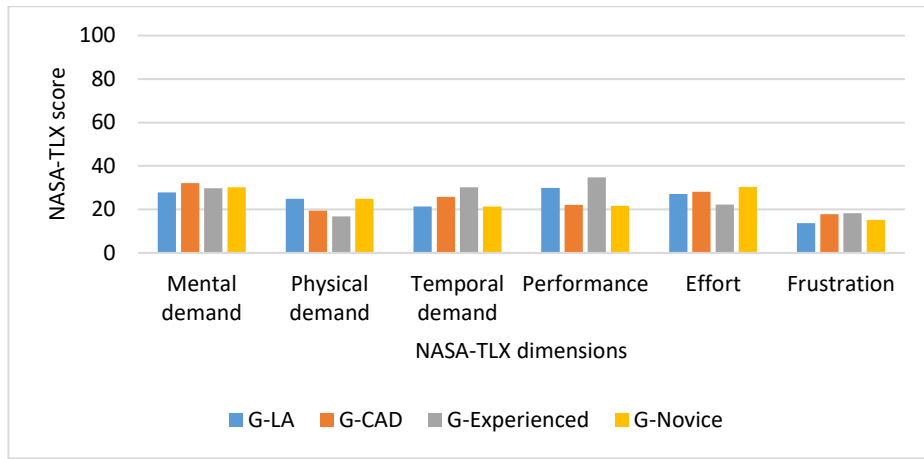
**Figure 7-9. T-FE2: reported ACT values per each assembly instruction, over the three cycles.**

In addition, we used the Pearson's correlation test to analyze the correlation of the ACT performances between the two sets of data reported by the two groups (LA and CAD), over the assembly operations of type AT2. We use the notation  $r$  for the Pearson correlation coefficient, which takes values between  $-1$  and  $1$ , measuring the strength and direction of the relationship between the two variables. Further, we conduct a statistical significance test (i.e., the *probability value* or *p-value* denoted by  $p$ ), with a threshold of  $0.05$  ( $p < 0.05$  is typically considered to be statistically significant), to measure the probability of whether the correlation between the two data sets occurred by chance. We note that a *p-value* of  $0.05$  means that there is only 5% chance that the results from the sample occurred due to chance. The test shows a high correlation between the mean ACT of the two groups, during all assembly cycles: [ $r=0.94$ ,  $p=0.142$ ] for the first cycle, [ $r=0.99$ ,  $p=0.003$ ] for the second one and finally [ $r=0.96$ ,  $p=0.007$ ] for the third assembly cycle. We observe a high correlation ( $r = 0.94$ ) without statistical significance ( $p=0.142$ ) during the first cycle. However, very strong correlations with statistical significance are reported for the second and third assembly cycles, respectively.

#### 7.1.2.2.3 Subjective measurements

We present subjective measurements organized by subgroups: G-LA, G-CAD, experienced (**G-Experienced**) and novice (**G-Novice**) participants.

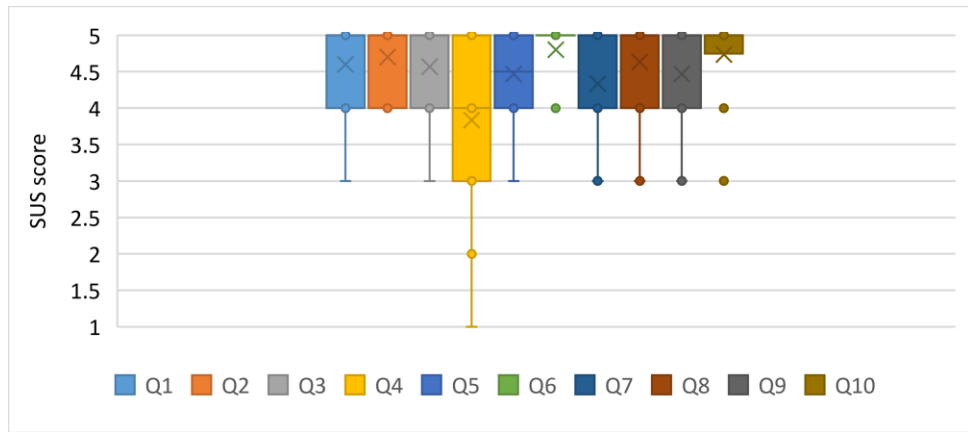
**NASA-TLX scores.** The NASA Task Load Index (NASA-TLX) is a subjective assessment tool that rates the perceived workload of the participants to assess the system's effectiveness. The NASA-TLX approach jointly considers the following 6 dimensions: mental demand (MD), physical demand (PD), temporal demand (TD), performance (P), effort (E), and frustration (F). Figure 7-10 presents the average reported NASA-TLX scores for each of the 4 subgroups, on each of the 6 dimensions: G-LA ( $S=24.53$ ,  $SD=4.94$ ), G-CAD ( $S=24.28$ ,  $SD=4.98$ ), G-Experienced ( $S=25.33$ ,  $SD=6.65$ ) and G-Novice ( $S=23.92$ ,  $SD=5.29$ ). We use the  $S$  notation to express the overall workload score and  $SD$  for the corresponding standard deviation.



**Figure 7-10. T-FE2: reported NASA-TLX scores per dimension and subgroups.**

In addition, a one-way analysis of variance, i.e., ANOVA approach, has been run to identify if a statistically significant difference between the mental workload reported in the four considered subgroups exists. This analysis finds no statistically significant differences between all subgroups ( $[F(1,29)=0.041, p=0.84]$  between G-LA and G-CAD,  $[F(1,29)=0.048, p=0.82]$  between G-Novice and G-Experienced) and on all dimensions as the reported probability value  $p$  was greater than 0.05, a threshold that is typically considered to be statistically significant. However, our post-experiment evaluation reveals that participants with assembly experience have higher expectations from a temporal perspective, affecting their perceived performance level. These claims are slightly visible in Figure 7-10, where the most notable differences can be observed between G-Experienced and G-Novice on the following dimensions: PD  $[F(1,29)=2.25, p=0.14]$ , T  $[F(1,29)=2.11, p=0.16]$  and P  $[F(1,29)=2.18, p=0.15]$ . We use the  $F(df1, df2)$  notation to express the score of the test;  $df1$  represents the number of degrees of freedom between groups - i.e., number of groups (2)–1;  $df2$  represents the total number of degrees of freedom – i.e., the total number of observations (30)–the number of groups (2).

**SUS results.** We used the SUS questionnaire to evaluate the overall usability of the proposed training method, by using a five-item Likert scale ranging from *strongly agree* (1) to *strongly disagree* (5). The reported average scores for each dimension of the questionnaire are illustrated in a *box and whiskers chart* (Figure 7-11). A one-way analysis of variance (i.e., ANOVA) reveals no significant differences between G-LA and G-CAD  $[F(1,29)=0.81, p=0.37]$  or between G-Experienced and G-Novice  $[F(1,29)=2.71, p=0.11]$ . The overall reported perceived usability for all the participants is 4.53 (that is 90.6 out of 100), with a standard deviation  $SD=0.15$ . This suggests that the proposed method validates the HCD principles discussed in Section 4.4. However, we observe that the lowest overall score is reported on the question Q4 of the SUS questionnaire (score=3.80), where some participants claimed that human supervision is required particularly during the WEC. The effectiveness measurements presented in the previous sections, and notably the error rates, support this theory. At the same time, Q4 reports a significant difference between G-LA vs. G-CAD  $[F(1,29)=6.1, p=0.02]$ , potentially indicating that CAD models lead to a higher user confidence, evidence supported by the IRT difference of assembly operations of type AT2 as well.



**Figure 7-11.** T-FE2: box and whiskers chart illustrating the raw SUS training scores reported by the participants on each dimension.

### 7.1.3 Discussion

#### 7.1.3.1 Relevant findings

The percentage of participants committing errors during the WEC, 66% in each of the two groups (L-GA and L-CAD), invalidates the hypothesis that novice workers can perform the training completely unsupervised. CAD-complemented AR instructions lead to faster ACT (-7% in the first assembly cycle, -20% in the second and -22% in the third), but to a higher (by 50%) error rate during the first assembly cycle. These results suggest that (1) FPV video demonstrations are more reliable for conveying error-prone assembly operations to novice workers and secondly, that (2) CAD-based instructions lead to faster assembly for operations of type AT2, especially after the WEC. It seems therefore that FPV video demonstrations are more effective for conveying assembly information during the WEC, however leading to higher assembly times in the following cycles, when compared to CAD models.

The reported IRT shows that G-CAD uses 28% less time for reading the instructions during the WEC, indicating that participants preferred CAD-based guidance to video demonstrations. By considering the higher error rate and the lower IRT of G-CAD during the WEC, we speculate that CAD models lead to a higher user confidence at the cost of a lower user attentiveness. The ACT decrease of only 7% in the favor of G-CAD and the error rate increase by 50% reported during the first assembly cycle support the hypothesis that video demonstrations are more effective than CAD models for conveying assembly instructions of type AT2 to novice workers. The overall decrease of the IRT by 51% and 45% in the second and third assembly cycles, respectively, suggest that participants became less dependent on the AR instructions rapidly, questioning therefore the worthiness of authoring CAD-based AR instructions.

By considering that only a single error of type non-AT2 was reported during all assembly cycles, we anticipate that low-cost visual assets can effectively convey assembly information of type AT1, AT3 and AT4 in similar assembly use cases, even during the WEC. Additionally, all participants, independently on their assembly performance reported during the experiment, believed that spatially registered low-cost visual assets were sufficient for conveying assembly instructions via AR.

Assembly experience leads to a better training performance. We observed during the experiment that participants with assembly experience performed significantly faster in AT3 tasks (e.g., screwing or riveting) and committed less errors (-52%) during the WEC. They did not commit errors during the following cycles and their ACT was better, independently on the group, with -9%, -19% and -16% on average over the three assembly cycles. AR experience however did not affect performances. Thus, we did not observe neither a lower mental workload nor usability advantages for participants with AR experience.

The ACT of G-CAD suggests that registered CAD models allow a faster identification of the position and orientation of the assembly components, as long as a precise spatial registration is guaranteed. However, the reliability of spatially registered, CAD-based AR instructions is questioned until an accurate continuous object registration technique will be provided. We note as well that CAD models seem to interfere visually with the assembly location, making it difficult to perform operations in non-obvious locations, as reported by few participants. A similar concern was observed in video-based instructions as well, where some of the participants spent more time than expected to identify the corresponding real world assembly location indicated in the demonstration video.

We note that successive and repetitive assembly operations like screwing and riveting can be grouped and conveyed as a single AR instruction. Participants performing at least three assembly cycles either suggested or agreed on this affirmation, indicating that the instruction chunking technique was not adapted for repetitive assembly operations, particularly after the WEC.

Finally, the evaluation results of the proposed training method (both instruction types) are promising. From all participants performing the third assembly cycle, none committed errors while ACT decreased on average by 59% and IRT by 73% compared to the WEC. The overall effectiveness and usability scores are favorable, indicating that the proposed method can be used in concrete real world industrial use cases.

#### 7.1.3.2 Objectives' conclusions

- **(O1)** Low-cost visual assets are sufficient to effectively describe manual assembly in AR: the evaluation results strongly support this hypothesis, particularly after the WEC.
- **(O2)** Benefits of CAD-complemented AR instructions: faster ACT especially after the WEC. Disadvantages: lower user attentiveness, eventually leading to an increased error rate especially for subtle assembly operations.
- **(O3)** Optimal visual assets for describing manual assembly in AR: spatially registered (1) visual cues (i.e., arrows) for indicating physical assembly locations, (2) text and/or FPV photo for easy assembly like picking, (3) text and FPV video demonstrations for complex assembly. CAD data can complement or even replace other visual assets if a reliable spatial registration is ensured.

#### 7.1.3.3 Final observations and lessons learnt

In a more qualitative manner, we have gathered the following observations:

- Not all participants follow instructions as requested. User attentiveness and cognitive skills definitely have an impact on the performance.

- In-app usability explanations are understood and assimilated significantly faster than the ones performed outside AR,
- Voice commands and button clicks can be (rarely) triggered without intention. The Hololens 2 menu can be triggered without intention if the device is not in “kiosk mode”.
- System-related errors occurred during the evaluations, without apparent reason, requiring restarting the device. No negative remarks regarding the comfort of the headset have been made.
- The most significant challenges encountered during the field experiment include the availability of the production area and of the personnel involved. Unexpected events like the lack of an internet connection on the shop floor also occurred.

The main limitation of our work is that the proposed training method was evaluated on a single assembly workstation. To obtain a full statistical validation of its effectiveness and usability, more complete training procedures involving novice assembly workers and multiple workstations is required. We anticipate that future evaluations, considering detailed user profiles and associated cognitive skills, will reveal important findings regarding the optimal ways of conveying profile-adapted instructions in AR. From a training perspective, we plan to conduct a large-scale evaluation in other manual assembly use cases and to extend the current evaluation to multiple workstations.

## 7.2 Authoring experiments

In this section, we present the evaluations of the proposed AR authoring system. A first one, described in the following section, concerns a comparison between ATOFIS and Microsoft Dynamics 365 Guides.

### 7.2.1 Comparative authoring evaluation: ATOFIS versus Guides (field study - AT-FS)

#### 7.2.1.1 Introduction and objective

The main criterion that we have considered in this study is the authoring time required for completing the authoring procedure. To this purpose, we have analyzed the creation of step-by-step, low-cost AR work instructions corresponding to the assembly tasks of five assembly workstations. We note that the current study did not deal with CAD data.

#### 7.2.1.2 Set-up and evaluation procedure

As in the case of the training experiment presented in Section 7.1, the authoring study was conducted at elm.leblanc, on the assembly line L2. The first five workstations of L2 have been chosen for authoring the corresponding step-by-step work instructions in AR. The numbers of assembly operations created for each workstation are 33, 11, 33, 20 and 35, respectively. The author of the AR instructions, myself, is expert in both performing the assembly tasks of the selected workstations and in manipulating the two considered AR authoring systems. We must note that both authoring tools have been extensively used by the author during the duration of the thesis, reason for which we can consider that no bias from the usage perspective can be implied. The author has performed the assembly of the considered workstations multiple times before the experiment, for the same reason to avoid an evaluation bias during the usage of the

tools. Finally, to ensure an objective evaluation of the two authoring tools, the creation of the AR work instructions corresponding to the 5 workstations was conducted as follows: (i) for each workstation  $W_N$ , the author performed the experiment with both tools, before going to the next workstation  $W_{N+1}$ ; (ii) for each workstation, the usage of the two authoring systems was inverted. Therefore, the experiment started with  $W_1$ , by using ATOFIS then Guides, then continued with  $W_2$ , by using Guides then ATOFIS and so on. We consider therefore that the study should provide an unbiased and reliable estimation regarding the authoring performances when performed by experienced shop floor workers (e.g., line manager) that are familiar with the authoring systems as well. Moreover, the study aims to provide preliminary results regarding authoring expectations from an industrial perspective, particularly regarding time constraints, adaptability, and robustness, before conducting a costly, large-scale shop floor field experiment.

The detailed authoring workflow of Guides is available at (*Authoring Overview - Dynamics 365 Guides PC and HoloLens | Microsoft Docs, 2021*). Let us just recall the elements considered in our study. Authoring with Guides includes three successive stages: media preparation, PC authoring and HMD authoring. These stages represent separate, independent procedures that must be performed in the considered order, not at the same time and therefore potentially realized by different individuals. The first authoring step consists in preparing the multimedia files by either collecting existing files or creating new ones (e.g., photo and video capture, 3D model creation). The second authoring stage consists of creating a new *guide*, adding instructions (steps), their associated text descriptions and linking the multimedia files prepared in the first stage to their corresponding instructions. It is important to note that this stage requires the utilization of a desktop application running on a PC/laptop under a Windows OS. The final stage concerns the spatial registration. It takes place *in-situ*, is performed by using the AR device (HoloLens 2) and consists in placing the virtual content (arrows, product models, animations) in the real-world environment.

The considered evaluation takes into consideration only the actual authoring of the AR work instructions, excluding the actions that must be performed additionally, including user login, workstation selection, fiducial marker placement and scanning, etc. To avoid bias during the evaluation of the two authoring systems, the assembly instructions for each workstation were printed similarly as paper instructions, together with the media type that needs to be captured and the number of location arrows. In this way, we ensure the creation of the exact same AR instructions with both authoring systems: text description, number of location arrows and photo/video capture. Every assembly instruction was therefore written down by respecting the following template: “*DESCRIPTION, M, N*”, where “*M*” is either a “*P*” (photo) or a “*V*” (video), represented the captured media type while *N* is an integer, representing the number of the location arrows of the corresponding assembly task. An example of such an assembly instruction is “*Screw the 4 screws at the indicated locations, P, 4*”. The choice between illustrating a specific assembly operation by using a photo or a video demonstration was entirely left to the author of the AR instructions, based on his assembly experience at the considered workstations.

### 7.2.1.3 Results and interpretation

Table 7-9 presents the number of AR instructions and their corresponding media assets (photo or video for *how*, location arrows for *where*) captured during the authoring procedure, together with the total number of characters (text description for *what*) used to describe all assembly operations, per workstation. Together with the information reported in Table 7-10, we can estimate the media composition, on average, of an AR work instruction. Furthermore, by considering the information reported in Table 7-12, we can as well estimate the average time required for authoring the AR work instructions of a new assembly workstation.

**Table 7-9. AT-FS: authoring data composition per each of the five considered workstations.**

Workstation	W1	W2	W3	W4	W5	Total
<b>No. of instructions</b>	33	11	33	20	35	132
<b>No. of videos</b>	12	6	13	9	9	49
<b>No. of photos</b>	17	5	16	11	19	68
<b>No. of arrows</b>	43	22	41	24	45	175
<b>No. of characters</b>	1214	358	794	347	1106	3819

Table 7-10 presents the average number of characters, location arrows, photos and demonstration video used to author an AR work instruction, for each of the five considered workstations.

**Table 7-10. AT-FS: average number of media assets used for authoring an AR work instruction, per each of the five considered workstations.**

Workstation	W1	W2	W3	W4	W5	Average
<b>Avg. per instruction</b>						
<b>No. of characters</b>	37	33	24	18	32	28.8
<b>No. of location arrows</b>	1.3	2	1.2	1.3	1.3	1.4
<b>No. of photos</b>	0.51	0.45	0.48	0.55	0.54	0.5
<b>No. of videos</b>	0.36	0.54	0.39	0.45	0.26	0.4

We used the reported preliminary data to estimate the composition, on average, of an AR instruction and the overall authoring time required for a new assembly workstation. The average number of videos per workstation indicates that the author of the AR instructions considers that 40% of the total assembly operations are rather difficult or error-prone, and therefore require a video demonstration. The average number of captured photos indicates that 50% of the total assembly tasks are relatively easy and that 10% of them are obvious and do not require a visual representation in addition to the text description and the location indication (i.e., arrow). Finally, the number of characters used to describe an assembly operation was 28.8, on average, as indicated in Table 7-10.

Table 7-11 presents the authoring times in seconds for both AR authoring systems under evaluation. In addition, the time differences between the two systems are presented.

**Table 7-11. AT-FS: a comparison between authoring times reported by ATOFIS and Guides, respectively, for each of the workstations.**

Workstation	W1	W2	W3	W4	W5	Total
Guides (seconds)	2822	991	2459	1804	2706	10782
ATOFIS (seconds)	1805	606	1663	1058	1837	6969
Difference (%)	-36%	-39%	-32%	-41%	-32%	-35%

We observe an authoring time improvement of 35% on average, in the favor of ATOFIS.

Furthermore, a remarkable finding observed from the time measurements can be observed in Table 7-12, which reports the average authoring times required to create a single AR work instruction, per workstation, by using ATOFIS and Guides, respectively. It is interesting to note the extremely low standard deviation ( $SD=1.21$ ) between the average authoring times of an AR work instruction, for all the considered workstations, reported by ATOFIS.

**Table 7-12. AT-FS: average AR work instruction authoring time reported by the two systems, per workstation.**

Workstation	W1	W2	W3	W4	W5	Total avg.	SD
ATOFIS (seconds)	54	55	51	53	53	53.2	1.21
Guides (seconds)	85	90	74	90	77	83.2	6.03

By considering that the evaluated workstations are representative, they were uniformly selected, and their cycle times is very similar (guaranteed by the assembly line balancing), we can estimate that the authoring of the corresponding AR work instructions of a new workstation,  $W_x$ , composed of  $N$  assembly operations, would approximately be  $N*53.2$  seconds. We note that a precise estimation of the time required to author a new workstation is an important planning aspect in the context of manufacturing industries.

#### 7.2.1.4 Discussion

The preliminary data collected during the authoring field study indicates that an expert in manipulating the proposed authoring system, would require approximately 30 minutes for creating the AR work instructions corresponding to an assembly workstation with a nominal cycle time of approximately 3 minutes and 30 seconds (i.e., generally 30 to 35 assembly operations), in a similar assembly context. The media types captured during the authoring of the five selected workstations for the evaluation, indicate that 40% of the assembly operations require video demonstrations (i.e., *how* the assembly is executed), 50% require a photo of the final assembly (i.e., *how* the assembly result looks like or a key assembly illustration), while 10% require solely a text description and an indication to the physical location of the assembly (i.e., *what* and *where*).

The proposed authoring method performs, on average, 35% faster than Guides when used by an expert in both the assembly procedure and the AR authoring system. The preliminary collected data suggests that a precise authoring time estimation of a manual assembly workstation, similar to our use case, can be obtained by using our proposed authoring approach. In addition to the authoring time gain, we expect our AR authoring proposal to provide other



benefits including faster learning curve, lower mental and physical effort, less expertise required and most importantly better adaptation to shop floor usage. These various hypotheses are discussed and evaluated in the following section.

### 7.2.2 Full comparative evaluation between ATOFIS and Guides (laboratory experiment - AT-LE)

In this section, we consider a complete evaluation of the proposed AR training system, jointly including both authoring and training phases. In most research works reported in the literature, the two main functionalities of AR training systems, authoring and learning, are most of the time separately evaluated. Even though the two functions can be designed, implemented, and evaluated separately, because of the holistic nature of an AR training system, an incremental evaluation of the two should be considered. This is a strong requirement for the industry, where the most fragile of the two will negatively impact the overall perceived utility, usability, and effectiveness of the system. We stress that evaluating an authoring system without assessing the obtained results (i.e., the AR work instructions) represents a critical concern. In addition to addressing this evaluation inadequacy, we aim as well to provide a comprehensive comparative evaluation between ATOFIS and Guides, including both authoring and training procedures.

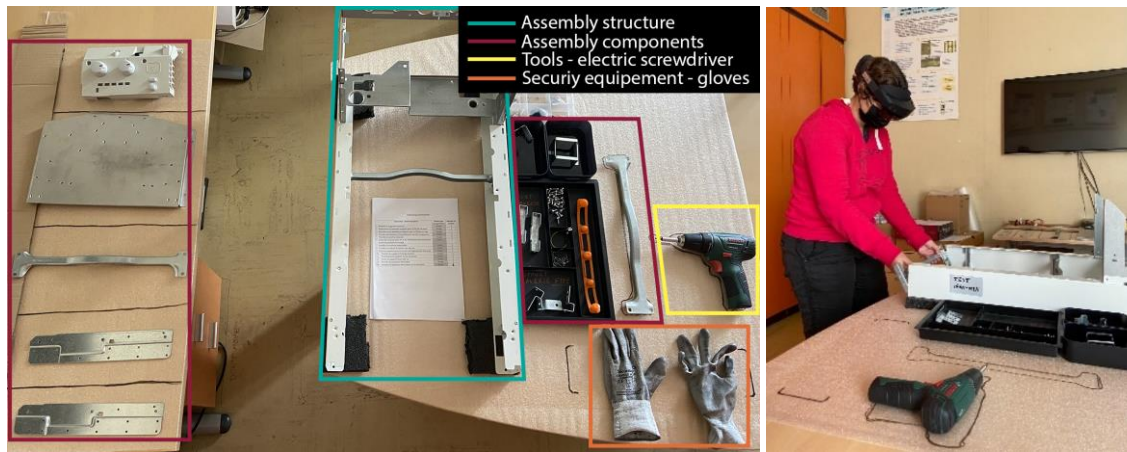
The main objectives of the experiment are the following:

- **O1:** Assess the overall usability and utility of the proposed AR training system with respect to the industrial requirements and challenges. Provide subjective and objective evaluation measurements relevant to industry (e.g., effectiveness, error rate, time estimations).
- **O2:** Assess the usefulness and effectiveness of the authored AR work instructions and the information conveyance approach.
- **O3:** Comparatively evaluate ATOFIS and Guides from both authoring and training perspectives.

#### 7.2.2.1 Set-up and preparation

Unlike all other evaluations conducted during this research work, this experiment has been performed in a laboratory setting. This was an organizational limitation, mainly due to the shop floor availability (i.e., assembly lines and workers, especially novices).

To simulate the shop floor conditions as close as possible, we have setup a partial assembly workstation by using a set of the components of W201, the first workstation of the assembly line L2 at elm.leblanc. The experiment has been conducted in a dedicated office at Telecom SudParis. Figure 7-12 illustrates the workstation with its main components and a participant performing an assembly operation during the experiment.



**Figure 7-12.** AT-LE: the assembly workstation setup with its main components and a participant performing an assembly operation.

We have involved 16 participants from Telecom SudParis to take part to the considered study: 4 females and 12 males, aged between 23 and 32, with an average of 27. One is a doctor, five are engineers, five are PhD students, and five are master's degree students. Table 7-13 illustrates the participants' AR and assembly experience. It is important to note that none of the participants have shop floor assembly experience, and none work with or studied AR in particular. Subjective post-experiment feedback suggests that participants rating AR experience "good" or "advanced" have basic knowledge about AR and are generally not familiar with advanced or technical AR topics. Most of them (13 out of 16) have used an AR device before (but not necessarily HoloLens 2).

**Table 7-13.** AT-LE: AR and assembly experience of participants to the experiment.

Experience	None	Elementary	Good	Advanced	Expert
Assembly	3	9	1	2	1
AR	3	6	2	5	0

#### 7.2.2.2 Procedure

To evaluate the proposed AR training system, we have conducted the experiment in two stages (i.e., authoring and training), as further described.

**Training procedure:** from *novice worker* to *shop floor expert*. The first stage of the experiment consisted in training the participants by using the two considered AR training systems. Each participant learned the assembly sequence by performing two assembly cycles, guided by ATOFIS and Guides, respectively. Once the training procedure was finished, the participant was considered a shop floor expert, meaning that he could assumingly perform the assembly operations unassisted.

**Authoring procedure:** authoring the corresponding AR work instructions for the considered assembly workstation. Once the participants learned how to execute the assembly operations, they were requested to replicate as close as possible the AR work instructions used during the first stage of the experiment (i.e., training), by using the two AR authoring systems.

It is worth mentioning the following observations:

- All participants were considered novice workers at first, as they did not have assembly experience neither any knowledge about the assembly operations of the experiment. The participants were not familiar with any of the AR training systems.
- As all participants were novice workers and followed the same learning process, their assembly performance from a time perspective did not bias the experiment.
- To ensure a fair evaluation, for both functions (i.e., training and authoring), we have employed the following strategy. For each participant, the training started with the last AR system used by the previous participant. The same was applied for the authoring procedure as well. Therefore, the first participant performed the first assembly cycle by using ATOFIS and the second one by using Guides. He then performed the authoring by using Guides, then ATOFIS. The 2<sup>nd</sup> participant started the training with Guides, then ATOFIS. For convenience, we describe this procedure by using the following notation: 1 ([A, G] [G, A]) -> 2 ([G, A,] [A, G]) ->...-> N ([A, G] [G, A]) -> etc., where N represents the number of the participant, the first pair of letters ([A, G]) refers to training while the second pair ([G, A]) refers to authoring. A stands for ATOFIS and G stands for Guides.
- To evaluate the usability and correctness of the authored content, every participant learned the assembly by using the AR work instructions created by the previous participant. The first participant used the AR work instructions created by us.

#### 7.2.2.2.1 Experimental procedure workflow

For each participant, the experiment has been conducted as indicated in Table 7-14.

**Table 7-14. AT-LE: experiment procedure workflow and description.**

Step	Objective	Description
1	Scope	Brief presentation of the objective of the evaluation, the research context, including industrial concerns and shop floor requirements.
2	AR device	Presentation of the AR device (i.e., Hololens 2), including visualization, basic interaction, and eye calibration.
<b>Training session</b>		
3	Training demo	Presentation of the two training systems, individually. Ask the participant to perform a short training session to get familiar with the functionalities, user interfaces and interactions. The examiner explains and interferes during the demonstration, which lasts until the participant feels comfortable using both training applications.
3.1	GUIDES	Participant N starts with Guides.
3.2	ATOFIS	Participant N+1 starts with ATOFIS.
4	Training evaluation	The participant performs 2 assembly cycles, each cycle guided by one of the two AR training systems. The examiner highlights the importance of performing the assembly operations correctly, a priority over the assembly completion time. For the 2 <sup>nd</sup> assembly cycle, the participant is asked to still follow the AR work instructions, even though some of the assembly tasks might be remembered from the 1 <sup>st</sup> cycle.
4.1	GUIDES	Participant N starts with ATOFIS
4.2	ATOFIS	Participant N+1 starts with Guides
5	NASA-TLX & SUS for training	At the end of the two assembly cycles, the participant is asked to fill in two questionnaires, NASA-TLX and SUS. The examiner explains the role of the two and explains the notation manner for each. The participant is asked to fill in both questionnaires at the same time, encouraging therefore a parallel evaluation of the

		two systems, on each dimension. Finally, the participant is asked to answer the set of questions related to the training procedure, presented in a specific questionnaire (step 11).
<b>Authoring session</b>		
<b>6</b>	<b>Authoring demo ATOFIS</b>	Present ATOFIS authoring. The participant performs a guided demo session to get familiar with the authoring functionalities, UI, and interactions. The examiner gives explanations and interferes during the demonstration whenever necessary. The session lasts until the participant feels comfortable using the authoring system.
<b>7</b>	<b>Authoring evaluation ATOFIS</b>	At this point, the participant is considered an expert in the assembly sequence used for the evaluation. However, to ensure the authoring of the exact same AR instructions among the participants and address potential failure to recall, a printed document containing the details of the assembly sequence (first 4 columns of Table 7-15) is provided to the participant. During authoring, the participant must use the text description indicated for each instruction, make sure to capture the indicated media file (i.e., photo or video), and finally use the indicated number of location arrows. We ensure therefore the authoring of the same AR instructions among the participants, to objectively evaluate the two systems and avoid bias.
<b>8</b>	<b>Authoring Demo GUIDES</b>	Presentation of Guides authoring system. The participant performs a guided demo session including the three authoring stages: (1) photo and video capture: (2) PC authoring and finally (3) HDM authoring. Similarity as for step 6, the session ends once the participant feels comfortable with the authoring procedure.
<b>9</b>	<b>Authoring Evaluation GUIDES</b>	Perform the authoring of the AR work instructions by using Guides. We note that step 9.2 is a preparatory stage, required for the PC authoring (step 9.3), considered, and evaluated as an individual step.
9.1	Photo & Video Capture	The participant starts the authoring by capturing the corresponding photos and video demonstrations of the assembly. To address the hands-free requirement, the participant uses the Hololens 2 for capturing the required media files.
9.2	Download, verify, rename & upload in app	A prerequisite for the next step. The participant goes, together with the examiner, to the office where the desktop PC dedicated to the experiment is installed. The participant downloads the captured media files, verifies their correctness, filters them by removing duplicates if multiple captures were taken for the same assembly, renames them and finally uploads them in the application. The renaming process was recommended by the examiner, for speeding up the media identification and linking during the PC authoring.
9.3	PC authoring	The participant chooses the registration method (i.e., QR code) and starts adding the steps and their associated text, guided by the paper instruction. For each step, the participant associates the corresponding media file captured and uploaded priorly, and adds a predefined arrow (i.e., the straight arrow) from the 3D toolkit library provided in Guides.
9.4	HMD authoring	Once the PC authoring is finished, the participant goes back to the office where the workstation is, to complete the last authoring stage, performed in Hololens 2. During this stage, the participant indicates the physical locations of each corresponding instruction by placing the virtual “gem” of the instruction panel (which links the panel with an animated dotted line) to the corresponding physical location of the assembly. The participant uses the predefined arrow in the same manner, to indicate additional assembly locations.
<b>10</b>	<b>NASA-TLX &amp; SUS</b>	The participant fills in the NASA-TLX and SUS questionnaires for the authoring system, by following the same procedure explained at step 5.

11	<b>Specific questionnaire</b>	Finally, the participant fills in a specific questionnaire which aims to determine the user profile, including AR and assembly experience, and collect feedback regarding each of the 4 procedures (i.e., 2 x training + 2 x authoring). The participant then indicates his preferred method of training and authoring between the 2 systems.
----	-------------------------------	---

We note that the text insertion for the ATOFIS authoring is performed by using the dictation mode. The virtual keyboard is displayed in the user interface for triggering or stopping the dictation mode, which at the time of the implementation represented a technical limitation. To avoid UI clutter, future implementations could either start the dictation automatically or provide a visual element (e.g., button) dedicated to starting and stopping the dictation. The examiner asks the participants to read the text description found on the printed document once and validate the input as is, even if dictation errors occur. In this way, we assessed the quality of the dictation service provided natively in Hololens 2.

#### 7.2.2.2.2 Assembly operations set

In addition to replicating the target assembly workstation, the definition of a set of assembly instructions used to evaluate the two AR systems is required. The participants learn performing the considered set of assembly operations during training, then create the corresponding AR work instructions as part of the authoring procedure, with both systems. The initial assembly set consisted of 30 operations. A full evaluation procedure, conducted with one participant, required approximately 6 hours, which was considered unreasonable. We decided to decrease the number of assembly operations to reduce the evaluation time to 4 hours or less, a timespan which was still considered high but feasible. We have reduced therefore the assembly set to 14, but diversified the assembly operations as much as possible, to imitate the real workstation in the best way possible and ensure the same level of complexity. Table 7-15 describes the assembly set by the number of tasks and their complexity, media type used (P – photo, V – video) to illustrate the assembly and number of location arrows. We note that the complexity level is a number between 1 and 5, subjectively assigned by the examiner. The complexity value is relative to the considered assembly set and reflects an average between the error susceptibility of the corresponding assembly operations and the physical difficulty to execute it. The following notations are used in Table 7-15, which summarizes the considered instruction set: I.N. - instruction number, M.T. - media type, A.N. - number of arrows and C.L. - complexity level.

**Table 7-15. AT-LE: instruction set used for the experiment, including the information associated to each corresponding AR work instruction.**

I.N.	Instruction description	M.T.	A.N.	C.L.	Assembly type. Notes.
1	Take 2 crosspiece supports	P	1	2	<b>Picking.</b> Complexity 2 because the two pieces should be picked
2	Position the first support with the arrow up	P	1	4	<b>Positioning.</b> This task is error prone because of the symmetrical nature of the assembly component.
3	Position the second support, arrow down	P	1	4	Same as above.
4	Take the crosspiece and position it on the 2 supports	V	2	4	<b>Positioning.</b> Error-prone because the crosspiece can be placed upside down.

5	Take 4 screws and the screwdriver	P	2	2	<b>Picking.</b> Two different components, slightly more difficult than assembly 1.
6	Screw the crosspiece with the 4 screws. Hold the crosspiece firmly	V	2	5	<b>Screwing.</b> High complexity mainly because of physical and time considerations. Correctly holding the crosspiece during the screwing is important.
7	Install the assembled crosspiece	V	1	4	<b>Installing.</b> Error prone because the crosspiece can be installed in 4 different positions.
8	Take a clip and put it on a screw	V	2	4	<b>Picking and assembling.</b> “Air assembly”, requiring the worker to pick and assembly two small components with both hands.
9	Fit the screw, tightening it lightly by hand	V	1	3	<b>Screwing/tightening.</b> The screwing location is error prone.
10	Take the turbine exchange support	P	1	1	<b>Picking.</b> Basic picking, not error prone nor difficult to perform.
11	Install the support on the structure	V	1	3	<b>Installing.</b> Specific gesture required for correctly positioning the piece.
12	Screw the exchange support with a screw	V	1	3	<b>Picking and screwing.</b> The instruction does not specify that a screw and the screwdriver shall be picked before, allowing the worker to reason, as a similar task was performed before.
13	Grab the electrical panel	P	1	1	<b>Picking.</b> Basic picking, not error prone nor difficult to perform.
14	Install the electrical panel on the structure	V	1	4	<b>Installing.</b> Specific manipulation. The structure allows the installation of the panel in 2 positions.

Even if the number of operations is not high – i.e., approximately a third of the total workstation assembly, we aim to reach a diversity level which is comparable with the complete set of instructions of the considered workstation. We have therefore included numerous assembly types including one-hand and two-hands picking, one and multiple components picking, basic and complex positioning/placing, installing, screwing, “air assembly” manipulation and tool usage. We note that only two assembly types could not be reproduced in our experiment, compared to the considered assembly workstation (i.e., W201) in the factory, including riveting and manipulating the mobile structure (Section 7.1.1).

We note a key aspect concerning the initial proposed authoring approach and the instruction set presented in Table 7-15. As discussed in Section 5.2.3.2, an implementation of the proposed authoring approach can easily extend the basic functionalities initially proposed, by changing the way the AR instructions are created. One such example is exemplified in the instruction set corresponding to the experiment discussed in this section, where the 2W1H principle is slightly adapted as follows. *What*, in this case can indicate the manipulation of one or multiple physical components. As an example, the first instruction of the assembly set, describes the picking of 2 components from the same physical location. This represents the basic, initial usage of *what*. However, instruction 5 asks picking different components from two different physical locations. This has repercussions on *where*, where the author should indicate two physical locations instead of one, as initially described by the 2W1H principle. As described in Section 5.2.3.2, our assumption is that extending the 2W1H as such for identical or very similar consecutive assembly operations that take place at different locations (still in the FoV of the user), allow combining

these operations into one AR instruction. This strategy not only allows a faster authoring, but seems to be preferred, as remarked by the participants during the shop floor training experiment described in Section 7.1.2. In addition to similar consecutive tasks, we assume that this strategy can be successfully applied for describing recent assembly tasks performed by the worker, explicitly described at their first appearance by following the initial 2W1H authoring technique. As an example, instruction 5 and 6 explicitly describe the picking of screws and the screwdriver and performing the screwing. Instruction 12 however, asks the worker to screw the exchange support, without explicitly indicating that a screw and the screwdriver must be first picked, which would have required the creation of an additional AR instruction, an operation that was already performed by the worker. The training evaluation feedback suggested that even novice workers prefer such guidance approach, considering that obvious, self-explanatory, or redundant instructions may cause frustration, in addition to longer assembly time. The validity of the 2W1H extension hypothesis is confirmed by the evaluation results discussed in Section 7.2.2.3.

#### 7.2.2.2.3 Evaluation metrics and data collection

We have evaluated the considered AR systems both objectively, by measuring completion times and error rates, and subjectively by measuring the perceived mental workload of the participants (NASA-TLX) and the perceived usability of the application (SUS).

**Completion times.** We have measured each stage of the experiment, including the familiarization with the AR headset and the demonstration sessions (Table 7-14). We have used this data not only to comparatively evaluate the two systems, but also to obtain time estimates for the individual stages of the entire evaluation procedure, and in particular the time needed for participants to get comfortable using the systems. We note that the time was manually measured by the examiner.

In addition, a log system was implemented in ATOFIS. The participants actions have been tracked and stored locally on the device, for both the authoring and training procedures. The log files have been used by the examiner to confirm the accuracy of the manual time measurements. A similar log system is not provided by Guides. However, the comparison with the time measurements reported in the log files confirmed the accuracy of the ones measured manually by the examiner.

**Error rates.** The error rate has also been measured by the examiner as follows. During the training procedure, the examiner stops the participant after advancing to the next instruction, if the previous assembly operation was executed incorrectly. In this case, the examiner briefly explains the assembly error while performing the task correctly. The participant continues afterwards the assembly procedure from where he was stopped. It is important to note that during the authoring phase, the examiner interferes to prevent errors in the assembly sequence or in performing the operations. We considered that committing such errors would have been related to the assembly experience of the participants or lack of attention, not by the authoring procedure itself. This hypothesis was confirmed at the end of the evaluation as only

3 authoring errors have been committed by different participants, one skipping an assembly operation and two performing the assembly incorrectly.

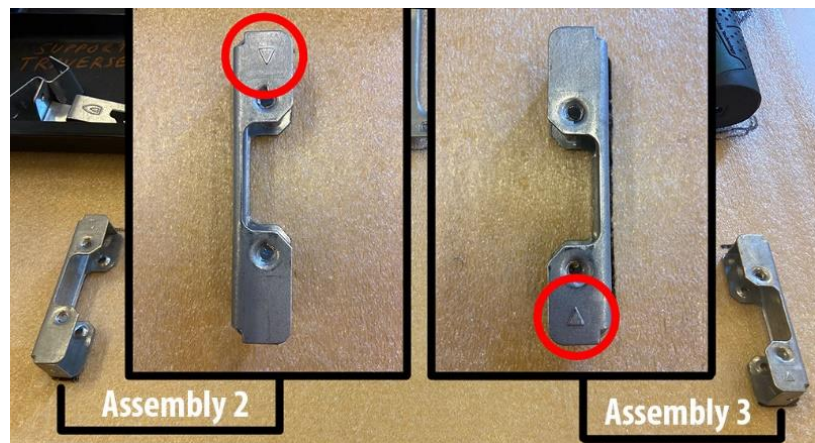
**Subjective evaluation.** Finally, we have used the NASA-TLX questionnaire for measuring the perceived mental workload of the participants, the SUS questionnaire for measuring the usability score and a specific questionnaire designed by us to collect user feedback (questions and free comments) as well as authoring and training preferences (ATOFIS vs. Guides).

### 7.2.2.3 Results and interpretation

The overall evaluation required 16 days (one day per participant), totaling approximately 53 hours, an evaluation that would not have been possible to conduct on the shop floor at elm.leblanc. All participants managed to use both systems successfully, without any significant concern to report. Therefore, none of them was considered outlier. In the following sections, the evaluation results for both training and authoring systems are discussed.

#### 7.2.2.3.1 Training procedure

**Error rates.** Three participants committed four assembly errors during their first assembly cycle with Guides, while ATOFIS reported no errors. Three errors out of the four were committed on the assembly operation number 3, specifically designed to test user attentiveness and information visualization effectiveness. We note that this assembly operation (Figure 7-13, step 3) corresponds to the 6<sup>th</sup> operation used in the shop floor training experiment, considered error-prone because of the symmetrical nature of the assembly component, that can be positioned in two different ways (upside-down).



**Figure 7-13.** AT-LE: assembly operations 2 and 3, considered error-prone because of the symmetrical nature of the assembly component.

To eliminate randomness, the second instruction asks the user to place the piece arrow-up, while the third, which expectedly would ask the user to place the second piece symmetrically to the first, asks the exact opposite. We assume, supported by the user feedback of T-FE1, that providing detailed text description for highlighting subtle assembly details, complemented by close-up photos of the final assembly, would significantly reduce the number of errors even during the workstation exploration cycle (WEC).



This hypothesis is partially validated by Guides, where 18.75% (3 out of 16) of the participants committed at least one error during the first assembly cycle. Guides was not evaluated in T-FE1; therefore, we based our claim by considering the error rate reported by ATOFIS during the WEC in T-FE1, which may be inaccurate. ATOFIS however reported no errors during the WEC, comparatively to 66% reported in T-FE1, validating the hypothesis that a detailed text description is sufficient for highlighting subtle assembly details. We believe that a reasonable explanation of the reported result can be given by the spatial contiguity principle discussed in Section 4.2.5. The information visualization method in Guides does not address this principle, as the visual information is displayed far from the physical location of the assembly, making it hard for the worker to make the correspondence between the two. Moreover, this information visualization approach can make it difficult for one to notice subtle assembly details captured in the photo or video demonstration. By considering that no errors have been reported by ATOFIS, we argue that contextualization and better media visualization enhance understanding and thus ensure lower error rates, particularly during the WEC.

No errors have been reported during the second assembly cycle.

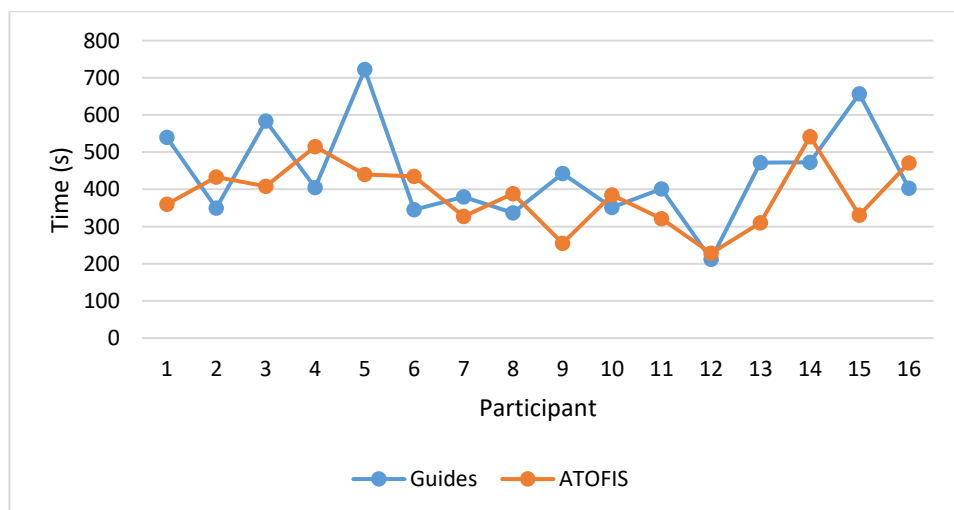
**Assembly time.** The assembly times measured during both cycles are reported in Table 7-16. They show that every participant had a better time performance in the second assembly cycle, independently on the AR training system, an information that is illustrated in Figure 7-14 as well. This is not surprising as all participants start the second cycle already familiarized with the assembly sequence and operations, leading therefore to a better assembly performance independently on the training system.

**Table 7-16. AT-LE: reported training times for both cycles (systems).**

Participant	Training order	Guides	ATOFIS	Difference
1	(G,A)	540	360	33.33%
2	(A,G)	350	433	-19.17%
3	(G,A)	584	408	30.14%
4	(A,G)	405	515	-21.36%
5	(G,A)	722	440	39.06%
6	(A,G)	345	435	-20.69%
7	(G,A)	380	327	13.95%
8	(A,G)	337	388	-13.14%
9	(G,A)	443	255	42.44%
10	(A,G)	351	385	-8.83%
11	(G,A)	401	321	19.95%
12	(A,G)	212	228	-7.02%
13	(G,A)	472	310	34.32%
14	(A,G)	473	542	-12.73%
15	(G,A)	657	331	49.62%
16	(A,G)	403	471	-14.44%
<b>Average</b>	Both cycles	442	384	13.09%
	1 <sup>st</sup> cycle	Guides 1 <sup>st</sup>	ATOFIS 1 <sup>st</sup>	
		524	424	<b>19.10%</b>
	2 <sup>nd</sup> cycle	Guides 2 <sup>nd</sup>	ATOFIS 2 <sup>nd</sup>	
359		344	4.31%	

However, ATOFIS reports an assembly time improvement by 13% overall (including all cycles), and approximately by 19% when taking into consideration the reported measurements when both systems were used during the first assembly cycle only. As expected, the time gap lowers significantly, to approximately 4% when considering the usage of the systems during the second assembly cycle, as the participants have already performed each assembly operation once. We have run a paired two-sample t-test to evaluate whether the means of the time values reported by the two training systems significantly differ from each other. We use the notation  $t(df)$  to express the t-test result, where  $df$  represents the number of degrees of freedom, and  $p$  to express the probability value (as detailed in Section 7.1.2.2.2). The t-test result ( $t(7)=2.48$ ,  $p=0.02$ ) shows a significant difference between the training times of the first assembly cycle, while no significant difference is reported for the second training cycle. These findings suggest that our proposed information conveyance approach leads to a significant assembly time improvement during the WEC only.

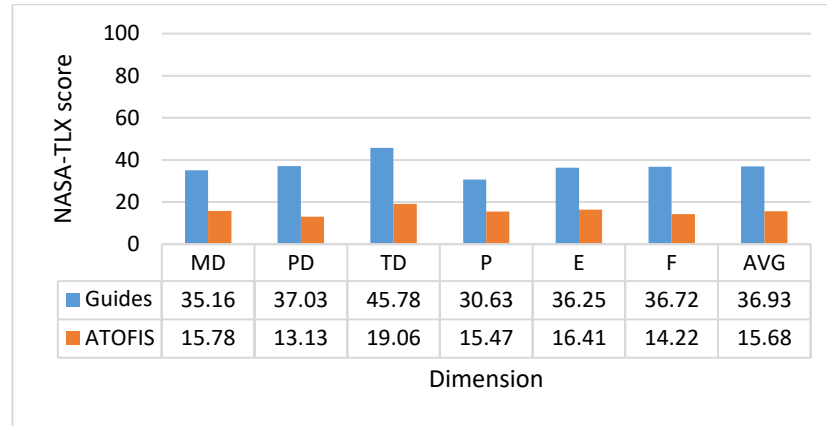
Figure 7-14 illustrates the assembly times for each of the two cycles, for each participant. For every other participant the training applications were switched, explaining the training timelines crossing every next evaluation. It is worth observing that ATOFIS leads to a significant faster overall assembly time (i.e., ~33% on average) when used during the second cycle, compared to Guides (i.e., ~15% on average). This indicates that participants spent less time in UI interaction with ATOFIS, validating the recommendation of providing a UI that makes the worker focus on the assembly work, while spending as little as possible on application usage, optimizing therefore the information conveyance and implicitly the training time.



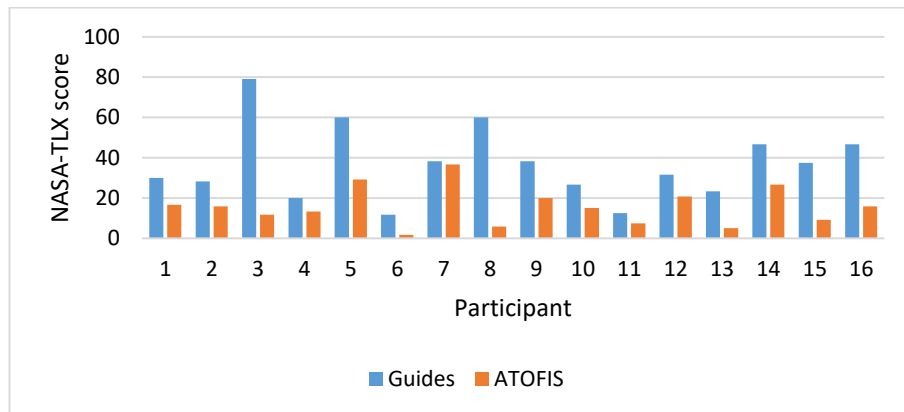
**Figure 7-14. AT-LE: reported training times per participant, per cycle (training system).**

**NASA-TLX scores.** Figure 7-15 illustrates the average NASA-TLX scores reported for Guides and ATOFIS, for each of the 6 dimensions. We note that the NASA-TLX scores are rated on a scale from 0 (low workload) to 100 (high workload). A low score signifies a low perceived mental workload. Therefore, the lower the score, the more positive the rating of the considered system. The result of a paired two-sample t-test ( $t(15)=4.75$ ,  $p=0.0001$ ) indicates that the mean values of the perceived mental workload of the participants was significantly higher for Guides ( $S=36.9$ ,

SD=17.7) than for ATOFIS ( $S=15.8$ ,  $SD=9.1$ ), where  $S$  represents the overall score of the NASA-TLX, while  $SD$  represents the standard deviation between the scores reported by each participant. ATOFIS reported better results on each of the 6 dimensions. These scores indicate that training with Guides was perceived by the users as being *somewhat high*, while the training with ATOFIS was considered *medium* from a mental workload perspective. The average score reported by each participant was as well in the favor of ATOFIS, as illustrated in Figure 7-16.

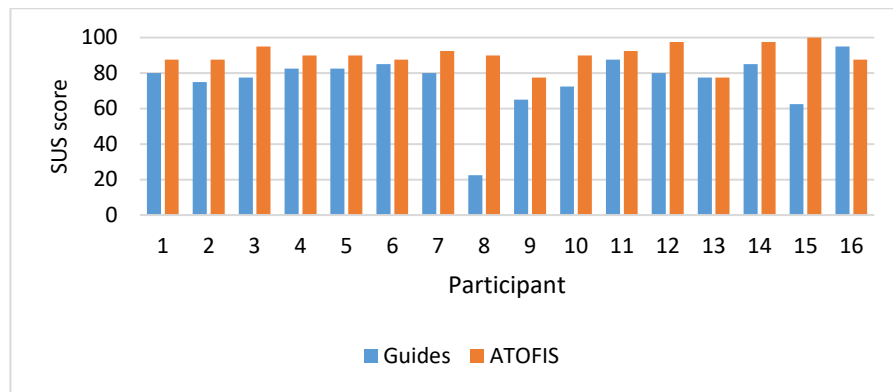


**Figure 7-15.** AT-LE: average NASA-TLX scores reported by participants on each dimension, for both training systems.



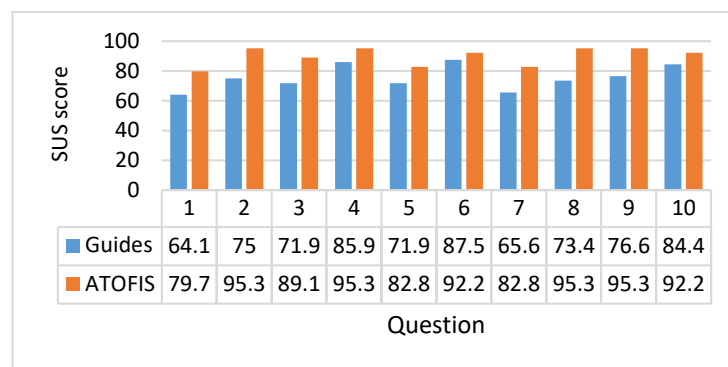
**Figure 7-16.** AT-LE: NASA-TLX scores reported by each participant, for both training systems.

**SUS results.** Figure 7-17 illustrates the average normalized SUS scores reported by each participant for the two training systems. We observe that all participants except one (i.e., P16), have rated ATOFIS better than Guides. The result of a paired two-sample t-test ( $t(15)=3.33$ ,  $p=0.002$ ) indicates that the average SUS scores reported by the participants were significantly lower for Guides ( $S=75.63$ ,  $SD=15.75$ ) than for ATOFIS ( $S=90$ ,  $SD=6.06$ ). The SUS adjective rating scale proposed in (Bangor et al., 2009) indicates that the rating for Guides is *good* ( $>71.4$ ), while for ATOFIS is *excellent* ( $>85.5$ ), close to *best imaginable* (90.9).



**Figure 7-17. AT-LE: normalized SUS scores reported by each participant, for both training systems.**

Figure 7-18 shows the average normalized SUS score reported for each question, for the two systems. We observe that ATOFIS reported a better average score on each of the 10 dimensions.



**Figure 7-18. AT-LE: average normalized SUS scores reported by all participants on each dimension, for both training systems.**

**User preferences and remarks.** Out of 16 participants, one (i.e., P6) has no preference, one (i.e., P16) prefers Guides and 14 prefer ATOFIS. We observe that from a mental workload perspective, P6 gave the lowest ratings from all participants, for both systems (see Figure 7-16). From a usability viewpoint, P6 considered the systems equivalent, reporting a slight edge of 2.5 (on a scale from 0 to 100) points in the favor of ATOFIS. P16 was the only participant for which the average reported usability score was in the favor of Guides (95 vs. 87.5). At the same time, P16 reported a much better mental workload in the favor of ATOFIS (15.83 vs. 46.67). Even if P16 could be considered an outlier, this finding is relevant because it suggests that user preferences can be strongly influenced by the perceived usability of the system rather than by the mental workload.

**ATOFIS.** Generally, participants found the interface simple to use, flexible and intuitive. Most of the participants appreciated particularly the automatic video interaction controlled by eye gaze, allowing them to follow the assembly demonstration with no interaction effort and high control. Watching short video sequences and performing the assembly operations at the same time was appreciated. Contextualization of media, especially photo and video, was also considered an advantage for ATOFIS.

Participants revealed few concerns as well regarding the information conveyance approach adopted in ATOFIS. A first remark concerns the location arrow, which, if not correctly positioned during the authoring, can obstruct the real-world assembly environment during training. This issue did not lead to assembly errors, as noted in the reported error rates, but may cause frustration and lower UI rating. This concern can be addressed by using transparency effects on the virtual elements based on user's hand position. For example, as the hand of the user is getting closer to the assembly location, the transparency of the virtual element may increase up to the point where it becomes barely visible, or even disappears completely.

Another UI concern is related to the visual appearances of the photo/video elements, users not being able to differentiate between the two. This concern can be addressed by displaying on top of the video element the classical playback interface with play and pause buttons, offering familiar visual cues that would probably be sufficient for differentiating between the two media types. We base our hypothesis on the fact that no similar feedback was reported for the video interface used in Guides, which integrates such a visual cueing technique. We estimate that addressing this concern does not represent a technical challenge.

Furthermore, many participants stated that the dotted line used in Guides to indicate the real-world position of an assembly was more intuitive than the arrow technique used in ATOFIS. The objective evaluation demonstrates that both techniques are effective, as no errors regarding the assembly location were committed. Participants preferred the dotted line because it offers a continuous visual cueing while moving the head towards the indicated location, making it easier to follow and potentially decreasing the mental workload.

Finally, few remarks relate to the way the text instructions are displayed by using head registration, requiring in some cases hiding the text panel for avoiding occlusion. We note that unlike the training field experiment T-FE2, the organization of the workstation replicated in the present experiment requires less head movement as it is more spatially condensed. This explains why the text panel sometimes overlaps with the regions of interest of the assembly. As discussed in Section 6.3, for seated assembly workstations, the way the information is conveyed can be slightly changed (e.g., media elements displayed at fixed world position, similarity to Guides). One option of displaying the text description differently, by using another registration technique than the one proposed in ATOFIS, is by attaching it in a subtitle-like manner to the corresponding photo or video, using therefore world registration. Such an approach will somehow break the concept of showing the assembly information successively, as initially proposed with the 2W1H principle for simulating human-to-human explanation, as the text description and the associated media will be displayed at the same time. The risk is that users may skip reading the text description, therefore overlooking important assembly details which may not be obvious in the visual illustration of the assembly.

An interesting remark was made by participant 16 (the only that preferred Guides training), stating that the way of displaying the AR instructions in ATOFIS is "chaotic". The participant claimed that the information conveyance displayed in a fixed layout and always at the same real-world position is more adapted than the spatial media contextualization approach.

Consequently, more time and physical effort would be spent in following the AR instructions in exchange of knowing that the information is always displayed at the anticipated location. This idea disagrees with the general feedback of the other participants. However, it supports a key aspect discussed in Section 4.2.6 in regarding user preferences, which should be considered during the design stage. We assume that, as for the displaying technique of the text panel discussed above, the perceived benefit of the media contextualization is diminished by the workstation organization.

**Guides.** Generally, the participants found the information conveyance approach adopted by Guides intuitive and well-organized. The UI design, the ease of use and especially the assembly location indication (i.e., the dotted line) have been appreciated the most by the participants.

As for ATOFIS, a few concerns have been pointed out as well. Participants generally agreed that reading the AR instructions in Guides requires additional effort and time, mainly because of the positioning of the instruction panel, which requires the user to change focus, sometimes multiple times, between the actual assembly and the information panel. This concern was identified specifically for the video element, where most participants suggested that following the video demonstration is demanding from both physical and mental perspective. Firstly, watching the video demonstration requires the user to specifically interact with it, by clicking the “play” button either by using hand gesture or eye gaze. Participants found this interaction requirement long and unnecessary, potentially because of comparing this technique with the one proposed in ATOFIS. They claimed that following the video demonstration and performing the assembly operation at the same time is generally not possible, as they cannot keep the pace with the speed of the assembly demonstrated in the video. To address this issue, participants preferred to watch the video several times, instead of playing and pausing the video multiple times, as identifying key points in the assembly operation is difficult for novice workers. This is in line with the literature findings discussed in Section 4.2.3 regarding the transient information effect and the benefits of segmentation. We note that the video demonstrations used during the experiment are short, generally illustrating a single assembly operation. However, it seems that particularly during the WEC, some participants still need video playback control and watching the video several times. Furthermore, some participants claimed that visually connecting the details presented in the video demonstration with the real assembly is difficult because of the distance between the two, requiring attention focus switching, eventually leading to longer assembly time, increased physical demand and mental workload. This is in line with the spatial contiguity principle discussed in Section 4.2.5. Finally, some participants claimed that assembly details (e.g., Figure 7-13) may not be spotted and could require additional visual cues if the instruction panel containing the corresponding photo or video is far from the user.

The most critical issue for Guides was related to the instruction panel, particularly to the video positioning and interaction technique. This was expected, as the information conveyance method proposed in ATOFIS consists of 3 individual visual elements (i.e., text description, location and sometimes directional arrow, and assembly demonstration – photo or video), whereas in Guides it consists of one main element (i.e., the instruction panel always displayed

in front of the user), which contains the other information (i.e., text and assembly demonstration) and the location indication illustrated by the dotted line.

As a general remark for both training systems, one participant suggested using audio instructions as a complement to the visual ones, while another one suggested as well using animated instructions for conveying non-obvious assembly manipulation. By considering the findings discussed in Sections 4.2.2.6 and 4.2.3, the transient nature of audio makes it inadequate for the considered use-case, while authoring animated content does not seem to be worth the effort, when considering the overall evaluation results reported by ATOFIS.

Finally, another remark for both training systems was that no global view of the whole set of instructions was displayed before starting the assembly procedure.

#### 7.2.2.3.2 Authoring procedure

**Error rates - assembly sequence.** Two participants committed one error each while using ATOFIS, by skipping the authoring of an assembly operation. We note that participants were asked to use a printed paper where all the assembly operations were listed, as a guideline during the authoring procedure, to ensure the correct assembly sequence, the same text description, number of arrows and assembly illustration (i.e., photo or video). As the assembly sequence was learned by the participants just before the authoring procedure, during the execution of the assembly cycle twice, it was expected that all the participants would follow the recommendation rather than relying on their short-term memory. However, two of them overlooked an assembly step while following the printed instructions and pursued with the authoring of the next assembly operation. We find it reasonable to believe that a shop floor expert would not have missed skipping the authoring of an assembly step. For this reason, as soon as the examiner noticed the dictation of the wrong assembly instruction, which represents the first step of the authoring, he would ask the participant to verify again the current assembly operation number and make sure that none was overlooked. Both participants realized and fixed their mistake.

We note that 3 similar errors have occurred during the authoring with Guides, one during the media capture stage and the other two during the PC authoring. The examiner proceeded in the same way as previously described during the media capture stage, warning the participant that he might have skipped an assembly operation and the corresponding media capture. During the PC authoring, the participants have identified by themselves their mistakes, noticing the text description and the corresponding instruction number in Guides not matching the ones in the printed document. By considering these aspects, particularly the number of interventions, the motivation, and the fact that the examiner did not influence the authoring of the AR content itself, we can fairly assume that no significant impact on the overall evaluation has been induced.

**Error rates - dictation and writing.** We report on the dictation error rate, applicable during ATOFIS authoring only, for two reasons. Firstly, we wanted to evaluate the reliability of dictation services available in state-of-the-art AR devices (i.e., HoloLens 2), and secondly to identify whether the dictation errors, if not rectified, would have a significant impact on the assembly information understanding and eventually on the error rate. The dictation of the text instructions was conducted in French. Out of 16 participants, 5 were not native speakers,

therefore their dictation errors were not considered in the considered error rate analysis. From the remaining 10 participants, 5 did not commit any error, while the other 5 committed in total 10 errors: 1, 2, 2, 2 and 3, respectively. 105 words were dictated during each authoring session, which translates to an overall error rate of only 0.95%. This demonstrates the reliability of the dictation service implemented in HoloLens 2, for native speakers. By considering that no assembly error was reported during the training with ATOFIS, we suppose that either the dictation errors did not have a significant impact on the overall understanding of the corresponding text instruction, or the complementing elements of the AR instruction (i.e., arrow and assembly illustration – photo or video) were sufficient for the participant to overcome a potential misunderstanding of the text.

The writing of the text instructions during the authoring with Guides was performed by using a PC keyboard and reported only 3 errors from the native speakers, corresponding to an error rate of 0.28%. During writing, we observed participants committing many more writing errors but as expected, most of them were easily identified and rectified. Even though 4 assembly errors have been committed during the training with Guides, it is completely reasonable to assume that they were not caused at all by the text description errors. If this was the case, assembly errors would have been expected to be committed during the training with ATOFIS where the dictation error rate was 3 times higher than the writing error rate.

We note that evaluating the quality of the AR content represented by the location arrows and the assembly illustrations was not considered, because of the lack of an evaluation methodology for such content in AR, arguably very difficult to assess. The way the present experiment was conducted, by using the content authored by the previous participant for training the next one, aims to partially address this concern, by indirectly evaluating the overall quality of the AR instructions, given primarily by the number of errors committed during training and the quality of the assembly.

**Authoring time.** The reported time during both cycles is presented in Table 7-17. Unlike the training procedure where every participant performed better during the 2<sup>nd</sup> assembly cycle, independently on the AR system, during the authoring, ATOFIS reported better times than Guides for all participants and in both cases, on average by approximately 42% (i.e., 1.72x faster), as illustrated in the last column of Table 7-17. When considering both authoring systems during the first authoring stage, ATOFIS performs approximately 36% (1.56x) faster than Guides. This gap increases when comparing the usage of the two systems during the second authoring, to approximately 48% (1.94x faster) in the favor of ATOFIS. A t-test on the authoring time values reported by the participants for the two systems, shows a significant difference in the favor of ATOFIS, for both authoring cycles. This information is illustrated in Table 7-18.



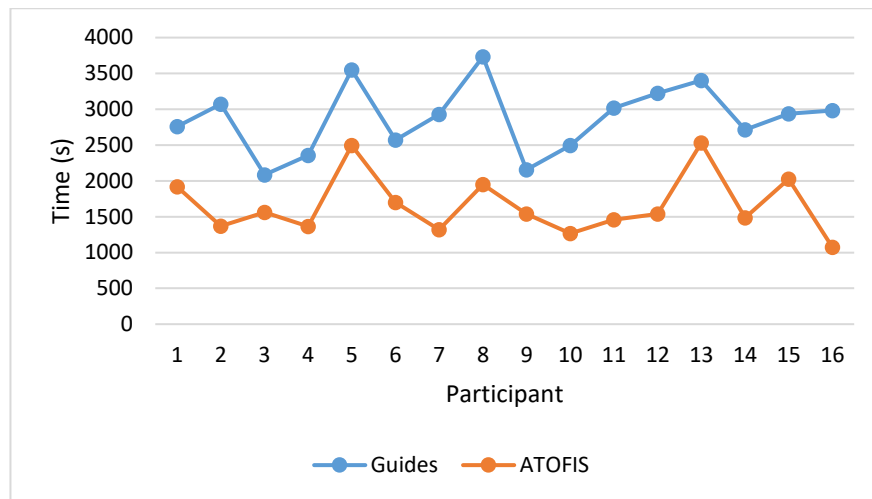
**Table 7-17. AT-LE: reported authoring times for both assembly cycles (systems).**

Participant	Training order	Guides	ATOFIS	Difference
1	(A,G)	2760	1920	30.43%
2	(G,A)	3073	1367	55.52%
3	(A,G)	2086	1560	25.22%
4	(G,A)	2356	1365	42.06%
5	(A,G)	3550	2495	29.72%
6	(G,A)	2569	1701	33.79%
7	(A,G)	2930	1318	55.02%
8	(G,A)	3733	1948	47.82%
9	(A,G)	2156	1540	28.57%
10	(G,A)	2494	1268	49.16%
11	(A,G)	3017	1457	51.71%
12	(G,A)	3224	1539	52.26%
13	(A,G)	3403	2529	25.68%
14	(G,A)	2713	1483	45.34%
15	(A,G)	2936	2026	30.99%
16	(G,A)	2984	1076	63.94%
<b>Average</b>	Both cycles	2874	1662	42.17%
	During 1 <sup>st</sup> cycle	Guides 1 <sup>st</sup>	ATOFIS 1 <sup>st</sup>	
		2854	1855	<b>35.86%</b>
	During 2 <sup>nd</sup> cycle	Guides 2 <sup>nd</sup>	ATOFIS 2 <sup>nd</sup>	
2893		1468	48.56%	

**Table 7-18. AT-LE: results of a paired two sample for means on the reported authoring times of the two systems, for both assembly cycles.**

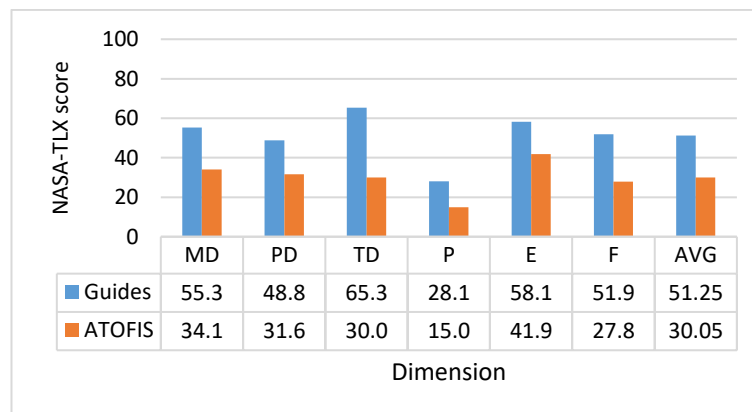
Cycle	1 <sup>st</sup>		2 <sup>nd</sup>	
	M	SD	M	SD
Guides	2893	453	2854	522
ATOFIS	1855	467	1468	268
T-test	t(7)=3.80, p=0.003		t(7)=8.23, p=0.00003	

We note that unlike the results reported for the training procedure, where all the participants reported better times during the second assembly cycle independently on the system, for authoring, all participants reported better times when using ATOFIS, independently on the assembly cycle. This information is illustrated, per participant, in Figure 7-19. It is interesting to note that ATOFIS reported a significant improvement, of approximately 21%, compared to its performance achieved by the participants during the first authoring cycle, whereas for Guides the improvement was only 1%. This potentially suggests that the authoring time can improve with experience in the case of ATOFIS, while not the same can be expected for Guides.

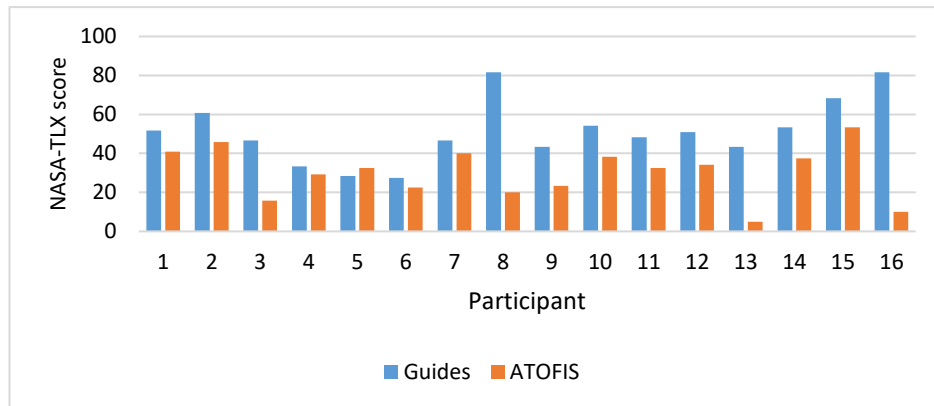


**Figure 7-19. AT-LE: authoring times reported by each participant, per cycle (authoring system).**

**NASA-TLX.** Figure 7-20 illustrates the average NASA-TLX authoring scores reported by the participants for Guides and ATOFIS, for each of the 6 dimensions. The result of a paired two-sample t-test ( $t(15)=4.14$ ,  $p=0.0004$ ) indicates that the mean values of the perceived mental workload of the participants was significantly higher for Guides ( $S=51.3$ ,  $SD=15.5$ ) than for ATOFIS ( $S=30.05$ ,  $SD=12.7$ ). Similarly, as for the training, ATOFIS reported better scores than Guides on each of the 6 dimensions as well. As expected, the overall mental workload is higher for both systems, when compared to the training, from  $S=37$  to  $S=51$  for Guides and from  $S=16$  to  $S=30$  for ATOFIS. These scores indicate that authoring with Guides was perceived by the users as being at the lower limit of *high*, while the authoring with ATOFIS was perceived as being between *medium* and *somewhat high*. As it can be observed in Figure 7-21, except one, all other participants reported a lower average perceived mental workload for ATOFIS.



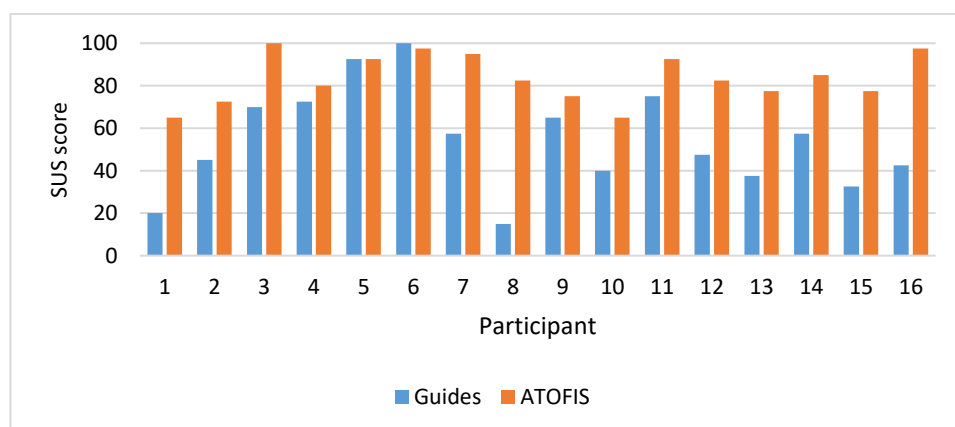
**Figure 7-20. AT-LE: average NASA-TLX scores reported by participants on each dimension, for both authoring systems.**



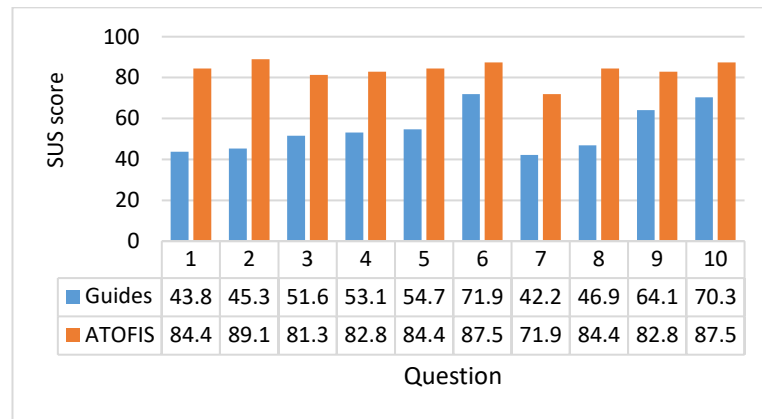
**Figure 7-21. AT-LE: NASA-TLX scores reported by each participant, for both authoring systems.**

**SUS results.** Figure 7-22 illustrates the average normalized SUS scores reported by all participants for each authoring system. The reported results indicate that from a usability perspective, 14 participants preferred ATOFIS over Guides, one (i.e., P6) had a slight preference for Guides, and one (i.e., P5) had no preference. Results of a paired t-test indicate that the average SUS authoring scores reported by the participants were significantly ( $t(15)=5.99$ ,  $p=1.23e-5$ ) lower for Guides ( $S=54.38$ ,  $SD=23.24$ ) than for ATOFIS ( $S=83.59$ ,  $SD=10.97$ ). The usability rating for Guides is therefore considered poor (between 51-68), while for ATOFIS is excellent ( $>80.3$ ). The SUS adjective rating scale proposed in (Bangor et al., 2009) indicates that the rating for Guides is *OK* (between 50.9 and 71.4), while for ATOFIS is *good* (between 71.4 and 85.5), close to *excellent*.

Figure 7-23 illustrates the average SUS authoring score reported by participants for each SUS question, for the two systems. We observe that each of the 10 SUS questions reported a better average value for ATOFIS than for Guides. For ATOFIS, 9 out of 10 questions (except Q7) reported a score higher than 80 (considered excellent); together with the overall usability score and the low value of the standard deviation ( $SD=4.78$ ) these indicate a high consistency of the proposed authoring on all SUS dimensions.



**Figure 7-22. AT-LE: normalized SUS scores reported by each participant, for both authoring systems.**



**Figure 7-23. AT-LE: average normalized SUS scores reported by all participants on each dimension, for both training systems.**

**User preferences and remarks.** All participants preferred the authoring with ATOFIS in comparison to Guides, a claim almost fully supported by the reported perceived mental workload and system usability scores. We note that P5 rated the two systems equally from a usability perspective and had a slight preference towards Guides from a mental workload viewpoint, considering that the effort with ATOFIS was higher ( $E=75$ ) than with Guides ( $E=50$ ). Overall, P5 still preferred ATOFIS. Similarly, P6 had a slight preference towards Guides from a usability perspective and vice versa when considering the perceived mental workload of the two systems, overall preferring ATOFIS as well.

It is worth noting that 62.5% (10 out of 16) of the participants rated ATOFIS with an excellent usability score ( $S>80$ ), while for Guides only 12.5% (i.e., 2 out of 16) reported a similar rating. The overall usability scores of the two systems,  $S=83$  for ATOFIS and  $S=54$  for Guides, indicate as well that our proposed authoring is generally preferred by participants by a high margin.

**ATOFIS.** The most frequent positive remarks reported by the participants regarding the authoring with ATOFIS include ease of use, speed, and flexibility. Generally, the participants appreciated that the authoring procedure is linear and clear, and entirely performed in the AR headset. For all these reasons, they considered it practical and efficient.

On the other hand, a set of negative remarks have been formulated. The most frequent comments refer to the UI design and interaction, including undetected user actions, difficult interaction with the graphical elements and the positioning of the virtual keyboard. The undetected user actions refer to those cases when user actions (e.g., button touch or voice command) are not intercepted by the system at first, generally requiring the user repeat the action or switch to another interaction technique if available. The most common example observed during the considered experiment is a user trying to interact with a virtual element by touching it, but not moving his finger enough in the direction of the element to actually touch it. This behavior can be induced by an incorrect 3D spatial estimation of the virtual element, which can be the result of the eye calibration process of Hololens 2, user's vision, or simply because of AR interaction unfamiliarity. The second most frequent example is represented by

unrecognized vocal commands, generally reported in one of the following two cases: wrong vocal command used (e.g., “*next step*” instead of “*next*”) or improper articulation (e.g., not loud enough, wrong pronunciation). We noticed that undetected user actions were generally reported at the beginning of the authoring procedure, during the creation of the first AR instructions, suggesting that some participants were not completely comfortable with the UI and the interaction techniques proposed in ATOFIS at this stage. Tracking all user actions in relation to time during the authoring, including undetected commands and other potential anomalies, would have provided more data to objectively evaluate the learning progress of the authoring application and how much time, on average, a user would need to operate the system as efficiently as possible. Based on our observation, we argue that all participants achieved a good level of expertise by the end of the authoring procedure with ATOFIS, which can roughly be translated to an average of 30 minutes at most.

The second most frequent concern identified by the participants was related to the interaction with the graphical elements, where some had difficulties at first in manipulating (e.g., moving) the captured media elements and validating them. This was due to the unfamiliarity of using the far interaction technique for grabbing and moving virtual elements in the 3D space. The concern regarding the validation button was a technical implementation issue; an incorrect “click” interaction (i.e., *long click + hand move* instead of *short click*) on this button would eventually lead to slightly translate the whole container (i.e., button + corresponding photo or video) instead of clicking the button to validate the position of the captured media.

Finally, the positioning of the virtual keyboard was a technical issue as well, not a concern specifically related to the proposed authoring approach itself. Activating the dictation mode in Hololens 2 required, at the time of the implementation, using the dictation button of the native virtual keyboard of the system. In addition, the position of the virtual keyboard was also controlled internally by the Hololens 2, by considering user’s head gaze and existing virtual elements in the AR scene. Consequently, the keyboard would not always be displayed where the participant would expect, sometimes out of the FoV or even overlapping with the authoring panel. As expected, most participants did not appreciate this behavior as occasionally, they felt the need of moving the virtual keyboard before interacting with the corresponding button for activating the dictation mode. We believe that a dedicated button for triggering the dictation mode, part of the first step authoring panel, would address this concern ideally.

Furthermore, a participant suggested hiding all visual elements, including the location arrows, while capturing the video demonstration with ATOFIS, to avoid obstruction of the assembly environment. The location arrows are always visible during the authoring of the corresponding AR instruction, to allow the author to identify whether these are obstructing relevant assembly areas while performing the assembly operation. If this was the case, the recommendation for the author would be repositioning the location arrow to avoid visual interference with the assembly location, which may lead to errors and even safety issues during training. Dynamic transparency of the virtual elements based on user’s hand position is an idea worth investigating in future work for addressing this concern.

**Guides.** Regarding the authoring with Guides, the general remark collected from the participants suggests that the overall procedure is time consuming and inefficient. The participants did not appreciate the authoring procedure being divided in 3 stages not only because of time but also because of physical considerations, as between each of the stages they had to move from one office to another. We note that the organization of the experiment was specifically designed to simulate as close as possible the real-world assembly environment where the training field study was conducted, reason for which the PC authoring was not performed in the same room with the assembly.

During the PC authoring, some participants found the matching of the captured media with the corresponding text instruction rather difficult. Other few negative remarks were made regarding the user interface. However, we have noticed that all participants have got comfortable using the PC authoring application quickly and started interacting with it efficiently very fast.

During the HMD authoring, as the participants were only required to indicate the physical locations of the assembly, not many remarks have been made. The most relevant concern was the need of redoing the assembly cycle while placing the location arrows, few of them reported having difficulty in interacting with them. This is in line with the remarks given for ATOFIS, where the far interaction technique was as well not understood by the participants right away. Unlike ATOFIS though, where moving the location arrows would be deactivated after validation, during Guides, some participants complained about moving the arrows involuntarily while performing the assembly operation.

#### 7.2.2.4 Discussion

Regarding the training procedures, the overall feedback strongly suggested that conveying manual assembly via AR should indeed consider media contextualization as well as natural and implicit interaction techniques, especially for video playback control. The chunking technique and the usage of short video demonstrations for non-obvious assembly operations or photos illustrating the final assembly, seems to be optimal to accommodate novice workers particularly during the workstation exploration cycle. Both out-of-view information access techniques reported good results, however the dotted line used in Guides was preferred by the participants overall.

Regarding the authoring, we have demonstrated that in a matter of minutes, people with neither AR expertise nor assembly experience, were able to effectively capture their assembly expertise as AR step-by-step work instructions, by using both Guides and ATOFIS authoring approaches. However, the comparative evaluation has shown that ATOFIS authoring is easier, much faster, requires less mental workload and physical effort, and preferred unanimously by the participants taking part to the experiment.

To provide unquestionable results however, replicating the full experiment in a real-world assembly environment could be necessary. We believe that evaluating the two authoring methods in a real-world environment, would lead to findings which are not obvious, like for example identifying whether the shop floor experts can easily remember the assembly sequence while not being in front of the workstation, which would have been the case during the PC

authoring with Guides. Our hypothesis is that an off-site authoring leads to more errors (skipping assembly operations), a theory that will be worth investigating in future work. Using a desktop PC might also be problematic in a shop floor environment, where assembly experts are potentially much less technical than the participants of this experiment. We believe that further similar comparative experiments will provide data that would allow improving the proposed AR information conveyance approach, potentially optimized for different assembly environments.

# Chapter 8. CONCLUSION AND FUTURE WORK

---



## 8.1 Conclusion

The main objective of this research work was to address expert knowledge transfer concerns in the context of manual assembly in Industry 4.0, by exploiting AR technologies.

### 8.1.1 Case study: manual assembly and industrial context

By considering the status of AR in industry and literature recommendations as well, we acknowledged the need of conducting an industrial case study in a manual assembly factory, to identify and define specific requirements and challenges for elaborating an optimal AR system adapted to the considered context. An in-depth analysis conducted over a period of approximately 3 years in the considered manufacturing context allowed us to test different hypothesis and approaches, and finally identify and classify the most relevant industrial requirements and challenges that a novel AR-based training system should address. In summary, it was suggested that the elaboration of the proposed AR system should prioritize usability, effectiveness, user acceptance and viability, without compromising shop floor safety and product quality. Technological and implementation concerns are not important at the organizational level, especially if the proposed tool delivers the expected results. Finally, the case study suggested that to be considered for adoption, an AR training tool should be optimally adapted to shop floor workers, generally people without AR or technical knowledge.

### 8.1.2 State-of-the-art in AR authoring and training

Further, we conducted a literature investigation regarding AR content authoring and training tools and methodologies, to understand the status of the research in the field and identify how existing AR-based training solutions address industrial concerns. We found that most proposals were elaborated outside the context of their eventual usage, without the direct implication of the potential end-users and generally evaluated in laboratory settings. Consequently, most proposals did not consider industrial challenges during design and implementation, making them inappropriate for real-world usage. It was therefore extremely difficult to identify design strategies and prioritize functional requirements for elaborating an optimal AR training system adapted to similar manufacturing context, reason for which conducting more research and especially expanding the area of research was required to address this concern.

### 8.1.3 Information representation guidelines and HCI considerations

Therefore, we have first conducted an in-depth exploration of information representation and visualization, aiming to answer at the same time an open research question which suggests that there is no agreement in the literature regarding best ways to convey instructions via AR. We have analyzed and classified multimedia data types with respect to their ability to convey information, by taking into consideration authoring concerns and most relevant AR technical challenges (e.g., precise spatial registration). To support optimal training of complex assembly operations, we have explored aspects including cognitive load, working memory and transient information effect. We have identified that short first-person video demonstrations (i.e., of single assembly tasks), complemented by spatially registered visual cues that indicate the real-world location of the corresponding assembly, represent the most reliable way of conveying complex assembly information to novice workers in AR. Furthermore, our analysis suggested

that using the three variables of the 2W1H principle formulated during the considered case study – *what*, *where* and *how*, would provide the shop floor experts with an intuitive and formalized manner of authoring the corresponding AR work instructions. To address organizational and technological concerns, we have concluded that visually representing the assembly operation by using low-cost visual assets including text, photo, video, and predefined visual assets (e.g., auxiliary 3D content) represented the optimal approach for the considered manufacturing use case. Consequently, the representation of any assembly operation, independently on the manufacturing process, environment, or user expertise, would be described in a formalized manner by using three visual assets: text (*what*), spatially registered arrow (*where*) and optionally a spatially registered FPV photo or video demonstration (*how*). To address human-centered principles and HCI considerations, we found that providing an *in-situ*, WYSIWYG authoring approach, is potentially the most adapted for shop floor experts, generally people without AR or technical knowledge. By proposing this method, we aimed as well to provide a familiar authoring paradigm in AR, by simulating human-to-human explanation of assembly operations as noted during the case study field observations. The information conveyance during the training session is presented in a similar way, for the same reason. In addition, to address cognitive considerations, we aimed to provide extremely fast learning (virtually on-the-spot) of the way the visual information is presented in AR and better retention of the assembly process as well.

Furthermore, to address user acceptance and human-centered design in the elaboration of the AR system, we have conducted a study on human-computer interactions. Our analysis suggested that system usability and usefulness are far more important than functionality and performance. We found that reducing system complexity and implicitly its power of expression to accommodate novice workers was the most advantageous approach for the considered research context. Providing an easy to use and intuitive user interface adapted to shop floor workers was identified as the top priority. From a functionality perspective, the system's reliability and predictability were the most important aspects that needed to be considered. To this purpose, we have proposed a multimodal interface that accommodates different environments and potentially user profiles. We have privileged as much as possible natural interaction techniques, while minimizing user input, as well as avoiding UI clutter and accompanying the user at any time during the procedure. The proposed approach depends only on AR functions that are available and reliable in state-of-the-art HMD devices, reason for which the usage of world registered CAD models and animations was considered as an extension of the proposed system.

#### **8.1.4 System implementation and evaluation**

A system implementation for Hololens 2 was proposed, for both authoring and training. We have conducted two field experiments (T-FE1 and T-FE2), an authoring field study and a laboratory experiment in which we evaluated the entire AR system.

The first field experiment (T-FE1) validated the hypothesis that spatially registered 2D low-cost visual assets are sufficient and effective for conveying manufacturing expertise to novice

workers via AR. In the second field experiment, T-FE2, we comparatively evaluated a CAD-complemented instruction set with the initial one (low-cost-based) to identify potential benefits of conveying assembly information by using non-animated, registered CAD models. We found that CAD data persuades lower user attentiveness, eventually leading to a higher error rate for components with a high degree of symmetry (error-prone), but to faster overall assembly completion times, particularly after the WEC. The overall reported effectiveness and usability scores are favorable, indicating that the proposed AR training method can potentially be used in concrete real world industrial use cases, with a remark that a better technique for underlining subtle assembly details is required for ensuring error-free, completely unsupervised AR training procedures. We expect that our approach can be generalized and adopted in other manufacturing use cases where the 2W1H principle can be applied.

The third experiment was a preliminary field study (AT-FS) that aimed at evaluating the authoring implementation from a time perspective. When compared to the most representative commercial AR authoring tool, Guides, the proposed authoring approach was by 35% faster. The evaluation results also showed that the proposal authoring system provides a much better time estimation for creating AR work instructions. These preliminary results indicated that the proposed authoring method eventually leads to a significant time improvement when compared to the traditional authoring and is more reliable for estimating authoring times for new assembly workstations for planning purposes.

The last study conducted during this thesis was the most relevant among all. It consisted in a complete evaluation of the proposed AR system, with a chained evaluation of both authoring and training procedures. Unlike previous evaluations, this experiment has been performed in a laboratory setting, on an improvised workstation which aimed to replicate the corresponding one from the assembly factory at elm.leblanc. All participants managed to use both systems, ATOFIS and Guides successfully, without any significant concern to report. Among the most relevant findings of the experiment, we note the following. ATOFIS reported no errors during the first cycle of the training session. This finding validated not only the correctness of the AR work instructions authored by the participants, but also the hypothesis made in FE2 which questioned whether highlighting subtle assembly details by using more detailed text descriptions would be sufficient for a zero-error report during the WEC. 18.75% of the participants committed at least one error during the WEC with Guides, suggesting that media contextualization in ATOFIS addresses the spatial contiguity principle, eventually leading to better user attentiveness and lower error rate. Regarding the assembly time, ATOFIS reported an improvement of 13% for all assembly cycles, and approximately of 19% when considering the systems during the WEC only. This result validates the time gain of the proposed information conveyance method, particularly thanks to the media contextualization and the video playback function. ATOFIS outperformed Guides in both perceived mental workload and system usability. Out of 16 participants, one had no preference, one preferred Guides and the rest preferred ATOFIS. These subjective evaluation results validate the design and usability of the system, and more importantly suggest a high user acceptance.

Considering the authoring evaluation, the most relevant findings are as follows. Regarding the error rate, few errors have been committed with both systems - i.e., 2 with ATOFIS and 3 with Guides. Such errors were committed due to the unfamiliarity of the participants with the assembly sequence, reason for which we considered the error rate insignificant. Furthermore, regarding the authoring time, ATOFIS reported better times than Guides for all participants and during both conditions - i.e., ATOFIS first and Guides second, and vice-versa, on average by approximately 42% (1.72x faster). This suggests that indeed the proposed AR authoring method (i.e., one-step, *in-situ*, HMD-only) is much faster than the traditional one which consists of multiple stages, devices, and software applications. From a mental workload perspective, ATOFIS reported a significantly better average score ( $S=30$ ) than Guides ( $S=51$ ). Similarly, the reported usability scores indicated a significant difference between Guides ( $S=54$ ) and ATOFIS ( $S=83$ ). Finally, all participants preferred authoring with ATOFIS, a claim fully supported by the reported results of NASA-TLX and SUS questionnaires. These subjective evaluation results indicate that an authoring procedure that requires a single hardware device and is executed by using a single software application in a one-stage process not only have a positive impact on the perceived usability of the system but decreases the mental effort of the user as well during the creation of the AR work instructions.

#### 8.1.5 Final remarks

Disregarding UI refining and other potential technical improvements, the overall reported results of the experiments conducted during this thesis suggest that the proposed AR training approach can be successfully implemented and used in other similar industrial contexts. We have demonstrated that low-cost visual assets, together with specific interaction and visualization techniques that are easily accessible in state-of-the-art AR devices, can provide extremely easy to use and effective AR authoring tools for capturing manual assembly expertise. Similarly, we have demonstrated that the proposed AR training system can be used effectively on the shop floor in a matter of minutes, by novice workers. At the same time, we exemplified that organizational and technical AR challenges can be surmounted, if the conception of the AR solution is elaborated and evaluated in the right context, with the direct involvement of the potential end users. We believe therefore that the industry does not need to wait for better registration techniques, improved 3D CAD content authoring, better interaction techniques and user interfaces for elaborating and adopting AR-based knowledge sharing tools for supporting human-operated shop floor processes.

Finally, the overall findings of this industrial research work suggests that a realistic usage of the proposed AR training system in a real-world use case like the manual assembly context at elm.leblanc, can be considered under the following settings. Firstly, and potentially the optimal setup would be the training of novice workers on an entire assembly line, a requirement that generally arises when a new assembly line is created. In this specific setup, the assembly completion time would eventually not represent a critical concern, as training in parallel an entire team would be advantageous overall, even though individual training times might be longer than in-person training. A second potential realistic setup would be during a half-team shift assembly, when a novice worker would have twice the nominal cycle time to execute the

assembly. In such scenario, one or more novice workers, up to half the number of the workstations of the assembly line, can perform the training at the same time. Finally, we note that the proposed AR training may be adopted in any manual assembly context where the completion time is not a constraint. A novice worker might need human assistance at the beginning of training, until he becomes comfortable following the AR instructions and executing the corresponding assembly operations. For the considered use case at elm.leblanc, this timespan would be around 10 to 15 minutes.

## 8.2 Future work

Future work should consider investigating two aspects that complement each other: (1) evaluate the proposed AR training system in large-scale field experiments including multiple assembly environments and (2) refine and potentially identify new principles and guidelines to adapt and improve the proposed AR training approach to optimally address different manual assembly contexts.

### 8.2.1 Large-scale system evaluation

To obtain unquestionable statistics regarding both the authoring and training, and eventually validate the ones reported during the experiments conducted during this thesis, replicating AT-LE experiment in a real-world assembly environment may be necessary, including shop floor experts and novice workers. Such an experiment should consist in authoring AR work instructions of at least one assembly workstation and conducting complete, but supervised training procedures with at least three novice workers. We believe that by only supervising the training procedure, shop floor experts in charge of the training would get a clear indication regarding the usability and effectiveness of the proposed AR system in a real-world use case. Furthermore, an extension of the suggested experiment should consider an entire assembly line, multiple shop floor experts for authoring the corresponding AR work instructions and a considerable number (i.e., >30) of novice workers to perform the training. Finally, the experiment should be reproduced in different manufacturing factories. We expect that such field experiments would reveal new requirements and functional priorities, that may require slight adaptations of the proposed AR training system to optimally address specific assembly environments (e.g., seated vs. standing and moving or cramped vs. ample workstation).

### 8.2.2 User interface optimization

Regarding the user interface and interaction techniques, some ideas would be worth investigating, such as implicit interaction based on proximity, for dynamically control the transparency and the scale of the virtual elements presented in the AR scene based on the user position, to address occlusions and UI clutter concerns. Future technical improvements should as well be considered to improving the implementation of the proposed system. Two of the most relevant topics include real time world registration and CAD data modeling, functionalities that can significantly improve the proposed training system. The first would remove the need of fiducial registration and ideally provide precise real time world registration, even in difficult assembly environments (e.g., bad lighting conditions, small assembly components). The second would potentially provide a much easier and intuitive way of creating CAD data, ideally without

the need of expert knowledge and by using the AR headset only. This would allow one to choose between authoring low-cost-based and CAD-based AR work instructions or both.

### 8.2.3 Automatic assembly validation and personalized training experiences

Another aspect worth investigating is providing personalized AR training experiences, by adapting the information conveyance in real-time, based on user preferences and performance. To implement such an approach, the AR system should be first able to identify and validate the actions performed by the worker during the training procedure. This would not only provide extremely valuable real-time feedback to the worker but would also provide data for objectively evaluating the quality of the AR instructions and the training performances of workers. Furthermore, an automatic assembly validation would potentially lead to a better user experience, as the information conveyance would be triggered by the system rather by worker.

### 8.2.4 Context-aware, adaptive, and automatic authoring

Finally, context-aware, and automatic authoring of AR work instructions would most definitely represent a key topic in future work. To optimally address industrial concerns, the ideal authoring system should be able to generate the AR work instructions in a fully automatic manner, perhaps just linking the system to a database where relevant information is stored. However, as shown in the literature, automatic authoring approaches are still far from industrial expectations.

An alternative, more realistic approach consists in an automatic authoring of 2W1H-based AR work instructions, *in-situ*, at the same time with the execution of a given assembly cycle. The authoring steps of this proposal, by considering the proposed 2W1H principle, are the following:

1. *What*. The system “listens” to the voice input of the expert always during the authoring. The expert provides brief vocal descriptions as he would do during a traditional training procedure, trying to synchronize the vocal input as much as possible with the assembly execution.
2. *Where*. Based on hands movement and position, voice input and eye gaze, the system detects the real-world assembly location.
3. *How*. The system records a FPV video during the whole assembly procedure. At the end, on a processing server, the video demonstration is automatically segmented per assembly operations, based on the same user input information. Further, by using voice recognition, the system decides if a photo or a video is more appropriate to illustrate the considered assembly.

We believe that such authoring approach is worth investigating first in future work, as this proposition seems feasible from an implementation perspective, without compromising the principles on which our initial AR training proposal was elaborated on.

## References

- Abraham, M., & Annunziata, M. (2017). *Augmented Reality Is Already Improving Worker Performance*. <https://hbr.org/2017/03/augmented-reality-is-already-improving-worker-performance>
- Agati, S. S., Bauer, R. D., Hounsell, M. D. S., & Paterno, A. S. (2020). Augmented Reality for Manual Assembly in Industry 4.0: Gathering Guidelines. *Proceedings - 2020 22nd Symposium on Virtual and Augmented Reality, SVR 2020, November*, 179–188. <https://doi.org/10.1109/SVR51698.2020.00039>
- Ahmadyan, A., Hou, T., Wei, J., Zhang, L., Ablavatski, A., & Grundmann, M. (2020). *Instant 3D Object Tracking with Applications in Augmented Reality*. <https://doi.org/10.48550/arxiv.2006.13194>
- Aliprantis, J., Konstantakis, M., Nikopoulou, R., Mylonas, P., & Caridakis, G. (2019). Natural interaction in augmented reality context. *CEUR Workshop Proceedings, 2320(February)*, 50–61.
- Angrisani, L., Arpaia, P., Esposito, A., & Moccaldi, N. (2020). A Wearable Brain-Computer Interface Instrument for Augmented Reality-Based Inspection in Industry 4.0. *IEEE Transactions on Instrumentation and Measurement, 69(4)*, 1530–1539. <https://doi.org/10.1109/TIM.2019.2914712>
- Arbeláez, J. C., Viganò, R., & Osorio-Gómez, G. (2019). Haptic Augmented Reality (HapticAR) for assembly guidance. *International Journal on Interactive Design and Manufacturing, 13(2)*, 673–687. <https://doi.org/10.1007/S12008-019-00532-3>
- Arf. (2019). *GR ARF 001 - V1.1.1 - Augmented Reality Framework (ARF); AR standards landscape*.
- Authoring overview - Dynamics 365 Guides PC and HoloLens | Microsoft Docs. (2021). <https://docs.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-overview>
- Ayres, P., & Paas, F. (2007a). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21(6)*, 695–700. <https://doi.org/10.1002/ACP.1343>
- Ayres, P., & Paas, F. (2007b). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology, 21(6)*, 811–820. <https://doi.org/10.1002/ACP.1351>
- Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments, 6*, 355–385.
- Baird, K. M., & Barfield, W. (1999). Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual Reality 1999 4:4, 4(4)*, 250–259. <https://doi.org/10.1007/BF01421808>
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean; adding an adjective rating. *Journal of Usability Studies, 4(3)*, 114–123.
- Barutçuoğlu, Ş., & Azizoğlu, M. (2011). Flexible assembly line design problem with fixed number of workstations. *International Journal of Production Research, 49(12)*, 3691–3714. <https://doi.org/10.1080/00207543.2010.492410>
- Battini, D., Calzavara, M., Persona, A., & Sgarbossa, F. (2015). A comparative analysis of different paperless picking systems. *Industrial Management and Data Systems, 115(3)*, 483–503. <https://doi.org/10.1108/IMDS-10-2014-0314>
- Bégout, P., Duval, T., Kubicki, S., Charbonnier, B., & Bricard, E. (2020). WAAT: a Workstation AR Authoring Tool for Industry 4.0. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12243*, 304–320. [https://doi.org/10.1007/978-3-030-58468-9\\_22](https://doi.org/10.1007/978-3-030-58468-9_22)
- Bellalouna, F. (2020). Industrial Use Cases for Augmented Reality Application. *11th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2020 - Proceedings, September*, 10–18. <https://doi.org/10.1109/CogInfoCom50765.2020.9237882>
- Berger, S., Deutch, J. M., Sanneman, L., Traficonte, D., Waldman-Brown, A., Wolters, L., & Yellen, M. (2020). *Manufacturing in America: A View from the Field*.
- Betrancourt, M. (2012). The Animation and Interactivity Principles in Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 287–296. <https://doi.org/10.1017/CBO9780511816819.019>
- Bhattacharya, B., & Winer, E. H. (2019). Augmented reality via expert demonstration authoring (AREDA). *Computers in Industry, 105*, 61–79. <https://doi.org/10.1016/j.compind.2018.04.021>

- Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior*, *89*, 411–417. <https://doi.org/10.1016/j.chb.2017.12.002>
- Billett, S. (1995). Workplace learning: Its potential and limitations. *Education + Training*, *37*(5), 20–27. <https://doi.org/10.1108/00400919510089103>
- Billinghurst, M., Kato, H., & Myojin, S. (2009). Advanced Interaction Techniques for Augmented Reality Applications. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 5622 LNCS* (Issue May 2014, pp. 13–22). [https://doi.org/10.1007/978-3-642-02771-0\\_2](https://doi.org/10.1007/978-3-642-02771-0_2)
- Biocca, F., Tang, A., Owen, C., Xiao, F., & Lab, M. (2006). Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/1124772>
- Blattgerste, J., Renner, P., Strenge, B., & Pfeiffer, T. (2018). In-Situ Instructions Exceed Side-by-Side Instructions in Augmented Reality Assisted Assembly. *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*, 133–140. <https://doi.org/10.1145/3197768.3197778>
- Blattgerste, J., Strenge, B., Renner, P., Pfeiffer, T., & Essig, K. (2017). Comparing conventional and augmented reality instructions for manual assembly tasks. *ACM International Conference Proceeding Series, Part F128530*, 75–82. <https://doi.org/10.1145/3056540.3056547>
- Boletsis, C., & Kongsvik, S. (2019). Text Input in Virtual Reality: A Preliminary Evaluation of the Drum-Like VR Keyboard. *Technologies*, *7*(2), 31. <https://doi.org/10.3390/technologies7020031>
- Bosch, T., Könemann, R., de Cock, H., & van Rhijn, G. (2017). The effects of projected versus display instructions on productivity, quality and workload in a simulated assembly task. *ACM International Conference Proceeding Series, Part F1285*, 412–415. <https://doi.org/10.1145/3056540.3076189>
- Bottani, E., & Vignali, G. (2019). Augmented reality technology in the manufacturing industry: A review of the last decade. *IISE Transactions*, *51*(3), 284–310. <https://doi.org/10.1080/24725854.2018.1493244>
- Brecher, C., Müller, S., Breitbach, T., & Lohse, W. (2013). Viable system model for manufacturing execution systems. *Procedia CIRP*, *7*, 461–466. <https://doi.org/10.1016/j.procir.2013.06.016>
- Breedveld, P. (1997). Observation, Manipulation, and Eye-Hand Coordination Problems in Minimally Invasive Surgery. IN *PROC XVI EUROPEAN ANNUAL CONFERENCE ON HUMAN DECISION MAKING AND MANUAL CONTROL*, 9–11.
- Bright, P. Lord, B., Forbes, H. Oprescu, F., Barr, N., Downer, T., Phillips, N., McTier, L., Simbag, V. & Alla, K. (2015). Expert in My Pocket: Creating First Person POV Videos to Enhance Mobile Learning. *THETA 2015: Create, Connect, Consume-Innovating Today for Tomorrow: Proceedings of the 2015 the Higher Education Technology Agenda Conference, May*, 1–15.
- Brinzer, B., Banerjee, A., & Hauth, M. (2017). Complexity Thinking and Cyber-Physical Systems. *International Journal of Industrial Engineering*, *4*(1), 14–20. <https://doi.org/10.14445/23499362/ijie-v4i1p103>
- Brooke, J. (2020). SUS: A “Quick and Dirty” Usability Scale. *Usability Evaluation In Industry, November 1995*, 207–212. <https://doi.org/10.1201/9781498710411-35>
- Bughin, J., Manyika, J., Woetzel, J., Mattern, F. M., Chui, S., Lund, A., Madgavkar, S., Ramaswamy, J., Cadena, A., Dobbs, R., George, K., Gupta, R., Hazan, E., Labaye, E., Leke, A., & Nyquist, S. (2017). A Future That Works: Automation, Employment, and Productivity. *McKinsey Global Institute, January*, 1–28.
- Bujak, K. R., Radu, I., Catrambone, R., MacIntyre, B., Zheng, R., & Golubski, G. (2013). A psychological perspective on augmented reality in the mathematics classroom. *Computers & Education*, *68*, 536–544. <https://doi.org/10.1016/J.COMPEDU.2013.02.017>
- Cao, Y., Wang, T., Qian, X., Rao, P. S., Wadhawan, M., Huo, K., & Ramani, K. (2019). GhostAR: A Time-space Editor for Embodied Authoring of Human-Robot Collaborative Task with Augmented Reality.



- Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 521–534. <https://doi.org/10.1145/3332165.3347902>
- Carroll, L., & Wiebe, E. N. (2004). Static versus Dynamic Presentation of Procedural Instruction: Investigating the Efficacy of Video-based Delivery. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(7), 1059–1063. <https://doi.org/10.1177/154193120404800712>
- Caudell, T. P., & Mizell, D. W. (1992). *Augmented reality: an application of heads-up display technology to manual manufacturing processes*. 659–669 vol.2. <https://doi.org/10.1109/HICSS.1992.183317>
- Caudell, T. P., & Mizell, D. W. (2003). *Augmented reality: an application of heads-up display technology to manual manufacturing processes*. February, 659–669 vol.2. <https://doi.org/10.1109/hicss.1992.183317>
- Ceruti, A., Marzocca, P., Liverani, A., & Bil, C. (2019). Maintenance in aeronautics in an Industry 4.0 context: The role of Augmented Reality and Additive Manufacturing. *Journal of Computational Design and Engineering*, 6(4), 516–526. <https://doi.org/10.1016/j.jcde.2019.02.001>
- Chang, M. M. L., Ong, S. K., & Nee, A. Y. C. (2017). AR-guided Product Disassembly for Maintenance and Remanufacturing. *Procedia CIRP*, 61, 299–304. <https://doi.org/10.1016/j.procir.2016.11.194>
- Clark, J. M., & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review*, 3(3).
- Clark, R. C., Nguyen, F., Sweller, J., & Baddeley, M. (2006). Efficiency in learning: Evidence-based guidelines to manage cognitive load. *Performance Improvement*, 45(9), 46–47. <https://doi.org/10.1002/PFI.4930450920>
- Closing the Industrial Skills Gap with Industrial AR | PTC*. (2019).
- Cohen, Y., Naseraldin, H., Chaudhuri, A., & Pilati, F. (2019). Assembly systems in Industry 4.0 era: a road map to understand Assembly 4.0. *International Journal of Advanced Manufacturing Technology*, 105(9), 4037–4054. <https://doi.org/10.1007/s00170-019-04203-1>
- Conway, S. (2019). *The Total Economic Impact™ Of PTC Vuforia - Cost Savings And Business Benefits Enabled By Industrial Augmented Reality* (Issue July).
- Corrêa, H. L. (1994). The Flexibility of Technological and Human Resources in Automotive Manufacturing. *Integrated Manufacturing Systems*, 5(1), 33–40. <https://doi.org/10.1108/09576069410050323>
- Courpasson, D., Clegg, F., & Clegg, S. (2012). Resisters at work: Generating productive resistance in the workplace. *Organization Science*, 23(3), 801–819. <https://doi.org/10.1287/orsc.1110.0657>
- Dabbagh, N. (2001). Authoring tools and learning systems: A historical perspective. *Proceedings of Selected Research and Development and Practice Papers Presented at the National Convention of the Association for Educational Communications and Technology*, 87–95.
- Dabbagh, N. (2002). The Evolution of Authoring Tools and Hypermedia Learning Systems: Current and Future Implications. *Educational Technology*, 42(4), 24–31.
- Damen, D., Leelasawassuk, T., Haines, O., Calway, A., & Mayol-Cuevas, W. (2014). Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video. *Proceedings of the British Machine Vision Conference 2014*, 30.1-30.13. <https://doi.org/10.5244/C.28.30>
- Daneshmand, M., Helmi, A., Avots, E., Noroozi, F., Alisinanoglu, F., Arslan, H. S., Gorbova, J., Haamer, R. E., Ozcinar, C., & Anbarjafari, G. (2018). *3D Scanning: A Comprehensive Survey*. January.
- Danielsson, O., Syberfeldt, A., Holm, M., & Wang, L. (2018). Operators perspective on augmented reality as a support tool in engine assembly. *Procedia CIRP*, 72, 45–50. <https://doi.org/10.1016/j.procir.2018.03.153>
- de Souza Cardoso, L. F., Mariano, F. C. M. Q., & Zorzal, E. R. (2020). A survey of industrial augmented reality. *Computers & Industrial Engineering*, 139(November 2019), 106159. <https://doi.org/10.1016/j.cie.2019.106159>
- Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development*, 24(3), 265–273. <https://doi.org/10.1016/j.cogdev.2009.05.002>

- Despotakis, T. C., Palaigeorgiou, G. E., & Tsoukalas, I. A. (2007). Students' attitudes towards animated demonstrations as computer learning tools. *Educational Technology and Society*, 10(1), 196–205.
- Dey, A., Billingham, M., Lindeman, R. W., & Swan, J. E. (2018). A systematic review of 10 Years of Augmented Reality usability studies: 2005 to 2014. *Frontiers Robotics AI*, 5(APR). <https://doi.org/10.3389/frobt.2018.00037>
- Doshi, A., Smith, R. T., Thomas, B. H., & Bouras, C. (2017). Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. *International Journal of Advanced Manufacturing Technology*, 89(5–8), 1279–1293. <https://doi.org/10.1007/s00170-016-9164-5>
- Dudley, J. J., Vertanen, K., & Ola Kristensson, P. (2018). Fast and precise touch-based text entry for head-mounted augmented reality with variable occlusion. *ACM Transactions on Computer-Human Interaction*, 25(6), 1–41. <https://doi.org/10.1145/3232163>
- Eckhoff, D., Sandor, C., Lins, C., Eck, U., Kalkofen, D., & Hein, A. (2018). TutAR: Augmented reality tutorials for hands-only procedures. *Proceedings - VRCAI 2018: 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, October 2020*. <https://doi.org/10.1145/3284398.3284399>
- Egger, J., & Masood, T. (2020). Augmented reality in support of intelligent manufacturing – A systematic literature review. *Computers & Industrial Engineering*, 140(February 2020), 106195. <https://doi.org/10.1016/j.cie.2019.106195>
- Elmaraghy, W., Elmaraghy, H., Tomiyama, T., & Monostori, L. (2012). Complexity in engineering design and manufacturing. *CIRP Annals - Manufacturing Technology*, 61(2), 793–814. <https://doi.org/10.1016/j.cirp.2012.05.001>
- Erkoyuncu, J. A., del Amo, I. F., Dalle Mura, M., Roy, R., & Dini, G. (2017). Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality. *CIRP Annals - Manufacturing Technology*, 66(1), 465–468. <https://doi.org/10.1016/j.cirp.2017.04.006>
- Ertelt, A., Renkl, A., & Spada, H. (2006). Making a difference: exploiting the full potential of instructionally designed on-screen videos. *Proceedings of the 7th International Conference on Learning Sciences (ICLS '06). International Society of the Learning Sciences*, 154–160. <https://doi.org/10.5555/1150034>
- Falck, A. C., Örtengren, R., Rosenqvist, M., & Söderberg, R. (2016). Criteria for Assessment of Basic Manual Assembly Complexity. *Procedia CIRP*, 44, 424–428. <https://doi.org/10.1016/j.procir.2016.02.152>
- Feldmann, K., & Slama, S. (2001). Highly flexible Assembly – Scope and Justification. *CIRP Annals*, 50(2), 489–498. [https://doi.org/https://doi.org/10.1016/S0007-8506\(07\)62987-4](https://doi.org/https://doi.org/10.1016/S0007-8506(07)62987-4)
- Fernández del Amo, I., Erkoyuncu, J. A., Roy, R., Palmarini, R., & Onoufriou, D. (2018). A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications. *Computers in Industry*, 103, 47–71. <https://doi.org/10.1016/j.compind.2018.08.007>
- Fiorentino, M., Uva, A. E., Gattullo, M., Debernardis, S., & Monno, G. (2014). Augmented reality on large screen for interactive maintenance instructions. *Computers in Industry*, 65(2), 270–278. <https://doi.org/10.1016/j.compind.2013.11.004>
- Fischer, G. (2001). User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction*, 11(1–2), 65–86. <https://doi.org/10.1023/A:1011145532042>
- Fite-Georgel, P. (2011). Is there a reality in Industrial Augmented Reality? *2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011*, 201–210. <https://doi.org/10.1109/ISMAR.2011.6092387>
- Foley D., J., van Dam, A., Feiner K., S., & Hughes F., J. (2004). *Computer Graphics: Principles and Practice*.
- Fournier-Viger, P., Nkambou, R., & Mephu Nguifo, E. (2009). Exploiting partial problem spaces learned from users' interactions to provide key tutoring services in procedural and ill-defined domains. *Frontiers in Artificial Intelligence and Applications*, 200(1), 383–390. <https://doi.org/10.3233/978-1-60750-028-5-383>

- Fraga-Lamas, P., Fernández-Caramés, T. M., Blanco-Novoa, Ó., & Vilar-Montesinos, M. A. (2018). A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard. *IEEE Access*, 6, 13358–13375. <https://doi.org/10.1109/ACCESS.2018.2808326>
- Friedrich, W. (2002). ARVIKA-augmented reality for development, production and service. *Proceedings - International Symposium on Mixed and Augmented Reality, ISMAR 2002*, 3–4. <https://doi.org/10.1109/ISMAR.2002.1115059>
- Funk, M., Kosch, T., & Schmidt, A. (2016). Interactive worker assistance: Comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. *UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, July 2018*, 934–939. <https://doi.org/10.1145/2971648.2971706>
- Funk, M., Mayer, S., & Schmidt, A. (2015). Using in-situ projection to support cognitively impaired workers at the workplace. *ASSETS 2015 - Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility*, 185–192. <https://doi.org/10.1145/2700648.2809853>
- Funk, M., Shirazi, A. S., Mayer, S., Lischke, L., & Schmidt, A. (2015). Pick from Here-An interactive mobile cart using in-situ projection for order picking. *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 601–609. <https://doi.org/10.1145/2750858.2804268>
- Gabbard, J. L., Fitch, G. M., & Kim, H. (2014). Behind the glass: Driver challenges and opportunities for AR automotive applications. *Proceedings of the IEEE*, 102(2), 124–136. <https://doi.org/10.1109/JPROC.2013.2294642>
- Gattullo, M., Evangelista, A., Uva, A. E., Fiorentino, M., & Gabbard, J. (2020). What, How, and Why are Visual Assets used in Industrial Augmented Reality? A Systematic Review and Classification in Maintenance, Assembly, and Training (from 1997 to 2019). *IEEE Transactions on Visualization and Computer Graphics*, 2626(c), 1–1. <https://doi.org/10.1109/tvcg.2020.3014614>
- Gattullo, M., Scurati, G. W., Evangelista, A., Ferrise, F., Fiorentino, M., & Uva, A. E. (2020). Informing the Use of Visual Assets in Industrial Augmented Reality. *Lecture Notes in Mechanical Engineering*, 106–117. [https://doi.org/10.1007/978-3-030-31154-4\\_10](https://doi.org/10.1007/978-3-030-31154-4_10)
- Gattullo, M., Scurati, G. W., Fiorentino, M., Uva, A. E., Ferrise, F., & Bordegoni, M. (2019). Towards augmented reality manuals for industry 4.0: A methodology. *Robotics and Computer-Integrated Manufacturing*, 56(October 2018), 276–286. <https://doi.org/10.1016/j.rcim.2018.10.001>
- 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. *Interactive Learning Environments*, 23(6), 778–798. <https://doi.org/10.1080/10494820.2013.815221>
- Gavrila, D. M. (1999). The Visual Analysis of Human Movement: A Survey. *Computer Vision and Image Understanding*, 73(1), 82–98. <https://doi.org/10.1006/CVIU.1998.0716>
- Geissbauer, R., Schrauf, S., Bertram, P., & Cheraghi, F. (2017). Digital Factories 2020: Shaping the future of manufacturing. *PricewaterhouseCoopers GmbH*, 48.
- Geng, J., Song, X., Pan, Y., Tang, J., Liu, Y., Zhao, D., & Ma, Y. (2020). A systematic design method of adaptive augmented reality work instruction for complex industrial operations. *Computers in Industry*, 119, 103229. <https://doi.org/10.1016/j.compind.2020.103229>
- Gleeson, B., Maclean, K., Croft, E., & Alcazar, J. (2013). *Gestures for Industry*. 349–356.
- Gonzalez, A. N. V., Kapalo, K., Koh, S., Sottolare, R., Garrity, P., & Laviola, J. J. (2019). A Comparison of Desktop and Augmented Reality Scenario Based Training Authoring Tools. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1199–1200. <https://doi.org/10.1109/VR.2019.8797973>
- Goto, M., Uematsu, Y., Saito, H., Senda, S., & Iketani, A. (2010). Task support system by displaying instructional video onto AR workspace. *2010 IEEE International Symposium on Mixed and Augmented Reality*, 83–90. <https://doi.org/10.1109/ISMAR.2010.5643554>

- Grabler, F., Agrawala, M., Li, W., Dontcheva, M., & Igarashi, T. (2009). Generating photo manipulation tutorials by demonstration. *ACM Transactions on Graphics (TOG)*, 28(3). <https://doi.org/10.1145/1531326.1531372>
- Gräßler, I., Roesmann, D., Cappello, C., & Steffen, E. (2021). Skill-based worker assignment in a manual assembly line. *Procedia CIRP*, 100, 433–438. <https://doi.org/10.1016/j.procir.2021.05.100>
- Grimin, P., Haller, M., & Reinhold, S. (2002). AMIRE - Authoring Mixed Reality. *Augmented Reality Toolkit, The First IEEE International Workshop*. <https://doi.org/doi:http://dx.doi.org/10.1109/ART.2002.1107008>
- Grosky, W. I. (2003). Hyper-Media Databases. In *Encyclopedia of Information Systems* (pp. 571–582). Elsevier. <https://doi.org/10.1016/b0-12-227240-4/00089-7>
- Grossman, T., & Fitzmaurice, G. (2010). ToolClips: An investigation of contextual video assistance for functionality understanding. *Conference on Human Factors in Computing Systems - Proceedings*, 3, 1515–1524. <https://doi.org/10.1145/1753326.1753552>
- Grote, G. (2004). Uncertainty management at the core of system design. *Annual Reviews in Control*, 28(2), 267–274. <https://doi.org/10.1016/j.arcontrol.2004.03.001>
- Grote, G., & Weichbrodt, J. C. (2007). *Uncertainty Management Through Flexible Routines in a High-Risk Organization Swiss Federal Institute of Technology (ETH Zürich) 2nd Annual Cambridge Conference on Regulation, Inspection & . October*.
- Gruenefeld, U., Ennenga, D., Ali, A. El, Heuten, W., & Boll, S. (2017). EyeSee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. *SUI 2017 - Proceedings of the 2017 Symposium on Spatial User Interaction*, 109–118. <https://doi.org/10.1145/3131277.3132175>
- Gupta, A., Fox, D., Curless, B., & Cohen, M. (2012). DuploTrack: A Real-time System for Authoring and Guiding Duplo Block Assembly. *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. <https://doi.org/10.1145/2380116>
- Ha, T., & Woo, W. (2007). Graphical tangible user interface for a AR authoring tool in product design environment. *CEUR Workshop Proceedings*, 260(June 2014).
- Hahn, J., Ludwig, B., & Wolff, C. (2015). Augmented reality-based training of the PCB assembly process. *ACM International Conference Proceeding Series, 30-Novembe(Mum)*, 395–399. <https://doi.org/10.1145/2836041.2841215>
- Haller, M. (2004). Photorealism or/and non-photorealism in Augmented Reality. *Proceedings VRCAI 2004 - ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry*, 189–196. <https://doi.org/10.1145/1044588.1044627>
- Hanson, R., Falkenström, W., & Miettinen, M. (2017). Augmented reality as a means of conveying picking information in kit preparation for mixed-model assembly. *Computers & Industrial Engineering*, 113(August), 570–575. <https://doi.org/10.1016/j.cie.2017.09.048>
- Haringer, M., & Regenbrecht, H. T. (2002). A pragmatic approach to augmented reality authoring. *Proceedings - International Symposium on Mixed and Augmented Reality, ISMAR 2002*, 237–246. <https://doi.org/10.1109/ISMAR.2002.1115093>
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, 89(1), 92–102. <https://doi.org/10.1037/0022-0663.89.1.92>
- Harrison, S. M. (1995). Comparison of still, animated, or nonillustrated on-line help with written or spoken instructions in a graphical user interface. *Conference on Human Factors in Computing Systems - Proceedings*, 1, 82–89. <https://doi.org/10.1145/223904.223915>
- Hart, S. (1980). *TLX @ NASA Ames - Home*. <https://humansystems.arc.nasa.gov/groups/tlx/>
- Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology*, 21(6), 713–729. <https://doi.org/10.1002/ACP.1345>

- Hategekimana, C., Gilbert, S., & Blessing, S. (2008). Effectiveness of using an intelligent tutoring system to train users on off-the-shelf software. *Society for Information Technology & Teacher Education International Conference*, 414–419.
- Haug, A. (2015). Work instruction quality in industrial management. *International Journal of Industrial Ergonomics*, 50, 170–177. <https://doi.org/10.1016/j.ergon.2015.09.015>
- Hegarty, M. (2012). Multimedia Learning About Physical Systems. *The Cambridge Handbook of Multimedia Learning*, 447–466. <https://doi.org/10.1017/CBO9780511816819.029>
- Heidemann, G., Bax, I., & Bekel, H. (2004). Multimodal interaction in an augmented reality scenario. *ICMI'04 - Sixth International Conference on Multimodal Interfaces*, 53–60. <https://doi.org/10.1145/1027933.1027944>
- Henderson, S., & Feiner, S. (2011). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics*, 17(10), 1355–1368. <https://doi.org/10.1109/TVCG.2010.245>
- Henderson, S. J., & Feiner, S. (2009). Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. *Science and Technology Proceedings - IEEE 2009 International Symposium on Mixed and Augmented Reality, ISMAR 2009*, 135–144. <https://doi.org/10.1109/ISMAR.2009.5336486>
- Hermann, M., Pentek, T., & Otto, B. (2016). *Design Principles for Industrie 4.0 Scenarios*. <https://doi.org/10.1109/HICSS.2016.488>
- Hill, E. J., Erickson, J. J., & Holmes, E. K. (2010). *Workplace Flexibility, Work Hours, and Work-Life Conflict : Finding an Extra Workplace Flexibility, Work Hours, and Work-Life Conflict : Finding an Extra Day or Two. March 2014*. <https://doi.org/10.1037/a0019282>
- Hodaie, Z., Haladjian, J., & Bruegge, B. (2019). ISAR: An authoring system for interactive tabletops. *ISS 2019 - Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, 355–360. <https://doi.org/10.1145/3343055.3360751>
- Höfler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations - Evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior*, 27(1), 209–216. <https://doi.org/10.1016/j.chb.2010.07.042>
- Höfler, T. N., Prechtel, H., & Nerdel, C. (2010). The influence of visual cognitive style when learning from instructional animations and static pictures. *Learning and Individual Differences*, 20(5), 479–483. <https://doi.org/10.1016/j.lindif.2010.03.001>
- Holm, M., Danielsson, O., Syberfeldt, A., Moore, P., & Wang, L. (2017). Adaptive instructions to novice shop-floor operators using Augmented Reality. *Journal of Industrial and Production Engineering*, 34(5), 362–374. <https://doi.org/10.1080/21681015.2017.1320592>
- Holmqvist, K., & Andersson, R. (2015). Paradigms and Measures. *Eye Tracking: A Comprehensive Guide to Methods and Measures*, 560.
- Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*, 23(9), 1215–1226. <https://doi.org/10.1002/acp.1525>
- Hu, S. J., Zhu, X., Wang, H., & Koren, Y. (2008). Product variety and manufacturing complexity in assembly systems and supply chains. *CIRP Annals - Manufacturing Technology*, 57(1), 45–48. <https://doi.org/10.1016/j.cirp.2008.03.138>
- Huang, H., Wei, Z., & Yao, L. (2019). A novel approach to component assembly inspection based on mask R-CNN and support vector machines. *Information (Switzerland)*, 10(9). <https://doi.org/10.3390/info10090282>
- Ibáñez, M. B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, 123, 109–123. <https://doi.org/10.1016/J.COMPEDU.2018.05.002>
- Irawati, S., Green, S., Billingham, M., Duenser, A., & Ko, H. (2006). An evaluation of an augmented reality multimodal interface using speech and paddle gestures. *Lecture Notes in Computer Science*

- (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4282 LNCS(January), 272–283. [https://doi.org/10.1007/11941354\\_28](https://doi.org/10.1007/11941354_28)
- ISO/IEC - ARAF. (2017). *ISO/IEC 23000-13:2017 - Information technology - Multimedia application format (MPEG-A) — Part 13: Augmented reality application format.*
- Jasche, F., Hofmann, S., & Ludwig, T. (2021, May 6). Comparison of different types of augmented reality visualizations for instructions. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411764.3445724>
- Jo, G., Oh, K., Ha, I., Lee, K., Hong, M., Neumann, U., & You, S. (2014). *A Unified Framework for Augmented Reality and Knowledge-based Systems in A Unified Framework for Augmented Reality and Knowledge-Based Systems in Maintaining Aircraft*. July.
- Kahn, S. (2013). Reducing the gap between Augmented Reality and 3D modeling with real-time depth imaging. *Virtual Reality*, 17(2), 111–123. <https://doi.org/10.1007/s10055-011-0203-0>
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539. <https://doi.org/10.1007/s10648-007-9054-3>
- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40(1), 1–17. <https://doi.org/10.1518/001872098779480587>
- Kara, S. (2004). *Manufacturing flexibility and variability: an overview*. 15(6), 466–478. <https://doi.org/10.1108/17410380410547870>
- Karray, F., Alemzadeh, M., Abou Saleh, J., & Nours Arab, M. (2008). Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent Systems*, 1(1), 137–159. <https://doi.org/10.21307/ijssis-2017-283>
- Kerbl, B., Kalkofen, D., Steinberger, M., & Schmalstieg, D. (2015). Interactive Disassembly Planning for Complex Objects. *Computer Graphics Forum*, 34(2), 287–297. <https://doi.org/10.1111/cgf.12560>
- Kim, K., Billinghamurst, M., Bruder, G., Duh, H. B. L., & Welch, G. F. (2018). Revisiting trends in augmented reality research: A review of the 2nd Decade of ISMAR (2008-2017). *IEEE Transactions on Visualization and Computer Graphics*, 24(11), 2947–2962. <https://doi.org/10.1109/TVCG.2018.2868591>
- Knierim, P., Schwind, V., Feit, A. M., Nieuwenhuizen, F., & Henze, N. (2018). Physical Keyboards in Virtual Reality. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, 2018-April*, 1–9. <https://doi.org/10.1145/3173574.3173919>
- Knöpfle, C., Weidenhausen, J., Chauvigné, L., & Stock, I. (2005). Template based authoring for AR based service scenarios. *Proceedings - IEEE Virtual Reality*, 237–240. <https://doi.org/10.1109/vr.2005.1492779>
- Koć-Januchta, M., Höffler, T., Thoma, G. B., Prectl, H., & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures – An eye-tracking study. *Computers in Human Behavior*, 68(February), 170–179. <https://doi.org/10.1016/j.chb.2016.11.028>
- Korhonen, P. J., & Wallenius, J. (2020). A picture is worth a thousand words. *International Series in Operations Research and Management Science*, 294(May), 63–74. [https://doi.org/10.1007/978-3-030-49459-9\\_8](https://doi.org/10.1007/978-3-030-49459-9_8)
- Kosslyn, S. M. (2007). *Clear and to the point: 8 psychological principles for compelling PowerPoint presentations*. 222.
- Koumaditis, K., Chinello, F., Mitkidis, P., & Karg, S. T. (2020). Effectiveness of Virtual vs. Physical Training: The Case of Assembly Tasks, Trainer’s Verbal Assistance and Task Complexity. *IEEE Computer Graphics and Applications*, July 2020, 41–56. <https://doi.org/10.1109/MCG.2020.3006330>
- Koumaditis, K., Venkute, S., Jensen, F. S., & Chinello, F. (2019). Immersive Training: Outcomes from Small Scale AR/VR Pilot-Studies. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, 2019-Janua(March 2020), 1894–1898. <https://doi.org/10.1109/VR44988.2019.9044162>

- Lai, Z. H., Tao, W., Leu, M. C., & Yin, Z. (2020). Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. *Journal of Manufacturing Systems*, 55(July 2019), 69–81. <https://doi.org/10.1016/j.jmsy.2020.02.010>
- Landauer, T. K. (1996). *The trouble with computers : usefulness, usability, and productivity*. MIT Press.
- Langlotz, T., Zingerle, M., Grasset, R., Kaufmann, H., & Reitmayr, G. (2012). AR record&replay: Situated compositing of video content in mobile augmented reality. *Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI 2012, June 2014*, 318–326. <https://doi.org/10.1145/2414536.2414588>
- Lavric, T., Bricard, E., Preda, M., & Zaharia, T. (2020). An AR Work Instructions Authoring Tool for Human-Operated Industrial Assembly Lines. *Proceedings - 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR 2020*. <https://doi.org/10.1109/AIVR50618.2020.00037>
- Lavric, T., Bricard, E., Preda, M., & Zaharia, T. (2021a). An Industry-Adapted AR Training Method for Manual Assembly Operations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13095 LNCS, 282–304. [https://doi.org/10.1007/978-3-030-90963-5\\_22](https://doi.org/10.1007/978-3-030-90963-5_22)
- Lavric, T., Bricard, E., Preda, M., & Zaharia, T. (2021b). Exploring low-cost visual assets for conveying assembly instructions in AR. *2021 International Conference on INnovations in Intelligent Systems and Applications, INISTA 2021 - Proceedings*. <https://doi.org/10.1109/INISTA52262.2021.9548570>
- Lavric, T., Bricard, E., Preda, M., & Zaharia, T. (2022a). A Low-Cost AR Training System for Manual Assembly. *Computer Science and Information Systems*, 20(1), 1–27. <https://doi.org/10.2298/CSIS211123013L>
- Lavric, T., Bricard, E., Preda, M., & Zaharia, T. (2022b). ATOFIS , an AR Training System for Manual Assembly : A Full Comparative Evaluation against Guides. *21st IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, n/a.
- Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science*, 44(1), 107–123. <https://doi.org/10.1007/s11251-015-9362-9>
- Lee, G. A., & Hoff, W. (2020). *Enhancing First-Person View Task Instruction Videos with Augmented Reality Cues*. 666–676. <https://doi.org/10.1109/ISMAR50242.2020.00078>
- Lee, G. A., & Kim, G. J. (2009). Immersive authoring of Tangible Augmented Reality content: A user study. *Journal of Visual Languages and Computing*, 20(2), 61–79. <https://doi.org/10.1016/j.jvlc.2008.07.001>
- Lee, G. A., Kim, G. J., & Billingham, M. (2005). Immersive authoring: What You eXperience Is What You Get (WYXIWYG). *Communications of the ACM*, 48(7), 76–81.
- Lee, G., Ahn, S., Hoff, W., & Billingham, M. (2019). AR Tips: Augmented First-Person View Task Instruction Videos. *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 34–36. <https://doi.org/10.1109/ISMAR-Adjunct.2019.00024>
- Lee, L. H., Yung Lam, K., Yau, Y. P., Braud, T., & Hui, P. (2019). HIBEY: Hide the keyboard in augmented reality. *2019 IEEE International Conference on Pervasive Computing and Communications, PerCom 2019, January*. <https://doi.org/10.1109/PERCOM.2019.8767420>
- Lee, Y., & Kim, G. J. (2017). Vitty: Virtual touch typing interface with added finger buttons. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10280, 111–119. [https://doi.org/10.1007/978-3-319-57987-0\\_9](https://doi.org/10.1007/978-3-319-57987-0_9)
- Li, D., Mattsson, S., Salunkhe, O., Fast-Berglund, A., Skoogh, A., & Broberg, J. (2018). Effects of Information Content in Work Instructions for Operator Performance. *Procedia Manufacturing*, 25, 628–635. <https://doi.org/10.1016/j.promfg.2018.06.092>
- Li, W., Agrawala, M., & Salesin, D. (2004). Interactive image-based exploded view diagrams. *Proceedings - Graphics Interface, June 2014*, 203–212.
- Li, W., Wang, J., Jiao, S., Wang, M., & Li, S. (2019). Research on the visual elements of augmented reality assembly processes. *Virtual Reality & Intelligent Hardware*, 1(6), 622–634. <https://doi.org/10.1016/j.vrih.2019.09.006>

- Lindberg, C. F., Tan, S., Yan, J., & Starfelt, F. (2015). Key Performance Indicators Improve Industrial Performance. *Energy Procedia*, 75, 1785–1790. <https://doi.org/10.1016/j.egypro.2015.07.474>
- Lines, B. C., Sullivan, K. T., Smithwick, J. B., & Mischung, J. (2015). Overcoming resistance to change in engineering and construction: Change management factors for owner organizations. *International Journal of Project Management*, 33(5), 1170–1179. <https://doi.org/10.1016/j.ijproman.2015.01.008>
- Liu, G., He, B., Liu, S., & Huang, J. (2019). Chassis assembly detection and identification based on deep learning component instance segmentation. *Symmetry*, 11(8). <https://doi.org/10.3390/sym11081001>
- Liu, L., Li, H., & Gruteser, M. (2019). Edge assisted real-time object detection for mobile augmented reality. *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*. <https://doi.org/10.1145/3300061.3300116>
- Loch, F., Quint, F., & Brishtel, I. (2016). Comparing Video and Augmented Reality Assistance in Manual Assembly. *2016 12th International Conference on Intelligent Environments (IE)*, 147–150. <https://doi.org/10.1109/IE.2016.31>
- Lorenz, M., Knopp, S., & Klimant, P. (2018). Industrial Augmented Reality: Requirements for an Augmented Reality Maintenance Worker Support System. *Adjunct Proceedings - 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2018*, 151–153. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00055>
- Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction*, 13(2), 157–176. [https://doi.org/10.1016/S0959-4752\(02\)00018-X](https://doi.org/10.1016/S0959-4752(02)00018-X)
- Lu, F., Davari, S., Lisle, L., Li, Y., & Bowman, D. A. (2020). Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 930–939. <https://doi.org/10.1109/VR46266.2020.00113>
- Lu, X., Yu, D., Liang, H.-N., Xu, W., Chen, Y., Li, X., & Hasan, K. (2020). *Exploration of Hands-free Text Entry Techniques For Virtual Reality*. 465–470. <https://doi.org/10.1109/ISMAR50242.2020.00061>
- Lund, S., James, M., Jonathan, W., Eduard, B., Mekala, K., Knut, A., Michael, B., Katy, G., Sven, S., Daniel, S., & Kyle, H. (2020). *Risk, resilience, and rebalancing in global value chains*.
- Lv, Z., Halawani, A., Feng, S., Li, H., & Réhman, S. U. (2014). Multimodal Hand and Foot Gesture Interaction for Handheld Devices. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11(10). <https://doi.org/10.1145/2645860>
- Manifest® AR Platform Makes Everyone an Expert | Taqtile. (2022, October 4). <https://taqtile.com/manifest/>
- Marcus, N., Cleary, B., Wong, A., & Ayres, P. (2013). Should hand actions be observed when learning hand motor skills from instructional animations? *Computers in Human Behavior*, 29(6), 2172–2178. <https://doi.org/10.1016/j.chb.2013.04.035>
- Markouzis, D., & Fessakis, G. (2015). Interactive storytelling and mobile augmented reality applications for learning and entertainment - A rapid prototyping perspective. *Proceedings of 2015 International Conference on Interactive Mobile Communication Technologies and Learning, IMCL 2015*, 4–8. <https://doi.org/10.1109/IMCTL.2015.7359544>
- Marr, B. (2018). *The Amazing Ways Honeywell Is Using Virtual And Augmented Reality To Transfer Skills To Millennials*. <https://www.forbes.com/sites/bernardmarr/2018/03/07/the-amazing-ways-honeywell-is-using-virtual-and-augmented-reality-to-transfer-skills-to-millennials/?sh=2eb075c5536a>
- Martinetti, A., Marques, H. C., Singh, S., & van Dongen, L. (2019). Reflections on the Limited Pervasiveness of Augmented Reality in Industrial Sectors. *Applied Sciences*, 9(16), 3382. <https://doi.org/10.3390/app9163382>
- Masood, T., & Egger, J. (2019). Augmented reality in support of Industry 4.0—Implementation challenges and success factors. *Robotics and Computer-Integrated Manufacturing*, 58(March), 181–195. <https://doi.org/10.1016/j.rcim.2019.02.003>



- Masood, T., & Egger, J. (2020). Adopting augmented reality in the age of industrial digitalisation. *Computers in Industry*, *115*, 103112. <https://doi.org/10.1016/j.compind.2019.07.002>
- Massa, L. J., & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences*, *16*(4), 321–335. <https://doi.org/10.1016/j.lindif.2006.10.001>
- Mattsson, S., Fast-Berglund, & Li, D. (2016). Evaluation of Guidelines for Assembly Instructions. *IFAC-PapersOnLine*, *49*(12), 209–214. <https://doi.org/10.1016/j.ifacol.2016.07.598>
- Mattsson, S., Li, D., & Fast-Berglund, Å. (2018). Application of design principles for assembly instructions – Evaluation of practitioner use. *Procedia CIRP*, *76*, 42–47. <https://doi.org/10.1016/j.procir.2018.02.011>
- Mavrikios, D., Karabatsou, V., Pappas, M., & Chryssolouris, G. (2007). An efficient approach to human motion modeling for the verification of human-centric product design and manufacturing in virtual environments. *Robotics and Computer-Integrated Manufacturing*, *23*(5), 533–543. <https://doi.org/10.1016/j.rcim.2006.05.009>
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.
- Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, *93*(2), 390–397. <https://doi.org/10.1037//0022-0663.93.2.390>
- Mayer, R. E., & Fiorella, L. (2014). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. *The Cambridge Handbook of Multimedia Learning, Second Edition*, 279–315. <https://doi.org/10.1017/CBO9781139547369.015>
- Mengoni, M., Ceccacci, S., Generosi, A., & Leopardi, A. (2018). Spatial Augmented Reality: An application for human work in smart manufacturing environment. *Procedia Manufacturing*, *17*, 476–483. <https://doi.org/10.1016/j.promfg.2018.10.072>
- Merino, L., Schwarzl, M., Kraus, M., Sedlmair, M., Schmalstieg, D., & Weiskopf, D. (2020). *Evaluating Mixed and Augmented Reality: A Systematic Literature Review (2009-2019)*. <https://doi.org/10.1109/ISMAR50242.2020.00069>
- Metters, R. (1998). General rules for production planning with seasonal demand. *International Journal of Production Research*, *36*(5), 1387–1399. <https://doi.org/10.1080/002075498193381>
- Michalos, G., Makris, S., Papakostas, N., Mourtzis, D., & Chryssolouris, G. (2010). Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. *CIRP Journal of Manufacturing Science and Technology*, *2*(2), 81–91. <https://doi.org/10.1016/j.cirpj.2009.12.001>
- Michalos, G., Makris, S., Rentzos, L., & Chryssolouris, G. (2010). CIRP Journal of Manufacturing Science and Technology Dynamic job rotation for workload balancing in human based assembly systems. *CIRP Journal of Manufacturing Science and Technology*, *2*(3), 153–160. <https://doi.org/10.1016/j.cirpj.2010.03.009>
- Microsoft Mixed Reality / AR Guides | Microsoft Dynamics 365. (2019). <https://dynamics.microsoft.com/en-us/mixed-reality/guides/>
- Microsoft MRTK v2.3.0. (2020). <https://github.com/microsoft/MixedRealityToolkit-Unity/releases/tag/v2.3.0>
- Microsoft MRTK v2.4.0. (2020). <https://github.com/microsoft/MixedRealityToolkit-Unity/releases/tag/v2.4.0>
- Mijksehaar, Paul., & Westendorp, Piet. (1999). *Open here : The art of instructional design*. 142.
- Miller, M., Griese, D., Peterson, M., Summers, J. D., & Mocko, G. M. (2012). Reasoning: Installation process step instructions as an automated assembly time estimation tool. *Proceedings of the ASME Design Engineering Technical Conference*, *2*(PARTS A AND B), 223–230. <https://doi.org/10.1115/DETC2012-70109>
- Miqueo, A., Torralba, M., & Yagüe-Fabra, J. A. (2020). Lean manual assembly 4.0: A systematic review. *Applied Sciences (Switzerland)*, *10*(23), 1–37. <https://doi.org/10.3390/app10238555>

- Mital, A., Pennathur, A., Huston, R. L., Thompson, D., Pittman, M., Markle, G., Kaber, D. B., Crumpton, L., Bishu, R. R., Rajurkar, K. P., Rajan, V., Fernandez, J. E., McMulkin, M., Deivanayagam, S., Ray, P. S., & Sule, D. (1999). The need for worker training in advanced manufacturing technology (AMT) environments: A white paper. *International Journal of Industrial Ergonomics*, 24(2), 173–184. [https://doi.org/10.1016/S0169-8141\(98\)00024-9](https://doi.org/10.1016/S0169-8141(98)00024-9)
- Mixed Reality direct manipulation with hands | Microsoft Docs.* (2020). <https://docs.microsoft.com/en-us/windows/mixed-reality/design/direct-manipulation#3d-object-manipulation>
- Mixed Reality instinctual interactions | Microsoft Docs.* (2020). <https://docs.microsoft.com/en-us/windows/mixed-reality/design/interaction-fundamentals>
- Mixed Reality Toolkit | Microsoft Docs.* (2020). <https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/>
- Moghaddam, M., Wilson, N. C., Modestino, A. S., Jona, K., & Marsella, S. C. (2021). Exploring augmented reality for worker assistance versus training. *Advanced Engineering Informatics*, 50(April), 101410. <https://doi.org/10.1016/j.aei.2021.101410>
- Mohr, P., Kerbl, B., Donoser, M., Schmalstieg, D., & Kalkofen, D. (2015). Retargeting Technical Documentation to Augmented Reality. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, 2015-April*, 3337–3346. <https://doi.org/10.1145/2702123.2702490>
- Mohr, P., Mandl, D., Tatzgern, M., Veas, E., Schmalstieg, D., & Kalkofen, D. (2017). Retargeting video tutorials showing tools with surface contact to augmented reality. *Conference on Human Factors in Computing Systems - Proceedings, 2017-May*, 6547–6558. <https://doi.org/10.1145/3025453.3025688>
- Mooser, J., You, S., & Neumann, U. (2007). Real-time object tracking for augmented reality combining graph cuts and optical flow. *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR*, 145–152. <https://doi.org/10.1109/ISMAR.2007.4538839>
- Mossel, A., & Venditti, B. (2013). *3DTouch and HOMER-S: Intuitive Manipulation Techniques for One-Handed Handheld Augmented Reality.*
- Mota, R. C., Roberto, R. A., & Teichrieb, V. (2015). [POSTER] Authoring Tools in Augmented Reality: An Analysis and Classification of Content Design Tools. *2015 IEEE International Symposium on Mixed and Augmented Reality, October 2017*, 164–167. <https://doi.org/10.1109/ISMAR.2015.47>
- Moura, G. D. S., Pessoa, S. A., Lima, J. P. S. D. M., Teichrieb, V., & Kelner, J. (2011). RPR-SORS: An authoring toolkit for photorealistic AR. *Proceedings - 2011 13th Symposium on Virtual Reality, SVR 2011*, 178–187. <https://doi.org/10.1109/SVR.2011.14>
- Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2020). A framework for automatic generation of augmented reality maintenance & repair instructions based on convolutional Neural networks. *Procedia CIRP*, 93, 977–982. <https://doi.org/10.1016/j.procir.2020.04.130>
- Mourtzis, D., Vlachou, A., & Zogopoulos, V. (2017). Cloud-based augmented reality remote maintenance through shop-floor monitoring: A product-service system approach. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 139(6). <https://doi.org/10.1115/1.4035721/366621>
- MTM (METHODS-TIME MEASUREMENT).* (n.d.). Retrieved March 21, 2022, from <https://mtm.org/en/about-mtm/mtm>
- MTM-UAS | MTM.* (n.d.). Retrieved March 21, 2022, from <https://www.ukmtm.co.uk/systems/mtm-uas/>
- Muhammad Nizam, S. S., Zainal Abidin, R., Che Hashim, N., Lam, M. C., Arshad, H., & Abd Majid, N. A. (2018). A Review of Multimodal Interaction Technique in Augmented Reality Environment. *International Journal on Advanced Science, Engineering and Information Technology*, 8(4–2), 1460. <https://doi.org/10.18517/ijaseit.8.4-2.6824>
- Murithi, J., & Lin, C. (2020). Markerless cooperative augmented reality-based smart manufacturing double-check system : Case of safe PCBA inspection following automatic optical inspection. *Robotics*

- and Computer Integrated Manufacturing*, 64(February), 101957.  
<https://doi.org/10.1016/j.rcim.2020.101957>
- NAM. (2020). *NAM Manufacturers' Outlook Survey Third Quarter 2020*.
- Nassani, A., Bai, H., Lee, G., & Billingham, M. (2015). Tag it!: AR annotation using wearable sensors. *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications, SA 2015*.  
<https://doi.org/10.1145/2818427.2818438>
- Nebeling, M., & Speicher, M. (2018). The Trouble with Augmented Reality/Virtual Reality Authoring Tools. *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 333–337. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00098>
- Nee, A. Y. C., Ong, S. K., Chryssolouris, G., & Mourtzis, D. (2012). Augmented reality applications in design and manufacturing. *CIRP Annals*, 61(2), 657–679. <https://doi.org/10.1016/j.cirp.2012.05.010>
- Neumann, W. P., Kolus, A., & Wells, R. W. (2016). Human Factors in Production System Design and Quality Performance – A Systematic Review. *IFAC-PapersOnLine*, 49(12), 1721–1724.  
<https://doi.org/10.1016/j.ifacol.2016.07.830>
- Nielsen, J. (1993). *Usability engineering*. Academic Press.
- Nilsson, S., & Johansson, B. (2007). Fun and usable: Augmented Reality instructions in a hospital setting. *Australasian Computer-Human Interaction Conference, OZCHI'07*, 123–130.
- Nishihara, A., & Okamoto, J. (2015). Object recognition in assembly assisted by augmented reality system. *IntelliSys 2015 - Proceedings of 2015 SAI Intelligent Systems Conference*, 400–407.  
<https://doi.org/10.1109/IntelliSys.2015.7361172>
- Nooruddin, N., Dembani, R., & Maitlo, N. (2020). HGR: Hand-Gesture-Recognition Based Text Input Method for AR/VR Wearable Devices. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020-Octob*, 744–751. <https://doi.org/10.1109/SMC42975.2020.9283348>
- Novaković, D., Milić, N., & Milosavljević, B. (2013). Animated vs. illustrated software tutorials: Screencasts for acquisition and screenshots for recalling. *International Journal of Engineering Education*, 29(4), 1013–1023.
- O'Donnell, R. (2018). *How Mercedes-Benz uses augmented reality to train employees of all types | HR Dive*. <https://www.hrdiver.com/news/how-mercedes-benz-uses-augmented-reality-to-train-employees-of-all-types/530425/>
- Oestreich, H., Töniges, T., Wojtynek, M., & Wrede, S. (2019). Interactive Learning of Assembly Processes using Digital Assistance. *Procedia Manufacturing*, 31(January), 14–19.  
<https://doi.org/10.1016/j.promfg.2019.03.003>
- OGC - ARML 2.0. (2012). *OGC® Augmented Reality Markup Language 2.0 (ARML 2.0) | OGC*.  
<https://www.ogc.org/standards/arml>
- Ong, S. K., Yuan, M. L., & Nee, A. Y. C. (2008). Augmented reality applications in manufacturing: a survey. <http://dx.doi.org/10.1080/00207540601064773>, 46(10), 2707–2742.  
<https://doi.org/10.1080/00207540601064773>
- Ong, S. K., & Zhu, J. (2013). CIRP Annals - Manufacturing Technology A novel maintenance system for equipment serviceability improvement. *CIRP Annals - Manufacturing Technology*, 62(1), 39–42.  
<https://doi.org/10.1016/j.cirp.2013.03.091>
- Ong, X. W. S. K., & Nee, A. Y. C. (2016). A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4(1), 1–22. <https://doi.org/10.1007/s40436-015-0131-4>
- Otte, A., Schneider, D., Menzner, T., Gesslein, T., Gagel, P., & Grubert, J. (2019). Evaluating text entry in virtual reality using a touch-sensitive physical keyboard. *Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2019*, 387–392.  
<https://doi.org/10.1109/ISMAR-Adjunct.2019.000-4>
- Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., & Ferro, D. (2000). Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. *Human-Computer Interaction*, 15(4), 263–322. [https://doi.org/10.1207/S15327051HCI1504\\_1](https://doi.org/10.1207/S15327051HCI1504_1)

- Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford University Press.
- Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, 5(2), 176–206. [https://doi.org/10.1016/0010-0285\(73\)90032-7](https://doi.org/10.1016/0010-0285(73)90032-7)
- Palladino, T. (2017). *Porsche Adopts Atheer's AR Platform to Connect Mechanics with Remote Experts | Next Reality*. <https://next.reality.news/news/porsche-adopts-atheers-ar-platform-connect-mechanics-with-remote-experts-0181255/>
- Palmarini, R., Erkoyuncu, J. A., Roy, R., & Torabmostaedi, H. (2018). A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, 49(March 2017), 215–228. <https://doi.org/10.1016/j.rcim.2017.06.002>
- Palmiter, S., & Elkerton, J. (1991). An evaluation of animated demonstrations for learning computer-based tasks. *Conference on Human Factors in Computing Systems - Proceedings*, 257–263. <https://doi.org/10.1145/108844.108906>
- Palmiter, S., & Elkerton, J. (1993). Animated demonstrations for learning procedural computer-based tasks. *Human-Computer Interaction*, 8(3), 193–216. [https://doi.org/10.1207/S15327051HCI0803\\_1](https://doi.org/10.1207/S15327051HCI0803_1)
- Paramarthalingam, A. (2016). Machine parts recognition and defect detection in automated assembly systems using computer vision techniques. *Rev. Téc. Ing. Univ. Zulia., January*. <https://doi.org/10.21311/001.39.1.08>
- Park, J., Kim, S. S., Park, H., & Woo, W. (2016). Dreamhouse: NUI-based photo-realistic AR authoring system for interior design. *ACM International Conference Proceeding Series*, 25-27-Febr. <https://doi.org/10.1145/2875194.2875221>
- Petersen, N., & Stricker, D. (2012). Learning task structure from video examples for workflow tracking and authoring. *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality 2012, Science and Technology Papers, November*, 237–246. <https://doi.org/10.1109/ISMAR.2012.6402562>
- Pilati, F., Faccio, M., Gamberi, M., & Regattieri, A. (2020). Learning manual assembly through real-time motion capture for operator training with augmented reality. *Procedia Manufacturing*, 45, 189–195. <https://doi.org/10.1016/j.promfg.2020.04.093>
- Plaisant, C., & Shneiderman, B. (2005). Show me! guidelines for producing recorded demonstrations. *Proceedings - 2005 IEEE Symposium on Visual Languages and Human-Centric Computing, 2005*, 171–178. <https://doi.org/10.1109/VLHCC.2005.57>
- Polvi, J., Taketomi, T., Moteki, A., Yoshitake, T., Fukuoka, T., Yamamoto, G., Sandor, C., & Kato, H. (2018). Handheld Guides in Inspection Tasks: Augmented Reality versus Picture. *IEEE Transactions on Visualization and Computer Graphics*, 24(7), 2118–2128. <https://doi.org/10.1109/TVCG.2017.2709746>
- Pongnumkul, S., Dontcheva, M., Li, W., Wang, J., Bourdev, L., Avidan, S., & Cohen, M. F. (2011). Pause-and-play: Automatically linking screencast video tutorials with applications. *UIST'11 - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 135–144. <https://doi.org/10.1145/2047196.2047213>
- Popelka, O., Prochazka, D., Kolomaznik, J., Landa, J., & Koubek, T. (2012). Adaptive real-time object recognition for augmented reality. *Mendel, May 2014*, 526–531.
- Raczynski, A., & Gussmann, P. (2004). Services and training through augmented reality. *IEE Conference Publication*, 263–271.
- Radkowski, R., Herrema, J., & Oliver, J. (2015). Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. *International Journal of Human-Computer Interaction*, 31(5), 337–349. <https://doi.org/10.1080/10447318.2014.994194>
- Rasmussen, J. (1983). and Symbols , and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3), 257–266.
- Reed, S. K. (2013). Thinking Visually. *Thinking Visually*, 1–222. <https://doi.org/10.4324/9780203848715/THINKING-VISUALLY-STEPHEN-REED>

- Reiners, D., Stricker, D., Klinker, G., & Stefan, M. (1999). Augmented Reality for Construction Tasks: Doorlock Assembly. *Augmented Reality*, November 1999, 51–66. <https://doi.org/10.1201/9781439863992-10>
- Robertson, C. M., MacIntyre, B., & Walker, B. N. (2008). An evaluation of graphical context when the graphics are outside of the task area. *Proceedings - 7th IEEE International Symposium on Mixed and Augmented Reality 2008, ISMAR 2008*, 73–76. <https://doi.org/10.1109/ISMAR.2008.4637328>
- Rolland, J. P., Holloway, R. L., & Fuchs, H. (1995). Comparison of optical and video see-through, head-mounted displays. *Telemicrooperator and Telepresence Technologies*, 2351, 293–307. <https://doi.org/10.1117/12.197322>
- Romero, D., Bernus, P., Noran, O., Stahre, J., & Berglund, Å. F. (2016). The Operator 4.0: Human Cyber-Physical Systems & Adaptive Automation Towards Human-Automation Symbiosis Work Systems. *IFIP Advances in Information and Communication Technology*, 488, 677–686. [https://doi.org/10.1007/978-3-319-51133-7\\_80](https://doi.org/10.1007/978-3-319-51133-7_80)
- Rumiński, D., & Walczak, K. (2013). Creation of interactive AR content on mobile devices. *Lecture Notes in Business Information Processing*, 160, 258–269. [https://doi.org/10.1007/978-3-642-41687-3\\_24](https://doi.org/10.1007/978-3-642-41687-3_24)
- Sahu, C. K., Young, C., & Rai, R. (2021). Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review. *International Journal of Production Research*, 59(16), 4903–4959. <https://doi.org/10.1080/00207543.2020.1859636>
- Sanna, A., Manuri, F., Lamberti, F., Paravati, G., & Pezzolla, P. (2015). Using handheld devices to support augmented reality-based maintenance and assembly tasks. *2015 IEEE International Conference on Consumer Electronics, ICCE 2015*, 178–179. <https://doi.org/10.1109/ICCE.2015.7066370>
- Sato, H., & Cohen, M. (2010). Using motion capture for real-time augmented reality scenes. *Proceedings of the 13th International Conference on ... , December 2010*, 58–61.
- Schneider, P. (2018). Managerial challenges of Industry 4.0: an empirically backed research agenda for a nascent field. In *Review of Managerial Science* (Vol. 12, Issue 3). Springer Berlin Heidelberg. <https://doi.org/10.1007/s11846-018-0283-2>
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Schwab, K. (2016). *The Fourth Industrial Revolution*.
- Schwald, B., & de Laval, B. D. (2003). An Augmented Reality System for Training and Assistance to Maintenance in the Industrial Context. *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, 425–432.
- Schwald, B., Figue, J., Chauvineau, E., Vu-Hong, F., Robert, A., Arbolino, M., Schnaider, M., Laval, B. de, Dumas de Raully, F., Anez, F. G., Baldo, O., & Santos, J. (2001). STARMATE: Using augmented reality technology for computer guided maintenance of complex mechanical elements. *EBusiness and EWork Conference*, 196–202.
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, 14(3), 293–305. <https://doi.org/10.1016/j.learninstruc.2004.06.005>
- Schwerdtfeger, B., Reif, R., Günthner, W. A., & Klinker, G. (2011). Pick-by-vision: There is something to pick at the end of the augmented tunnel. *Virtual Reality*, 15(2–3), 213–223. <https://doi.org/10.1007/S10055-011-0187-9>
- Scurati, G. W., Gattullo, M., Fiorentino, M., Ferrise, F., Bordegoni, M., & Uva, A. E. (2018). Converting maintenance actions into standard symbols for Augmented Reality applications in Industry 4.0. *Computers in Industry*, 98, 68–79. <https://doi.org/10.1016/j.compind.2018.02.001>
- Seppänen, J. (2006). *The Power of the Gaze: An Introduction to Visual Literacy*. 145.
- Sethi, A. K., & Sethi, S. P. (1990). Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems*, 2(4), 289–328. <https://doi.org/10.1007/BF00186471>
- Shao, T., Li, W., Zhou, K., Xu, W., Guo, B., & Mitra, N. J. (2013). Interpreting concept sketches. *ACM Transactions on Graphics (TOG)*, 32(4). <https://doi.org/10.1145/2461912.2462003>

- Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2018). *Designing the User Interface Strategies for Effective Human-Computer Interaction by Ben Shneiderman Catherine Plaisant Maxine Cohen Steven Jacobs Niklas Elmqvist Nicholas Diakopoulos*. 623.
- Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. *Conference on Human Factors in Computing Systems - Proceedings*, 281–288. <https://doi.org/10.1145/332040.332445>
- Siew, C. Y., Ong, S. K., & Nee, A. Y. C. (2019). A practical augmented reality-assisted maintenance system framework for adaptive user support. *Robotics and Computer-Integrated Manufacturing*, 59, 115–129. <https://doi.org/10.1016/j.rcim.2019.03.010>
- Singh, A. M., Marcus, N., & Ayres, P. (2012). The transient information effect: Investigating the impact of segmentation on spoken and written text. *Applied Cognitive Psychology*, 26(6), 848–853. <https://doi.org/10.1002/acp.2885>
- Smith, E., McRae, K., Semple, G., Welsh, H., Evans, D., & Blackwell, P. (2021). Enhancing Vocational Training in the Post-COVID Era through Mobile Mixed Reality. *Sustainability 2021, Vol. 13, Page 6144*, 13(11), 6144. <https://doi.org/10.3390/SU13116144>
- Smith, E., Semple, G., Evans, D., McRae, K., & Blackwell, P. (2020). Augmented Instructions: Analysis of Performance and Efficiency of Assembly Tasks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12191 LNCS, 166–177. [https://doi.org/10.1007/978-3-030-49698-2\\_12](https://doi.org/10.1007/978-3-030-49698-2_12)
- Souza, J. M. B. de, & Dyson, M. C. (2008). Are animated demonstrations the clearest and most comfortable way to communicate on-screen instructions? *Information Design Journal*, 16(2), 107–124. <https://doi.org/10.1075/IDJ.16.2.03BEZ>
- SPA, E. I., AE, E. T., ZOO, H. S. P., & KUENSTLICHE, D. F. F. (2014). *INTERACT–Interactive Manual Assembly Operations for the Human-Centered Workplaces of the Future*. 611007.
- Spanjers, I. A. E., Wouters, P., Van Gog, T., & Van Merriënboer, J. J. G. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior*, 27(1), 46–52. <https://doi.org/10.1016/j.chb.2010.05.011>
- Spannagel, C., Girwidz, R., Löthe, H., Zender, A., & Schroeder, U. (2008). Animated demonstrations and training wheels interfaces in a complex learning environment. *Interacting with Computers*, 20(1), 97–111. <https://doi.org/10.1016/J.INTCOM.2007.08.002>
- Steinberg, G. (2012). Natural User Interfaces. *Computer Science*. <https://www.semanticscholar.org/paper/Natural-User-Interfaces-Steinberg/a173b16449b44f8fa877fea686c54ed6cf09eb10>
- Stock, I., Weber, M., & Steinmeier, E. (2005). Metadata based Authoring for Technical Documentation. *Proceedings of the 23rd Annual International Conference on Design of Communication Documenting & Designing for Pervasive Information - SIGDOC '05*. <https://doi.org/10.1145/1085313>
- Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I on - AFIPS '68 (Fall, Part I)*, 33(pt 1), 757–764. <https://doi.org/10.1145/1476589>
- Swan, J. E., & Gabbard, J. L. (2005). Survey of User - Based Experimentation in Augmented Reality. *Proceedings of 1st International Conference on Virtual Reality*, 1–9.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). The Expertise Reversal Effect. *Cognitive Load Theory*, 155–170. [https://doi.org/10.1007/978-1-4419-8126-4\\_12](https://doi.org/10.1007/978-1-4419-8126-4_12)
- Syberfeldt, A., Danielsson, O., & Gustavsson, P. (2017). Augmented Reality Smart Glasses in the Smart Factory: Product Evaluation Guidelines and Review of Available Products. *IEEE Access*, 5(May), 9118–9130. <https://doi.org/10.1109/ACCESS.2017.2703952>
- Syberfeldt, A., Danielsson, O., Holm, M., & Wang, L. (2015). Visual Assembling Guidance Using Augmented Reality. *Procedia Manufacturing*, 1, 98–109. <https://doi.org/10.1016/j.promfg.2015.09.068>
- System Usability Scale (SUS) | Usability.gov. (n.d.). Retrieved March 22, 2022, from <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

- Tainaka, K. (2020). *Guideline and Tool for Designing an Assembly Task Support System Using Augmented Reality*. 654–665. <https://doi.org/10.1109/ISMAR50242.2020.00077>
- Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative Effectiveness of Augmented Reality in Object Assembly. *Proceedings of the Conference on Human Factors in Computing Systems - CHI '03*. <https://doi.org/10.1145/642611>
- Toro, C., Barandiaran, I., & Posada, J. (2015). A perspective on knowledge based and intelligent systems implementation in industrie 4.0. *Procedia Computer Science*, 60(1), 362–370. <https://doi.org/10.1016/j.procs.2015.08.143>
- Tsihrintzis, G. A., Damiani, E., Virvou, M., Howlett, R. J., & Jain, L. C. (2010). Intelligent interactive multimedia systems and services. *Smart Innovation, Systems and Technologies*, 6. <https://doi.org/10.1007/978-3-642-14619-0>
- Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters*, 36, 189–195. <https://doi.org/10.1016/j.patrec.2013.07.003>
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4), 247–262. <https://doi.org/10.1006/IJHC.2002.1017>
- Uchiyama, H., Marchand, E., Uchiyama, H., Marchand, E., Detection, O., Tracking, P., & Uchiyama, H. (2012). *Object Detection and Pose Tracking for Augmented Reality: Recent Approaches To cite this version: HAL Id : hal-00751704 Object Detection and Pose Tracking for Augmented Reality: Recent Approaches*.
- Unity 2019.4.10. (2019). <https://unity3d.com/unity/whats-new/2019.4.10>
- Urbas, U., Vrabič, R., & Vukašinović, N. (2019). Displaying Product Manufacturing Information in Augmented Reality for Inspection. *Procedia CIRP*, 81, 832–837. <https://doi.org/10.1016/J.PROCIR.2019.03.208>
- Uva, A. E., Gattullo, M., Manghisi, V. M., Spagnulo, D., Cascella, G. L., & Fiorentino, M. (2018). Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. *International Journal of Advanced Manufacturing Technology*, 94(1–4), 509–521. <https://doi.org/10.1007/s00170-017-0846-4>
- van Marrewijk, A. (2018). Digging for Change: Change and Resistance in Interorganizational Projects in the Utilities Sector. *Project Management Journal*, 49(3), 34–45. <https://doi.org/10.1177/8756972818770590>
- Vanneste, P., Huang, Y., Park, J. Y., Cornillie, F., Declodt, B., & Van den Noortgate, W. (2020). Cognitive support for assembly operations by means of augmented reality: an exploratory study. *International Journal of Human Computer Studies*, 143(October 2019), 102480. <https://doi.org/10.1016/j.ijhcs.2020.102480>
- Vuforia Expert Capture | PTC. (2022). <https://www.ptc.com/en/products/vuforia/vuforia-expert-capture>
- Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., & Schmalstieg, D. (2010). Real-time detection and tracking for augmented reality on mobile phones. *IEEE Transactions on Visualization and Computer Graphics*, 16(3), 355–368. <https://doi.org/10.1109/TVCG.2009.99>
- Wang, T., Qian, X., He, F., Hu, X., Huo, K., Cao, Y., & Ramani, K. (2020). CAPtURAR: An augmented reality tool for authoring human-involved context-aware applications. *UIST 2020 - Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 328–341. <https://doi.org/10.1145/3379337.3415815>
- Wang, X., Ong, S. K., & Nee, A. Y. C. (2015). Real-virtual interaction in AR assembly simulation based on component contact handling strategy. *Assembly Automation*, 35(4), 376–394. <https://doi.org/10.1108/AA-02-2015-012>
- Wang, X., Ong, S. K., & Nee, A. Y. C. (2016). Advanced Engineering Informatics Multi-modal augmented-reality assembly guidance based on bare-hand interface. *Advanced Engineering Informatics*, 30(3), 406–421. <https://doi.org/10.1016/j.aei.2016.05.004>

- Wang, Y., Zhang, S., Wan, B., He, W., & Bai, X. (2018). Point cloud and visual feature-based tracking method for an augmented reality-aided mechanical assembly system. *International Journal of Advanced Manufacturing Technology*, 99(9–12), 2341–2352. <https://doi.org/10.1007/s00170-018-2575-8>
- Warnecke, H. J., & Hüser, M. (1995). Lean production. *International Journal of Production Economics*, 41(1–3), 37–43. [https://doi.org/10.1016/0925-5273\(95\)00080-1](https://doi.org/10.1016/0925-5273(95)00080-1)
- Watson, G., Curran, R., Butterfield, J., & Craig, C. (2008). The Effect of Using Animated Work Instructions Over Text and Static Graphics When Performing a Small Scale Engineering Assembly. *Collaborative Product and Service Life Cycle Management for a Sustainable World - Proceedings of the 15th ISPE International Conference on Concurrent Engineering, CE 2008*, 541–550. [https://doi.org/10.1007/978-1-84800-972-1\\_51](https://doi.org/10.1007/978-1-84800-972-1_51)
- Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., & Preusche, C. (2013). An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems*, 61(4), 398–403. <https://doi.org/10.1016/j.robot.2012.09.013>
- WebXR. (2018). *WebXR Device API - Web APIs | MDN*. [https://developer.mozilla.org/en-US/docs/Web/API/WebXR\\_Device\\_API](https://developer.mozilla.org/en-US/docs/Web/API/WebXR_Device_API)
- Werrlich, S., Daniel, A., Ginger, A., Nguyen, P. A., & Notni, G. (2019). Comparing HMD-Based and Paper-Based Training. *Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2018*, 134–142. <https://doi.org/10.1109/ISMAR.2018.00046>
- Westerfield, G., Mitrovic, A., & Billinghamurst, M. (2015). Intelligent Augmented Reality Training for Motherboard Assembly. *International Journal of Artificial Intelligence in Education*, 25(1), 157–172. <https://doi.org/10.1007/s40593-014-0032-x>
- Weyer, S., Schmitt, M., Ohmer, M., Gorecky, D., Weyer, S., Schmitt, M., Ohmer, M., & Gorecky, D. (2015). ScienceDirect Standardization as the crucial challenge Towards Standardization as the crucial challenge for highly production systems for highly modular , multi-vendor production systems for highly modular , multi-vendor productio. *IFAC-PapersOnLine*, 48(3), 579–584. <https://doi.org/10.1016/j.ifacol.2015.06.143>
- Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). The development of the picture-superiority effect. *British Journal of Developmental Psychology*, 24(4), 767–773. <https://doi.org/10.1348/026151005X74153>
- Whitmire, E., Jain, M., Jain, D., Nelson, G., Karkar, R., Patel, S., & Goel, M. (2017). DigiTouch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3), 1–21. <https://doi.org/10.1145/3130978>
- Wiedenmaier, S., Oehme, O., Schmidt, L., & Luczak, H. (2009). Augmented Reality (AR) for Assembly Processes Design and Experimental Evaluation. *Http://Dx.Doi.Org/10.1207/S15327590IJHC1603\_7*, 16(3), 497–514. [https://doi.org/10.1207/S15327590IJHC1603\\_7](https://doi.org/10.1207/S15327590IJHC1603_7)
- Wittrick, M. C. (2010). Generative Processes of Comprehension. *Http://Dx.Doi.Org/10.1207/S15326985ep2404\_2*, 24(4), 345–376. [https://doi.org/10.1207/S15326985EP2404\\_2](https://doi.org/10.1207/S15326985EP2404_2)
- Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction*, 22(6), 449–457. <https://doi.org/10.1016/j.learninstruc.2012.05.004>
- Wouters, P., Tabbers, H. K., & Paas, F. (2007). Interactivity in video-based models. *Educational Psychology Review*, 19(3), 327–342. <https://doi.org/10.1007/s10648-007-9045-4>
- Wu, J., Zhou, B., Russell, R., Kee, V., Wagner, S., Hebert, M., Torralba, A., & Johnson, D. M. S. (2018). Real-Time Object Pose Estimation with Pose Interpreter Networks. *IEEE International Conference on Intelligent Robots and Systems*, 6798–6805. <https://doi.org/10.1109/IROS.2018.8593662>
- Wu, L. C., Lin, I. C., & Tsai, M. H. (2016). Augmented reality instruction for object assembly based on markerless tracking. *Proceedings - 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D 2016*, 95–102. <https://doi.org/10.1145/2856400.2856416>
- Wyman, O. (2017). *Surprise: Robots Aren't Replacing Humans In Key Areas Of Manufacturing | Forbes*.



- Yamaguchi, M., Mori, S., Mohr, P., Tatzgern, M., Stanescu, A., Saito, H., & Kalkofen, D. (2020). Video-annotated augmented reality assembly tutorials. *UIST 2020 - Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 1010–1022. <https://doi.org/10.1145/3379337.3415819>
- Yang, Y., Shim, J., Chae, S., & Han, T. D. (2016). Mobile Augmented Reality Authoring Tool. *Proceedings - 2016 IEEE 10th International Conference on Semantic Computing, ICSC 2016*, 358–361. <https://doi.org/10.1109/ICSC.2016.42>
- Yu, J., Jeon, J. U., Park, G., Kim, H. II, & Woo, W. (2016). A Unified Framework for Remote Collaboration Using Interactive AR Authoring and Hands Tracking. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9749, 132–141. [https://doi.org/10.1007/978-3-319-39862-4\\_13](https://doi.org/10.1007/978-3-319-39862-4_13)
- Zauner, J., Haller, M., Brandl, A., & Hartman, W. (2003). Authoring of a mixed reality assembly instructor for hierarchical structures. *Proceedings - 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2003*, 237–246. <https://doi.org/10.1109/ISMAR.2003.1240707>
- Zhao, J., Sensibaugh, T., Bodenheimer, B., McNamara, T. P., Nazareth, A., Newcombe, N., Minear, M., & Klippel, A. (2020). Desktop versus immersive virtual environments: effects on spatial learning. *Spatial Cognition and Computation*, 20(4), 328–363. <https://doi.org/10.1080/13875868.2020.1817925>
- Zhu, J., Ong, S. K., & Nee, A. Y. C. (2013). An authorable context-aware augmented reality system to assist the maintenance technicians. *International Journal of Advanced Manufacturing Technology*, 66(9–12), 1699–1714. <https://doi.org/10.1007/s00170-012-4451-2>
- Zhu, J., Ong, S. K., & Nee, A. Y. C. (2014). A context-aware augmented reality assisted maintenance system. <https://doi.org/10.1080/0951192X.2013.874589>, 28(2), 213–225. <https://doi.org/10.1080/0951192X.2013.874589>
- Zydney, J. M., & Warner, Z. (2016). Mobile apps for science learning: Review of research. *Computers & Education*, 94, 1–17. <https://doi.org/10.1016/J.COMPEDU.2015.11.001>

**Titre :** Méthodologies et outils pour le partage des connaissances spécialisées dans les industries d'assemblage manuel en utilisant la réalité augmentée.

**Mots clés :** réalité augmentée, formation, outil auteur, instructions de travail, assemblage manuel, industrie.

**Résumé :** Pour rester compétitifs dans la révolution industrielle en cours, c'est-à-dire l'industrie 4.0, les secteurs manufacturiers doivent assurer une grande flexibilité au niveau de la production, un besoin mieux satisfait par des travailleurs qualifiés. Alors que la formation traditionnelle devient de plus en plus inefficace, trouver une meilleure façon de former les travailleurs novices devient une exigence critique. La littérature suggère que la réalité augmentée (RA), une technologie émergente proposée par l'industrie 4.0, peut potentiellement répondre à cette question.

Les avantages des outils de partage des connaissances basés sur la RA ont été démontrés dans divers domaines, y compris l'industrie, de la fabrication à la validation et à la maintenance. Cependant, malgré les progrès de la RA ces dernières années, aucune percée industrielle significative n'est à noter. Nous avons constaté que la plupart des systèmes RA sont élaborés et évalués dans des environnements contrôlés, sans l'implication des éventuels utilisateurs finaux.

Guidés par les recommandations de la littérature, nous avons mené une étude de cas au long cours dans une usine d'assemblage manuel, afin d'identifier les besoins et les attentes auxquels un système de formation en RA devrait répondre, pour adresser de manière optimale le secteur industriel considéré.

De plus, nous avons mené une analyse approfondie de la représentation et de la transmission de l'information dans la RA, en ce qui concerne les implications cognitives et les efforts de création de contenu. Nous avons également exploré les paradigmes d'interaction homme-machine pour identifier les principes et les lignes directrices de conception pour l'élaboration d'un outil de RA dédié au contexte atelier. Nous avons constaté que la représentation visuelle de l'expertise d'assemblage en RA peut s'appuyer sur des éléments visuels peu coûteux (c'est-à-dire texte, photo, vidéo et données auxiliaires prédéfinies), tandis qu'une conception centrée sur l'humain devrait être adoptée lors de l'élaboration du système, en privilégiant l'utilisabilité et l'utilité plutôt que la performance.

Nous avons défini une représentation visuelle formalisée des opérations d'assemblage en RA, qui prend en compte les problématiques de création et soutient les performances de formation. Nous avons proposé une approche de création immersive qui permet de capturer son expertise d'assemblage *in-situ*, lors de l'assemblage. La création des instructions en RA est un processus en une étape, ne s'appuie pas sur des données existantes ou des services externes et ne nécessite pas d'expertise AR ou technique, de pré- ou post-traitement des données. Pendant la formation, les informations d'assemblage sont transmises en suivant le principe 2W1H, conçu pour guider les novices de manière naturelle et non intrusive, en minimisant les entrées de l'utilisateur et l'encombrement de l'interface utilisateur, et en visant à maximiser la compréhension et l'apprentissage.

Nous avons évalué notre proposition en menant plusieurs études. La première, menée sur un poste de travail d'assemblage réel, a confirmé l'hypothèse selon laquelle des éléments visuels à faible coût enregistrés dans l'espace peuvent transmettre efficacement l'expertise d'assemblage manuel aux travailleurs novices dans une configuration industrielle. Les résultats de la deuxième étude ont soutenu l'hypothèse selon laquelle la valeur de la création d'instructions AR basées sur la CAO dans un contexte industriel similaire est discutable.

Une dernière étude a prouvé que le système RA proposé, comprenant à la fois la procédure de création et de formation, peut être utilisé efficacement par des novices en quelques minutes. Dans l'ensemble, les résultats rapportés ont démontré la usabilité et l'efficacité de l'approche de formation en RA proposée, indiquant qu'une mise en œuvre de système similaire pourrait être adoptée avec succès dans les secteurs manufacturiers.

Les travaux futurs devraient valider les résultats expérimentaux rapportés dans des évaluations industrielles à grande échelle et proposer des modules "intelligents" (par exemple, validation d'assemblage) pour mieux aider les novices pendant leur formation et optimiser également la procédure de création.

**Title:** Methodologies and tools for expert knowledge sharing in manual assembly industries by using AR.

**Keywords:** augmented reality, training, authoring tool, work instructions, manual assembly, industry.

**Abstract:** To remain competitive in the ongoing industrial revolution (i.e., Industry 4.0), manufacturing sectors must ensure high flexibility at the production level, a need best addressed by skilled human workforce. As traditional training becomes increasingly inefficient, finding a better way of training novice workers becomes a critical requirement. Literature suggests that augmented reality (AR), an emerging technology proposed by Industry 4.0, can potentially address this concern.

The benefits of AR-based knowledge sharing tools have been demonstrated in a variety of domains, including industry, from manufacturing to validation and maintenance. However, despite the progress of AR in recent years, no significant industrial breakthrough can be noted. We found that most AR systems are elaborated and evaluated under controlled settings, without the implication of the eventual end users. Guided by literature recommendations, we conducted a long-term case study in a manual assembly factory, to identify needs and expectations that an AR training system should meet, to optimally address the considered industrial sector.

Further, we conducted an in-depth analysis on information representation and conveyance in AR, with respect to cognitive implications and content authoring efforts. We explored as well human-computer interaction paradigms to identify principles and design guidelines for elaborating an AR tool dedicated to the shop floor context. We found that the visual representation of the assembly expertise in AR can rely on spatially registered low-cost visual assets (i.e., text, photo, video, and predefined auxiliary data), while a human-centered design should be adopted during the elaboration of the AR system, prioritizing usability and usefulness rather than performance.

We defined a formalized visual representation (i.e., 2W1H principle) of assembly operations in AR, that considers authoring concerns and supports training performance. We proposed an HMD-only immersive authoring that allows one to capture his assembly expertise *in-situ*, during the assembly itself. The authoring is a one-step process, does not rely on existing data or external services and does not require AR or technical expertise, pre- or post-processing of data. During training, the assembly information is conveyed via AR by following the 2W1H principle, designed to guide novice workers in a natural, non-intrusive manner, minimizing user input and UI clutter, and aiming to optimize comprehension and learning.

We evaluated our proposal by conducting several experiments. The first, conducted on a real-world assembly workstation, confirmed the hypothesis that spatially registered low-cost visual assets can effectively convey manual assembly expertise to novice workers via AR in an industrial setup. The findings of the second experiment supported the assumption that the worthiness of authoring CAD-based AR instructions in similar industrial context is questionable. A final experiment proved that the proposed AR system, including both authoring and training procedures, can be used effectively by novices in a matter of minutes. The overall reported feedback demonstrated the usability and efficiency of the proposed AR training approach, indicating that a similar system implementation could be successfully adopted in shop floor environments.

Future work should validate the reported experimental findings in large-scale industrial evaluations and propose reliable “intelligent” modules (e.g., assembly validation and feedback) to better assist novice workers during training and optimize the authoring procedure as well.