Asynchronous Collaboration in Physical Engineering Environments using Immersive Technologies

Zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN
(Dr.-Ing.)

von der KIT-Fakultät für
Maschinenbau
des Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

M.Sc. Anjela Mayer

Tag der mündlichen Prüfung:	21.11.2025
Hauptreferent:	Prof. Dr. Dr.-Ing. Dr. h. c. Jivka Ovtcharova
Korreferent:	Prof. habil. Dr.-Ing. Jean-Rémy Chardonnet

"Time is not a line but a dimension, like the dimensions of space. If you can bend space you can bend time also, and if you knew enough and could move faster than light you could travel backwards in time and exist in two places at once."

– Margaret Atwood

Abstract

Collaborative engineering tasks are not only cognitively demanding, but often involve physical, spatial, and technical complexity. While established computer-supported cooperative work (CSCW) technologies provide effective support for synchronous collaboration, asynchronous collaboration – especially on spatially procedural tasks – remains unaddressed; this gap is particularly evident in industrial domains where task continuity, embodied interaction, and spatial context are critical.

To address this gap, this work applies the Design Science Research (DSR) methodology to develop and evaluate a framework for asynchronous collaboration in Extended Reality (XR). XR technologies enable the spatial integration of digital content with the physical environment, allowing users to interact immersively with task-relevant data and retain procedural context over time.

Grounded in theoretical foundations and related work, a conceptual model is proposed to address the three dimensions – communication, coordination, and cooperation – to support asynchronous collaboration. These components were implemented and evaluated individually in the design cycle.

In the rigor cycle, the modular components were consolidated into a holistic XR system and applied in a robotic manufacturing scenario at the Karlsruhe Research Factory. The integrated solution was evaluated through a comparative user study against a conventional video-based workflow. Results show improvements in spatial understanding, procedural clarity, co-presence and user engagement under the XR condition, although limitations remain in task efficiency, familiarity with XR, and hardware constraints such as image resolution.

This research contributes (i) a theory-informed conceptual model that operationalizes communication, coordination, and cooperation for asynchronous XR, along with design considerations for spatio-temporal work; (ii) a modular system architecture and a replicable evaluation protocol; and (iii) empirical evidence from a comparative study in an industrial setting. Together, these contributions address a critical gap in CSCW tools for engineering and provide practical guidance for integrating asynchronous spatial collaboration across co-located and remote settings.

Kurzfassung (German)

Kollaborative Aufgaben im Ingenieurwesen sind nicht nur kognitiv anspruchsvoll, sondern häufig physisch, räumlich und technisch komplex. Während konventionelle Technologien der Computerunterstützten Kooperativen Arbeit (CSCW) eine effektive synchrone Zusammenarbeit ermöglichen, bleibt asynchrone Kollaboration – insbesondere bei räumlich-prozeduralen Tätigkeiten – weitgehend unberücksichtigt. Diese Lücke zeigt sich besonders in industriellen Anwendungsfeldern, in denen Aufgabenkontinuität, verkörperte Interaktion und räumlicher Kontext entscheidend sind.

In dieser Arbeit wurde die Design-Science-Research-Methodik angewendet, um ein Rahmenwerk für asynchrone Zusammenarbeit in Extended Reality (XR) zu entwickeln und zu evaluieren. XR-Technologien ermöglichen die nahtlose räumliche Integration digitaler Inhalte in physische Umgebungen, unterstützen verkörperte Interaktion mit aufgabenrelevanten Daten und sichern die zeitliche Persistenz prozeduraler Kontexte.

Auf Basis theoretischer Grundlagen und verwandter Arbeiten wurde ein konzeptuelles Modell formuliert, das asynchrone Kommunikation, Koordination und Kooperation (3C) im XR-Kontext präzisiert und operationalisiert. Die Komponenten wurden im Design-Zyklus als modulare Prototypen implementiert und jeweils einzeln evaluiert. Im Rigor-Zyklus wurden die Module zu einem ganzheitlichen XR-System zusammengeführt und in einem robotergestützten Fertigungsszenario in der Karlsruher Forschungsfabrik angewendet. Die integrierte Lösung wurde in einer vergleichenden Nutzerstudie gegenüber einem konventionellen videobasierten Arbeitsablauf evaluiert. Die Ergebnisse zeigen Verbesserungen in räumlichem Verständnis, prozeduraler Klarheit, Co-Präsenz

und Nutzerengagement mit dem XR-Prototyp, bei gleichzeitigen Einschränkungen in Arbeitseffizienz, Technologievertrautheit sowie hardwarebedingten Limitierungen (z. B. Bildauflösung).

Der Beitrag der Arbeit liegt in: (i) einer konzeptionellen Fundierung asynchroner XR-Zusammenarbeit entlang der 3C einschließlich Begriffspräzisierung und Gestaltungsprinzipien, (ii) dem ACIE-Rahmenwerk mit Referenzarchitektur für räumlich-prozedurale Asynchronität sowie (iii) der empirischen Prüfung durch modulare und integrierte Evaluation in einem realitätsnahen Industrieszenario. Damit wird eine zentrale Lücke in der aktuellen CSCW-Forschung und in bestehenden Ingenieurwerkzeugen adressiert und die Grundlage für flexiblere Formen kollaborativer Arbeit gestärkt – als Basis für die umfassende Unterstützung und sukzessive Integration aller Kooperationsformen: co-lokal und remote, synchron und asynchron.

Acknowledgements

I was supervised by Prof. Dr. Dr.-Ing. Dr. h. c. Jivka Ovtcharova, my first supervisor, who founded the Institute for Information Management in Engineering (IMI) at the Karlsruhe Institute of Technology (KIT). She created a supportive environment for thriving, where I enjoyed studying, later working, and in the end also researching, teaching, and completing my PhD. She inspired me to do research, encouraged me to grow, give my best, and to expand my horizons – cognitively and geographically.

I also thank Prof. habil. Dr.-Ing. Jean-Rémy Chardonnet, my second supervisor, for support from the very beginning of my PhD. He introduced me to international research and invited me to France for my first research stay abroad. Throughout the PhD, he offered sustained scientific guidance, navigated me through the publication landscape, and helped me to develop a sense for academic best practices.

Two special collaborations became friendships beyond space and project timelines: with Inés Miguel Alonso, my "research tandem" from my first conference onward, and with Kevin Kastner from my first XR project onward. I always enjoy exchanging ideas and bringing them to life together. With them, the PhD felt more like an adventure shared than a solitary path.

At KIT and IMI, I was surrounded by inspiring, professional, and kind colleagues – whose absence was keenly felt during the lockdowns. Although our topics seem different, we collaborated closely and exchanged a lot. I thank Jakob Bönsch for introducing me to the VSK activities at Karlsruher Forschungsfabrik, which later became part of an international project between KIT and ENSAM and included an ideal evaluation use case for my research. Thanks also to colleagues at wbk, Edgar Mühlbeier, Simon Otto, and their students for the on-boarding and technical assistance throughout my experiments in the production cell. For reliable administrative support at IMI, I want to thank Karin Jenewein, Sonja Ohler, and Thomas Maier.

Furthermore, I am grateful for the opportunities I had for several research stays at institut Arts et Métiers de Chalon-sur-Saône in France, Vicomtech in Spain, and at Trakia University Stara Sagora in Bulgaria. I had opportunities to connect with peers in my research domain, to discuss XR capabilities in education and professional work, and to learn about their projects and research landscapes. These experiences are very valuable for me.

My sincere thanks go to all study participants and assistants, as well as to the students I had the pleasure to supervise; their perspectives were very valuable for my research and the broader view. I am grateful for partner organizations – Gabler Engineering and the ATS Lab at TH Mannheim – for kindly providing data for the experiments.

I also want to thank the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the InnovationCampus Future Mobility (ICM) for funding the KoLabBW and XIRCON projects, as well as awarding me an ICM mobility grant. These projects helped to identify realistic validation use cases.

Finally, I thank my family and friends – especially my parents, Natalia and Michael, and my husband, Sören, for encouraging curiosity, supporting my path, and consistently offering in-depth perspective and encouragement when doubt arose. Their support is consistent throughout good and difficult times.

Table of Contents

Abstract . . . iii

Kurzfassung (German) . . . v

Acknowledgements . . . vii

List of Figures . . . xiii

List of Tables . . . xix

Abbreviations and Symbols . . . xxiii

1 Introduction . . . **1**
- 1.1 Motivation . . . 1
- 1.2 Problem Definition . . . 3
- 1.3 Research Objective . . . 6
- 1.4 Research Methodology . . . 10
  - 1.4.1 Design Science Research Framework . . . 11
- 1.5 Research Contributions . . . 13
- 1.6 Thesis Structure . . . 14
- 1.7 Summary and Outlook . . . 15

2 Theoretical Foundation . . . **17**
- 2.1 Computer-Supported Cooperative Work . . . 18
  - 2.1.1 Immersive Asynchronous Collaboration . . . 20
  - 2.1.2 The 3C Model . . . 22
- 2.2 Extended Reality . . . 26
  - 2.2.1 Key Concepts for XR Collaboration . . . 28

2.2.2 XR Technology Considerations 31
2.2.3 Technical Suitability for Engineering Use Cases 35
2.2.4 Conceptual Challenges and Affordances for Asynchronous XR Collaboration 36
2.2.5 Technical Constraints and System Capabilities for Asynchronous XR 37
2.2.6 Evaluation Construct 38
2.3 Summary and Outlook 42

3 State of the Art 45
3.1 Findings from Prior Reviews 45
3.2 Structured Literature Review 49
3.3 SLR Synthesis 55
3.3.1 Embodiment 56
3.3.2 Task Complexity 56
3.3.3 Communication 57
3.3.4 Coordination 58
3.3.5 Cooperation 59
3.3.6 XR Continuum Analysis 61
3.3.7 System Analysis 62
3.4 Research Gaps 63
3.5 Summary and Outlook 66

4 Conceptual Design for Asynchronous Collaboration in XR 69
4.1 ACIE Framework 70
4.2 Asynchronous Communication 73
4.2.1 Immersive Messages 74
4.2.2 Immersive Communication Protocol 77
4.3 Asynchronous Cooperation 79
4.3.1 Role Management 79
4.3.2 Message Categorization 80
4.3.3 Interactive Proxies 81
4.3.4 Asynchronous Interaction 82

4.4 Asynchronous Coordination 84
4.4.1 Message Coordination 84
4.4.2 Work Coordination 86
4.5 Asymmetric Collaboration 89
4.6 Summary and Outlook 90

5 Implementation 93
5.1 System Overview 93
5.2 Asynchronous Communication 96
5.2.1 XR Messages 96
5.2.2 XR Messaging Protocol 101
5.2.3 Evaluation of Asynchronous Communication 102
5.3 Asynchronous Cooperation 107
5.3.1 Proxy Avatars 107
5.3.2 Role-Aware Proxy Interaction 111
5.3.3 Evaluation of Proxy-Based Asynchronous Cooperation . . 112
5.4 Asynchronous Coordination 115
5.4.1 Message Navigation and Referencing 115
5.4.2 Immersive Version Control System (iVCS) 118
5.4.3 Evaluation of Asynchronous Coordination 120
5.5 Cross-Reality Collaboration 122
5.5.1 Scene Understanding and Spatial Anchoring 122
5.5.2 Performance Limitations and Optimization 124
5.5.3 Evaluation of Mixed Reality Modalities 125
5.6 Summary and Outlook 127

6 Evaluation 131
6.1 Practical Field Use Case 132
6.1.1 Commissioning and Safety Procedure 134
6.1.2 Challenges in the Existing Workflow 136
6.1.3 Integrating Asynchronous Collaboration 137
6.2 Study Design 140
6.2.1 Experimental Conditions 140
6.2.2 Participants 141
6.2.3 Experimental Procedure 143

6.2.4 Task and System Setup . . . 146
6.2.5 Ethical Considerations . . . 147
6.2.6 Measurements . . . 147
6.3 Results . . . 152
6.3.1 Usability and Workload Results . . . 152
6.3.2 Collaboration Results . . . 160
6.3.3 Spatial and Social Presence Results . . . 162
6.3.4 Efficiency Results . . . 169
6.3.5 Accuracy Results . . . 173
6.3.6 Qualitative Results . . . 176
6.4 Major Findings and Implications . . . 179
6.5 Summary and Outlook . . . 184

7 Conclusion . . . 187
7.1 Discussion . . . 189
7.2 Contributions . . . 192
7.3 Limitations . . . 194
7.4 Outlook . . . 198
7.5 Closing Remarks . . . 202

A Appendix . . . 203
A.1 Search Keywords and Boolean Strings . . . 203
A.2 Scoring of XR AC Systems . . . 204
A.3 Avatar Bone and Expression Mappings . . . 205
A.4 Evaluation Appendix . . . 211

Own Publications . . . 217
Journal Articles . . . 217
Book Chapters . . . 217
Conference Papers . . . 218
Source Code . . . 221

References . . . 223

List of Figures

1.1 Overview of the identified research gaps, objective, and research questions . . . 9
1.2 Adapted Design Science Research framework illustrating the relevance, rigor, and design cycles guiding this thesis, based on [72] 10
2.1 Extended CSCW design space with the additional XR dimension distinguishing symmetric (same reality) and asymmetric (mixed realities) collaboration. Adapted from Johansen's time–space matrix [82]. . . . 18
2.2 The 3C model of collaborative work, illustrating communication, coordination, and cooperation as foundational dimensions of collaboration . . . 23
2.3 The Reality–Virtuality Continuum, adapted from Milgram et al. [113], with example imagery for Augmented Reality (AR), Augmented Virtuality (AV), and Virtual Reality (VR), and Extended Reality (XR) as the umbrella term. . . . 27
2.4 Presence and its subdimensions in XR collaboration with associated research questions . . . 30
2.5 XR devices: (a) haptic vest with a VR HMD, (b) exoskeleton glove for force feedback, (c) Disney's HoloTile omnidirectional floor, and (d) olfactory headset . . . 32
2.6 Decomposition of usability according to ISO 9241-11 . . . 40
3.1 Literature-search workflow: from conceptual grounding to systematic search, opportunistic expansion, and unified synthesis. Stages are detailed in ❯ Section 3.1, ❯ Section 3.2, and ❯ Section 3.3. . . . 46

3.2 PRISMA-based flow diagram of the structured literature review process for identifying relevant articles on asynchronous collaboration in XR [122] 50
4.1 Conceptual model of asynchronous collaboration structured into communication, cooperation, and coordination mechanisms, with related research questions and corresponding sections. . 71
4.2 Decomposition of **RQ1** into sub-questions (communication, cooperation, coordination). 72
4.3 Workflow of an Immersive Message Exchange in ACIE: (1) Avatar Recording captures a collaborator's spatial motion and voice; (2) the data are serialized into an Immersive Message file and exchanged between the collaborators; (3) on reception, Proxy Playback re-embodies the sender in the recipient's scene, enabling spatially aligned, temporarily shifted collaboration. 74
4.4 Record data structure (ICP implementation). 77
4.5 Enhanced asynchronous sequence diagram with explicit message types and report interaction step. 83
4.6 Asymmetric collaboration across XR modalities. 90
5.1 System architecture overview of the ACIE framework, illustrating key modules for XR interaction, asynchronous messaging, and file-based coordination. Arrows indicate data and control flow between runtime and support components. 94
5.2 Detailed overview of the ACIE components involved in avatar capture and proxy playback 97
5.3 Comparison of skeletal mesh structures of MetaXR and UE avatars, used for retargeting in avatar motion replay. Detailed joint structures are provided in Appendix A, Figures A.1 and A.2 . 98
5.4 Sequence diagram of the avatar recording process 99
5.5 Sequence diagram of the proxy avatar playback process 101
5.6 Data structures used in the immersive message protocol 103
5.7 Decomposition of **RQ1.1** with corresponding hypothesis and measurements. 104
5.8 NASA-TLX workload ratings from the SalentoXR study, showing reduced cognitive demand on most subscales for the ghost avatar condition 105

5.9 Examples of proxy avatars implemented in the framework. From left to right: a half-body proxy (head and hands) in see-through AR; a full-body proxy avatar in passthrough MR; and a Metahuman proxy avatar in passthrough MR. 108
5.10 Creation of a personalized Metahuman avatar using photogrammetry 110
5.11 Ghost avatars and object proxies for non-intrusive replay 110
5.12 Role-specific proxy menus for asynchronous interaction. The novice records and previews a message on the expert's proxy (e.g. a question) and later plays back the expert's response. The expert reviews the original issue and records an answer on the novice's proxy. 111
5.13 Decomposition of **RQ1.2** with corresponding hypothesis and measurements. 112
5.14 Industrial task guidance using three different asynchronous instruction modalities: (1) 2D video on smartphone (top right), (2) AR video embedded in physical space via HoloLens (top left and bottom left), and (3) immersive avatar-based playback in VR using Meta Quest Pro (top center and bottom right) 113
5.15 Decomposition of **RQ1.3** with corresponding hypothesis and measurements. 116
5.16 Directory structure of asynchronous collaboration session recordings.117
5.17 Hand menu variations depending on user role (novice or expert) and the current state of the replay system (e.g. recording, playing, stopped) 118
5.18 Immersive Version Control System (iVCS): Users can spatially explore and manage scene states through an interactive version tree (b), anchored in a collaborative virtual environment (a) 120
5.19 iVCS evaluation task: Tasks T1–T12 were performed by each participant using the Oculus Quest 1 HMD 121
5.20 Cross-reality assembly application, allowing the recording of assembly instruction in VR and playback in AR and vice versa . . . 126

5.21 Performance profile of the ACIE system during mixed reality (MR) operation on a Meta Quest Pro. The top curve (in red) shows the frame rate, fluctuating between 41–65 FPS. The middle group of curves depicts CPU thread utilization across both system and application cores. The bottom curves represent GPU activity and scheduling performance. 127
6.1 The Value Stream Kinematics (VSK) evaluation environment – a flexible production system at the Karlsruhe Research Factory . . 132
6.2 Examples of commissioning steps performed before operating the VSK system 134
6.3 Top-down CAD view of the VSK cell 138
6.4 Hierarchical overview for **RQ2** (Usability and Presence): research questions, hypotheses, and associated measurements. 139
6.5 Study task showing three instruction steps (top) presented by the MR avatar and participants (bottom) following the avatar's guidance 141
6.6 Participant following asynchronous task instructions in the 2D Video condition: using a handheld smartphone (left) and the video instruction showing the current step (right). 142
6.7 Participant demographic distribution by gender and condition . . . 142
6.8 Technology familiarity, experience, and comfort level 143
6.9 Training scenario used to familiarize participants with the experiment conditions and asynchronous workflow 144
6.10 Overview of the experimental procedure, including the sequence of tasks (T0–T3) and questionnaires (Q1–Q4). Left brace labels include per-phase completion times ($M \pm SD$, minutes) for *MR Avatar* and *2D Video*; training was constrained to 5–10 minutes. 145
6.11 Experiment task: participants conducted steps 1-8 to start up the production system, whereby tasks 3, 4, 5, 7, and 8 had intended errors to trigger asynchronous assistance using the Q&A function. 146
6.12 Average ASQ scores for MR Avatar and 2D Video conditions . . . 153
6.13 System Usability Scale (SUS) scores comparing MR Avatar and 2D Video conditions 154

6.14 NASA-TLX results across all six workload subscales: Mental Demand, Physical Demand, Temporal Demand, Effort, Performance and Frustration. 156
6.15 User Experience ratings for MR Avatar and 2D Video across dimensions 157
6.16 Detailed UEQ subscale comparisons between MR Avatar and 2D Video conditions 159
6.17 Collaboration ratings across three subcomponents: Communication, Cooperation, and Coordination 161
6.18 Comparison of IPQ subscales (Involvement, Realism, Spatial Presence) between MR Avatar and 2D Video conditions 163
6.19 Distribution of individual IPQ scores across subscales and conditions165
6.20 NMSPI results by subscale scores showing social presence perceptions in both conditions 166
6.21 Boxplot visualization of completion times across task phases for MR Avatar and 2D Video conditions 170
6.22 Boxplot visualization of accuracy across task phases for MR Avatar and 2D Video conditions 174
7.1 Application of immersive avatar-based instruction for pump-station procedures at the ATS Institute at TH Mannheim 197
7.2 Overview of the Sustainable Development Goals (SDGs), including SDG 4, SDG 9, and SDG 12 to which this work contributes 198
7.3 Volumetric human motion capture: compared to conventional avatars, volumetric recordings provide a holistic, trackerless, and highly accurate representation of both movement and appearance.201
A.1 MetaXR skeletal mesh structure in the Unreal Engine Editor view . 205
A.2 UE Quinn skeletal mesh structure in the Unreal Engine Editor view 206
A.3 Hand bones and names reference illustration 209
A.4 Decision logic for selecting appropriate statistical tests based on data distribution and variance assumptions 212

List of Tables

1.1 Asynchronous collaboration on spatial/procedural tasks: capability comparison of immersive and "flat" representation media 5
2.1 Key challenges of asynchronous collaboration in XR environments 20
2.2 XR technology requirements for asynchronous collaboration in practical engineering contexts 35
3.1 Overview of review papers on XR collaboration, including the number of articles addressing asynchronous collaboration (AC) . . 47
3.2 Existing technical approaches supporting asynchronous collaboration in XR (adapted from Letter et al. [97]) 48
3.3 Number of papers retrieved per database using the defined search expression 51
3.4 Categorization of screened articles by origin (review articles, structured literature review, and snowballing), grouped into excluded, peripheral, related, and core categories 52
3.5 Conceptual mapping of representative XR approaches for Asynchronous Collaboration on Complex Tasks (ACCT) across five dimensions: embodiment, task complexity, communication, coordination, and cooperation 55
4.1 Key components of a frame in the Immersive Communication Protocol. 78
4.2 Role-Based permissions in asynchronous immersive cooperation. . 80
4.3 Message types in asynchronous immersive cooperation. 81
5.1 Usability (SUS) of Avatar Capture and Playback across evaluated implementations. 106
5.2 Usability (SUS) across modalities. Means are reported as $M \pm SD$. 114

5.3 Validation of XR-related hypotheses in the Design Cycle based on implemented system modules and formative results 129
6.1 Analyzed sample size and gender distribution by condition 142
6.2 Scoring system used to assess task accuracy 148
6.3 Overview of questionnaires used to evaluate system usability, user experience, presence, collaboration, and participant demographics 150
6.4 Summary of NASA-TLX workload ratings comparing MR Avatar and 2D Video conditions 156
6.5 Statistical comparison of UEQ dimension ratings between MR Avatar and 2D Video conditions 158
6.6 Collaboration results summary comparing MR Avatar and 2D Video conditions. 161
6.7 IPQ results summary comparing MR Avatar and 2D Video conditions.164
6.8 Statistical comparison of NMSPI subscale scores between MR Avatar and 2D Video conditions 167
6.9 Mean task completion times (in minutes) and standard deviations across phases for MR Avatar and 2D Video conditions 170
6.10 Task accuracy comparison across phases between MR Avatar and 2D Video conditions 173
6.11 Summary of core findings across presence, collaboration, performance, and experience dimensions 182
6.12 Empirical validation of XR-related hypotheses in the Rigor Cycle, based on observed efficiency, accuracy, presence, and satisfaction outcomes 183
7.1 Comparison of core XR systems for asynchronous collaboration using a five-dimension scoring framework, including the proposed ACIE system 193
A.1 This scoring was used for the abstract comparison of XR asynchronous collaboration systems in ❯ **Chapter 3** 204
A.2 Skeletal joint mapping between MetaXR Avatar, UE Metahuman, and UE Mannequin. Body Joint Visual Reference for Movement SDK for Unreal: `https://developers.meta.com/horizon/documentation/unreal/move-ref-body-joints` 207

A.3 Skeletal mapping between UE5 hands and Meta Quest Pro hand tracking. Body Joint Visual Reference for Movement SDK for Unreal: https://developers.meta.com/horizon/documentation/unreal/move-ref-body-joints 208
A.4 Facial expression mapping between Metahuman expressions and MetaXR blendshapes. MetaXR Blendshape Visual Reference—Movement SDK for Unreal: https://developers.meta.com/horizon/documentation/unreal/move-ref-blendshapes 210
A.5 Overview of statistical tests used for data analysis, including checks for normality, variance, and appropriate group comparison methods 211
A.6 Demographics and Preferences Questionnaire 213
A.7 Custom Collaboration Questionnaire Items 214
A.8 Comprehensive results summary of MR Avatar and 2D Video conditions 215

Abbreviations and Symbols

Abbreviations

IMI	Institute for Information Management in Engineering
KIT	Karlsruhe Institute of Technology
DSR	Design Science Research methodology
ACIE	Asynchronous Collaboration in Immersive Environments (proposed framework)
AC	Asynchronous Collaboration
ACCT	Asynchronous Collaboration on Complex Tasks
XR	Extended Reality
VR	Virtual Reality
AR	Augmented Reality
AV	Augmented Virtuality
MR	Mixed Reality
UE	Unreal Engine
IM	Immersive Message
IVCS	Immersive Version Control System
SCT	Spatially Complex Tasks

CT Complex Tasks

CSCW Computer-Supported Cooperative Work

HCI Human-Computer Interaction

SUS System Usability Scale

NASA-TLX
NASA Task Load Index

DoF Degree of Freedom

CAD Computer-Aided Design

HMD Head-Mounted Display

FoV Field of View

FPS Frames Per Second

Latin Symbols and Variables

M Median

SD Standard Deviation

p Probability value for hypothesis testing

Δt Time interval between two frames or events

x, y, z 3D spatial coordinates in virtual environments

t Time

r Correlation coefficient

U Mann–Whitney U test statistic

1 Introduction

In this chapter, the motivation, problem space and research direction of the thesis are outlined. Current challenges in engineering collaboration are introduced, and the need for asynchronous support in spatial and procedural tasks is emphasized. In Section 1.1, the context and rationale for this research are presented. Section 1.2 defines the core issues related to spatial and embodied collaboration, highlights limitations of existing tools, and introduces the concept of complex tasks. The overarching research objective and corresponding research questions are formulated in Section 1.3. The Design Science Research (DSR) methodology guiding this work is described in Section 1.4. Lastly, an overview of the thesis structure is provided in Section 1.6, outlining how each chapter contributes to answering the research questions.

1.1 Motivation

Engineering work has always been inherently collaborative, involving actors with different roles, skills and responsibilities who jointly contribute to the development, optimization and execution of complex systems [38, 69, 168]. In modern engineering domains, this collaboration increasingly unfolds across spatial, temporal and organizational boundaries [1, 60, 119]. Digitalization, globally distributed projects and more flexible forms of work have led to highly dynamic, multi-modal and location-independent collaboration settings [27, 74, 121]. At the same time, the growing complexity and specialization of tasks make it increasingly difficult for all contributors to be present at the same time and place – whether virtually or physically [79, 109, 143].

While conventional Computer Supported Cooperative Work (CSCW) technologies such as video conferencing tools (e.g. Zoom, MS Teams), collaborative documents (e.g. Google Docs, Miro) and documentation systems (e.g. Confluence, Git) provide effective support for text- and screen-based collaboration, they fall short in supporting asynchronous collaboration that involves spatial, procedural or embodied interactions [36, 38, 48, 55]. They are well suited for co-authoring documents, conducting meetings or coordinating tasks via messaging or static annotations, but insufficient for communicating spatially situated procedural knowledge or intent within physical environments across time [26, 97, 162].

This limitation becomes particularly evident in spatially anchored and cyber-physical workflows – such as those found in flexible production environments or configuration of technical systems [43, 77, 139]. These engineering tasks demand interaction that is not only cognitive but also spatially precise, ordered and embodied [162]. They typically involve room-scale environments, 3D information, physical tooling and interdependent manual operations [2, 43]. Supporting such multidimensional and multi-modal interaction – spatial, procedural and physical – requires more than screen-based or textual communication [26, 162].

Although efforts have been made to support asynchronous collaboration using recordings, screenshots and comments, these media fail to capture the dynamic, contextual and procedural nature of spatial engineering tasks [38, 48, 162]. Version control systems and document annotations provide value in software-centric workflows but do not translate well to physically grounded, spatially complex activities [50, 172].

Recent advances in Extended Reality (XR), particularly in Augmented Reality (AR) and Mixed Reality (MR), offer promising capabilities to bridge this gap. XR technologies enable the spatial integration of digital information into the physical environment, supporting intuitive visualization, annotation and manipulation of 3D content in situ [1, 97, 162]. Beyond spatial immersion, XR also opens new dimensions in the time continuum: actions can be captured, persisted and replayed, allowing asynchronous users to interact with past activity traces in a shared spatial context [38, 48]. This dual capability – situating information in both space and time – positions XR as a powerful medium for asynchronous collaboration on spatially procedural engineering tasks [38, 48, 97].

However, despite this potential, XR systems today are mainly used for training scenarios or synchronous remote assistance. Their application for enabling asynchronous collaboration – particularly in engineering contexts that involve procedural workflows and physical interaction – remains underexplored [38, 97, 139].

This thesis addresses the gap by investigating how asynchronous collaboration in XR can be systematically designed and evaluated to support engineering workflows that depend on spatial understanding, procedural execution and embodied interaction.

1.2 Problem Definition

Collaboration in engineering involves not only distributed actors, but also complex tasks that require spatial, procedural, and physical interaction (➔ Section 1.1). While there are tools for synchronous collaboration (e.g. calls, video meetings) and basic asynchronous collaboration (e.g. emails, shared documents), most platforms lack seamless transitions across the collaboration matrix [48, 162]. Although XR offers suitable affordances, current practice remains oriented toward training and synchronous remote assistance, leaving generalizable methods and evaluation protocols for asynchronous, spatially procedural collaboration limited. Reviews highlight that asynchronous collaboration – despite its benefits such as flexibility and time-zone independence – remains significantly underexplored [38, 139].

Considerations – such as why asynchronous support is necessary for physical engineering tasks, whether existing solutions are sufficient, and which task properties make asynchronous support difficult – motivate a research need to specify and evaluate representations and interaction methods suited to asynchronous, spatially procedural collaboration in engineering.

Spatial and Physical Interaction Challenges

Existing CSCW tools fall short in supporting spatial and physical interaction, particularly for asynchronous collaboration. Engineering tasks often involve:

- 3D data and information that must be visualized, manipulated or interpreted [2, 16, 38, 162],
- room-scale or spatial interactions requiring an understanding of physical environments and actions [2, 20, 35, 38, 43],
- mapping digital information into real spaces and vice versa – bridging physical and digital worlds [2, 16, 43, 162].

A comparison of immersive and non-immersive CSCW representations for Asynchronous Collaboration (AC) on spatial/procedural tasks is provided in Table 1.1.

> ⚠ **P1 Representation Gap:** Existing asynchronous collaboration tools flatten spatial tasks into 2D, which impairs understanding.

Existing asynchronous collaboration tools are limited in their ability to represent spatial or physical tasks [77, 156]. They often reduce 3D data to 2D on-screen representations, which can be difficult to interpret and interact with [56, 105, 154]. For the engineering perspective, digital twins of machines and automated processes are becoming common – yet the human component is often left out Mayer et al. [9]. There are no sufficient methods to capture and document human tasks or embodied actions within these processes [36, 153]. In many engineering contexts it would be valuable to review previous actions

Table 1.1: Asynchronous collaboration on spatial/procedural tasks: capability comparison of immersive and "flat" representation media

Information Characteristic	**Immersive (XR)**	**Non-immersive (video/documents)**
In-situ spatial anchoring: information appears at its physical location	Provided via MR overlays & spatial anchors; depends on tracking and calibration	Not provided; depends on user's mental spatial mapping
Embodied demonstration: actions supported by hands/pose/gaze information	3D demonstration; avatar/volumetric/hand replays in context support imitation	"flat" demonstration; via video/images; embodiment remains indirect
Temporal navigation: playback control & history of actions	In-context temporal navigation; immersive 3D review of actions; requires new interaction patterns	On-screen temporal navigation; "flattened" review of actions; familiar temporal navigation patterns
Spatial navigation: spatial information distribution	possible; information is embedded all around the space	limited; 3D information gets mapped to 2D media
Persistence & history: trace of actions/states	embedded; traceable within the work environment	abstracted; traceable in repository environment
Safety: Situational awareness during physical work	Embedded information; Eyes-on-task & information simultaneously; FOV/occlusion risks must be addressed	Split information; attention shifts between environment and information; situational awareness may be reduced
Accessibility: deployment at scale	Device cost/fit/training required; benefits grow with routine use	Ubiquitous devices; low entry barrier and broad reach

for error checking, training or collaborative understanding – but existing tools offer limited support [38, 77].

> ⚠ **P2 Continuity Gap:** Lack of seamless integration between digital and real-world contexts disrupts workflow and comprehension, increasing cognitive workload.

Current approaches typically require users to look away from the real machine or environment – for example, checking a 2D screen or paper-based instructions – instead of seeing digital guidance overlaid directly onto the physical workspace [105, 154, 156]. Traditional collaboration tools rely primarily on touch, or keyboard and mouse interactions; they are less intuitive than embodied interaction when working with spatial or physical content [16, 36]. As engineering processes become increasingly digital, there is a growing need for intuitive methods that allow human operators to engage with digital information in situ [77, 162]. This represents a missed opportunity for immersive approaches, where digital and physical information could be integrated to improve presence, reduce cognitive load and support more natural user interaction [35, 38, 43].

Additionally, related technical limitations include:

- **Limited Support for Version Control in Immersive Environments:** While version control is a well-established concept in software development, it has not been widely adopted in XR environments [50, 172]. This complicates tracking and managing changes over time.
- **Absence of a Standard Exchange Format for Collaborative Data:** There is no common format for recording, storing and sharing user interaction data in XR environments – hindering collaboration [36, 172].

1.3 Research Objective

This thesis aims to facilitate asynchronous collaboration on complex engineering tasks using XR technologies. Immersive systems are well suited for spatial collaboration and offer key advantages over flat media – in particular for conveying spatially contextualized and embodied information. To address the gaps identified above, the work proposes a modular framework and supporting system for asynchronous, embodied and spatial collaboration.

◎ **Research Objective:** Propose an XR-based approach to facilitate ACCT through improved representation and communication of spatial context.

In this research, Complex Tasks (CT)[1] are defined as tasks that require multiple cognitive and physical efforts, often in dynamic or technical environments [47]. They are characterized by the following demand dimensions:

1. **Physical Demand**: Tasks involve physical objects or take place in real-world environments, requiring bodily movement and interaction with the surroundings [97, 100].
2. **Spatial Demand**: Tasks are embedded in a three-dimensional space, necessitating navigation, spatial awareness, and understanding of spatial relationships between objects, actors or work areas [13, 16, 35, 158].
3. **Cognitive Demand**: Tasks demand significant mental processing, including problem-solving, decision-making and procedural understanding [16, 87, 100].
4. **Technical Demand**: Tasks depend on specialized knowledge or skills, often involving complex tools, machinery or domain-specific expertise [8, 18, 131].

This multidimensional task characterization builds on prior research in HCI and CSCW [2, 35, 43]. Cognitive and technical complexity have been discussed in instructional and systems design literature [87, 131]. The embodied and spatial nature of interaction is emphasized in the foundations of embodied

[1] Here, *"complex"* refers to the joint presence of substantial embodied/spatial, cognitive, and technical demands, rather than to formal taxonomies of task complexity in organizational psychology (e.g. multiple solution paths, uncertainty, interdependence) [28, 165].

interaction [44] and spatial computing [13]. Finally, engineering collaboration often involves hybrid environments requiring shared spatial reasoning and coordination [92, 137].

In the remainder of this work, such scenarios are referred to as Asynchronous Collaboration on Complex Tasks (ACCT), with a particular focus on the engineering domain. The interplay of spatial, physical, cognitive, and technical complexities in collaborative work that occurs across time is captured by ACCT. Given these characteristics, Asynchronous Collaboration (AC) in complex tasks poses unique challenges, particularly in maintaining shared understanding, continuity and context without real-time interaction. Immersive technologies, such as XR, offer potential solutions by embedding information directly into the spatial and physical context of work.

Therefore, this research investigates the following overarching question:

> ❓ **RQ:** How do immersive representations in XR affect ACCT compared to non-immersive representations?

To address this overarching question, the research is further structured around two interrelated sub-questions. The first (**RQ1**) explores whether XR technologies can be purposefully designed to support asynchronous collaboration on complex engineering tasks.

> ❓ **RQ1:** Can XR be used to improve ACCT?

This question focuses on the design space: how XR systems should be structured, which interaction modalities are most effective, and how core mechanisms like avatar capture or spatial guidance can improve communication and continuity.

The second question (**RQ2**) investigates how such XR-based approaches perform in comparison to conventional non-immersive tools.

> ❓ **RQ2:** How does ACCT perform with XR compared to non-immersive representations?

It evaluates the resulting system in terms of usability, task efficiency, user experience and perceived presence. Together, **RQ1** and **RQ2** guide the design and validation process and ensure that the proposed framework is both conceptually sound and practically effective.

Figure 1.1 summarizes the overview of the main objectives and research questions for this thesis.

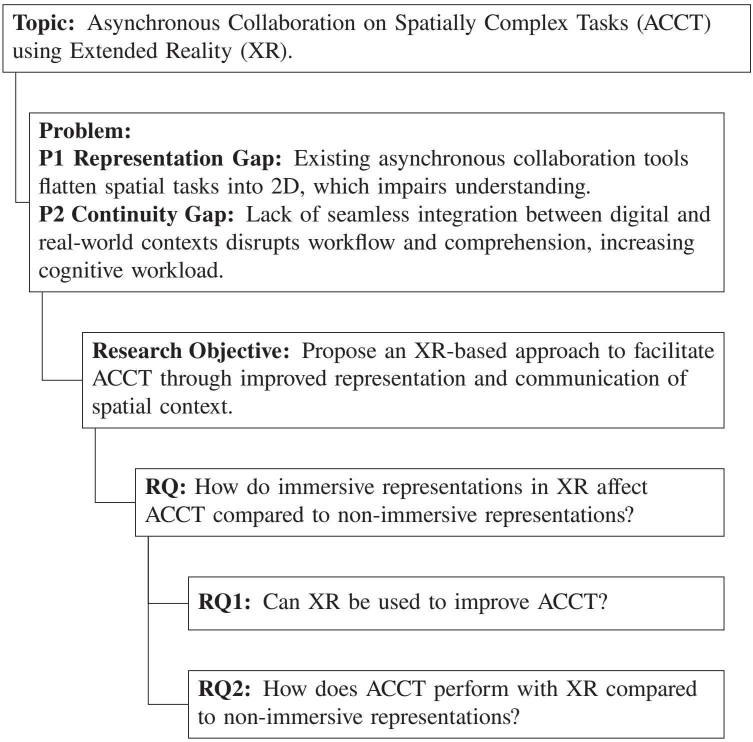


Figure 1.1: Overview of the identified research gaps, objective, and research questions

1.4 Research Methodology

To address the research questions and bridge the identified gaps, this work adopts the DSR methodology [72]. As depicted in Figure 1.2, DSR provides a structured approach for developing, evaluating and refining innovative artifacts that solve real-world problems while contributing to scientific knowledge.

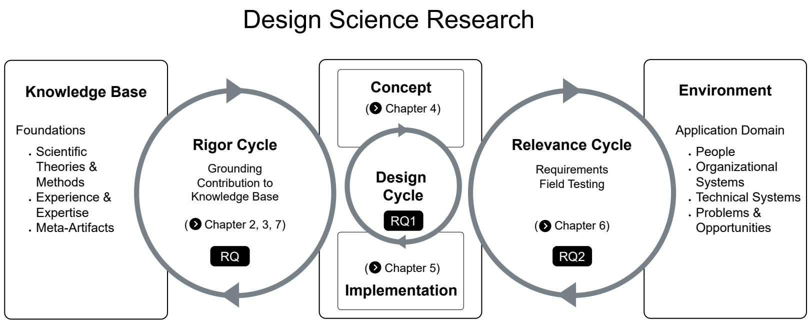


Figure 1.2: Adapted Design Science Research framework illustrating the relevance, rigor, and design cycles guiding this thesis, based on [72]

This thesis introduces a modular framework designed to support asynchronous collaboration on complex engineering tasks through immersive technologies. The DSR methodology provides a fitting foundation for this effort, as it combines systematic design with rigorous evaluation across three interrelated cycles: the Relevance Cycle, the Design Cycle and the Rigor Cycle.

In addition, the methodology incorporates experimental research and a mixed-methods approach, which enables both objective evaluation and contextual understanding. Quantitative methods provide measurable comparisons (e.g. performance, usability), while qualitative insights capture user experience and situational factors that are critical in human-centered systems.

1.4.1 Design Science Research Framework

The subsequent sections further describe the the Rigor, Design, and Relevance Cycles within this research.

Rigor Cycle

The Rigor Cycle ensures that the research is grounded in and contributes to the broader scientific Knowledge Base. This work draws on theoretical foundations (❯ **Chapter 2**) as well as related work on collaborative work using immersive environments (❯ **Chapter 3**) that provide a basis for framing and addressing the main research problem **RQ**.

This cycle also encompasses the derivation of theoretical and design implications based on the outcomes of this research (❯ **Chapter 7**). The findings from the experimental evaluation, when interpreted through the theoretical lens established earlier, contribute back to the Knowledge Base. These contributions include both conceptual insights and practical design guidelines for asynchronous collaboration in immersive environments – thereby enriching the scientific understanding of ACCT and informing future research and system development.

Design Cycle

The Design Cycle addresses **RQ1**, which investigates how immersive technologies can be designed to facilitate asynchronous collaboration in complex engineering tasks.

This process involves iterative development of the immersive collaboration framework through multiple prototyping stages. It begins with conceptual modelling (❯ **Chapter 4**), followed by the implementation of core components across XR platforms (VR, AR, MR), and the evaluation of individual modules and interaction techniques (➔ **Chapter 5**).

Each stage builds on the previous and feeds back into the next, forming a continuous loop of design, formative testing and refinement. Evaluation is conducted not only for technical feasibility, but also to assess usability, user experience and alignment with spatial, procedural and embodied characteristics of asynchronous collaboration in engineering contexts.

This iterative approach ensures that the resulting framework is both functionally robust and contextually appropriate for its intended domain.

Relevance Cycle

The Relevance Cycle connects the research to the practical application domain. It ensures that the developed solution addresses real-world needs by deriving problem relevance from the environment and validating the artifact in authentic settings through field-testing and stakeholder engagement. This alignment with real-world practice is a key principle in DSR, as it ensures both the utility and contextual validity of the proposed solution.

In this thesis, the application domain involves engineering collaboration scenarios that are asynchronous, spatially complex and physically embedded. These contexts pose unique challenges, due to the lack of co-presence, disrupted workflow continuity and inadequate spatial representation across time.

To address **RQ2**, the proposed framework is evaluated in a representative industrial setting using a robotic cell (➔ **Chapter 6**). Participants perform engineering tasks under immersive and non-immersive conditions, enabling a comparative assessment. This real-world validation assesses the system's applicability to practical engineering workflows, thereby closing the loop between theoretical insight, technical design and application context.

1.5 Research Contributions

This dissertation makes the following key contributions to the field of asynchronous collaboration in immersive engineering environments:

- **Theoretical Contribution:** A theoretical framing connecting embodied interaction with spatial and temporal continuity in asynchronous XR collaboration for physical engineering work, deriving implications for design and evaluation (**Chapters 2 and 7**).
- **Conceptual Contribution:** A concept to address asynchronous communication, coordination, cooperation – individually and combined – tailored to spatially complex engineering tasks (**Chapter 4**).
- **Technical Contribution:** A functional XR-based collaboration system, integrating immersive messaging, proxy interaction and spatial coordination tools to support asynchronous workflows (**Chapter 5**).
- **Methodological Contribution:** An evaluation blueprint that links research questions to validation instruments (e.g. workload and presence measures), documents evaluation design and protocols, and specifies analysis procedures, enabling reproducible assessment in comparable asynchronous, spatially procedural XR settings (**Chapters 4 and 6**).
- **Empirical Contribution:** A comparative user study in an engineering context, providing evidence of improved spatial understanding, procedural clarity, and user engagement under the XR condition versus traditional video-based workflows (**Chapter 6**).

These contributions are developed and validated using the DSR methodology and address current limitations of XR tools for asynchronous, spatial and procedural engineering collaboration.

1.6 Thesis Structure

The structure of this thesis is guided by the overarching goal of supporting asynchronous collaboration in complex engineering scenarios through immersive technologies. Each chapter contributes to this objective as part of the DSR methodology, as outlined in Figure 1.2.

Following the introduction (❯ **Chapter 1**) which outlines the motivation, problem space and research approach, ❯ **Chapter 2** establishes the theoretical foundations by defining key concepts for asynchronous collaboration in immersive environments. This conceptual groundwork is expanded in ❯ **Chapter 3**, which surveys related work in XR-based collaboration and engineering support tools. Together, these chapters identify the design space and highlight gaps that the proposed framework addresses.

Building on these insights, ❯ **Chapter 4** introduces the conceptual design of the *ACIE Framework* which structures asynchronous collaboration across the dimensions of communication, coordination and cooperation. The modular components are motivated by theoretical foundations, related work and domain-specific requirements. ❯ **Chapter 5** details the technical realization of the framework across XR platforms. Each module is implemented and evaluated in formative studies, with a focus on system usability.

The integrated system is then applied in a realistic engineering setting in ❯ **Chapter 6**. A comparative field study evaluates the immersive system against a conventional video-based baseline, assessing performance, user experience and perceived presence. Finally, ❯ **Chapter 7** integrates a critical discussion of the findings, summarizes the contributions and outlines directions for future work toward more flexible, embodied and time-independent collaboration in engineering and beyond.

1.7 Summary and Outlook

This chapter outlined the motivation, problem space and methodological orientation of the thesis. It identified limitations in current collaboration tools for supporting asynchronous, embodied and spatial engineering tasks – especially those involving complex physical environments. Extended Reality (XR) technologies were introduced as a promising avenue to address these gaps by enabling immersive interaction across both spatial and temporal dimensions.

The chapter established the main research objective: to develop and evaluate a framework that supports asynchronous collaboration on complex tasks using XR. It introduced the concept of Asynchronous Collaboration on Complex Tasks (ACCT) and formulated two guiding research questions – one focused on the design of XR-based collaboration mechanisms, the other on their evaluation.

To address these questions, the work adopts a DSR methodology, combining iterative design, technical implementation and empirical validation in realistic engineering contexts.

In Simpler Terms

This chapter explained why engineers need better ways to work together when they are not in the same place or working at the same time. While today's tools like video calls and shared documents help with basic tasks, they do not support hands-on spatial work – like configuring machines or solving problems in real physical spaces.

Engineering tasks are often complex: they involve 3D environments, technical steps, and physical actions. Right now, there is no effective way to capture and share this kind of work over time. That is where XR (Extended Reality) comes in – it can transfer digital instructions or past actions directly into the real world.

This thesis explores how XR can help people collaborate across time and space in engineering. This is done by designing and testing a new framework that supports spatial, procedural and embodied collaboration – even when people are not working at the same moment.

How to Proceed

The central research problem identified in this chapter is addressed through a conceptual and technical framework introduced in the upcoming chapters. While the theoretical and empirical groundwork is detailed in the following sections, readers interested in the core design contribution may proceed directly to:

For a full understanding of the underlying principles and design rationale, the following chapters provide the necessary foundation:

❯ **Chapter 2** outlines the theoretical underpinnings from CSCW, HCI and XR.

❯ **Chapter 3** reviews related work and current systems, identifying open challenges and gaps addressed by the framework.

❯ **Chapter 4** introducing the *ACIE Framework*, a modular concept for enabling asynchronous collaboration in XR across communication, cooperation and coordination.

2 Theoretical Foundation

In this chapter, the theoretical foundations relevant to the design and evaluation of asynchronous collaboration in XR environments are established. Core concepts and models from the fields of Computer Supported Cooperative Work (CSCW), Human Computer Interaction (HCI) and immersive technology research are synthesized to provide the conceptual basis for this thesis. Emphasis is placed on collaboration frameworks, spatial interaction, embodiment and usability, which are essential for framing the research questions and informing subsequent design and evaluation.

This chapter defines the central constructs and analytical lenses used throughout the work. It lays the groundwork for understanding asynchronous interaction in XR environments and supports the methodological and conceptual decisions developed in later chapters.

The remainder of the chapter is structured as follows: ❯ Section 2.1 introduces foundational CSCW models, including the 3C framework and its relevance to asynchronous collaboration. ❯ Section 2.2 covers key principles of immersive technologies, including embodiment, presence, and related evaluation concepts such as usability and cognitive load. Finally, ❯ Section 2.3 summarizes the key insights and their relevance to the remainder of this dissertation.

2.1 Computer-Supported Cooperative Work

CSCW investigates how collaborative activities and their coordination can be supported by computer technology. The foundational CSCW matrix by Johansen [82] categorizes collaboration based on whether activities occur simultaneously (synchronous) or at different times (asynchronous), and whether participants are co-located or distributed. As shown in Figure 2.1, the original matrix spans the time and space dimensions.

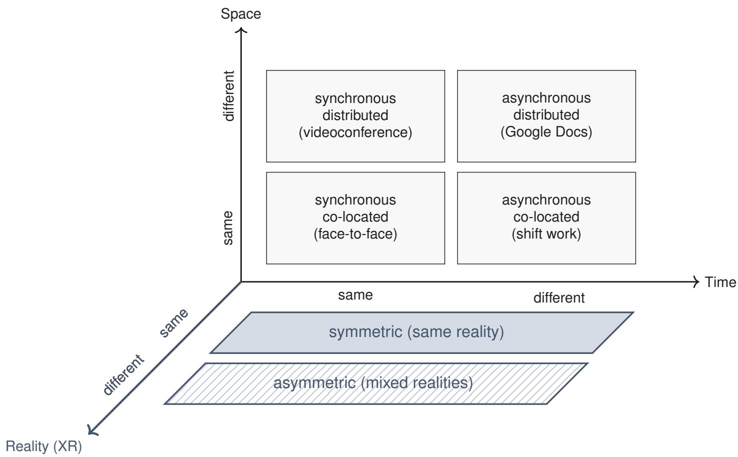


Figure 2.1: Extended CSCW design space with the additional XR dimension distinguishing symmetric (same reality) and asymmetric (mixed realities) collaboration. Adapted from Johansen's time–space matrix [82].

In the space dimension, co-located collaboration takes place at the same physical location, while remote collaboration involves participants contributing from distributed locations, connected through a shared digital environment. In the time dimension, collaboration is synchronous when participants work simultaneously, and asynchronous when they contribute at different points in time – as is often the case in globally distributed teams. The original matrix was extended by the XR dimension, as highlighted in Figure 2.1. It expands the design space of CSCW by differentiating whether collaborators are present in the same reality or across mixed realities.

In the XR dimension, symmetric collaboration means that all users share the same level of immersion – such as all being in VR – and is the most common form of XR CSCW as it is easier to implement [38, 48]. Asymmetric collaboration emerges due to differences in technological setups or user roles [11]. An example scenario of asymmetric collaboration in immersive environments is a co-located team using AR devices. These users can view their physical surroundings, shared virtual content, and remotely generated annotations or instructions directly integrated into their work environment. Simultaneously, remote collaborators join through VR or traditional video conferencing systems. In such setups, remote users interact via video streams or 3D reconstructions of the work environment, observing co-located users' avatars or virtual representations, and jointly manipulating shared digital objects. Such hybrid setups leverage the complementary capabilities of different technologies to facilitate collaboration effectively. Asymmetric collaboration is discussed in more detail in ❯ **Chapter 4**, Section 4.5.

Transitions across dimensions – between synchronous and asynchronous, co-located and remote setups, and varying XR configurations – remain a challenge and are still underexplored in current research [48]. This thesis focuses on Asynchronous Collaboration (AC) in immersive environments, with particular attention to how it unfolds across spatial and XR configurations.

> **Literature:** The original CSCW matrix by Johansen [82] was extended for immersive environments in Mayer et al. [2], additionally distinguishing between symmetric (same reality) and asymmetric (different realities) collaboration.

2.1.1 Immersive Asynchronous Collaboration

Asynchronous collaboration refers to cooperative work where participants contribute at different times. While this temporal decoupling allows for flexibility and supports distributed workflows, it introduces several challenges – particularly when tasks are practical or spatially complex [135]. Traditional tools, such as video streaming, often fail to preserve spatial context or make embodied interaction visible across time [56, 153]. In this context, XR technologies offer new possibilities: by enabling embodied interaction, spatial annotations, and immersive representations, they can support continuity and coordination in asynchronous settings embedded directly within the work environment [38, 48]. However, despite this potential, realizing effective AC in immersive environments remains a significant challenge [36, 77]. This section outlines key challenges, as summarized in Table 2.1, that need to be addressed to make asynchronous collaboration in XR both usable and effective.

Table 2.1: Key challenges of asynchronous collaboration in XR environments

Challenge	Description
Temporal Disconnection	Lack of real-time interaction complicates feedback, coordination, and task handovers.
Loss of Continuity	Difficulty reconstructing task state, rationale, and progress across sessions.
Spatial and Embodied Disconnection	Limited access to spatial cues and embodied actions in the absence of co-presence.
Tooling Limitations	Missing support for annotations, history, and versioning in asynchronous XR workflows.

Temporal Disconnection: The absence of temporal co-presence eliminates immediate feedback, negotiation, and real-time interaction [77]. Misunderstandings cannot be resolved in the moment, and collaborators are required to anticipate what information others might need [154]. Task handovers become fragile and error-prone, especially when sequential steps or shared decisions are involved [38, 77]. In traditional tools, such challenges are amplified by limited perception of others' actions due to constraints in visualization and interaction [52]. Immersive technologies offer new opportunities to mitigate this issue by enabling spatial anchoring of contributions embedded in the work environment, embodied replays of interactions instead of flat visualizations on 2D screens, and persistent representations of sensory input that can be used to reconstruct work processes across time [77, 172]. However, such features are not yet widely available or integrated, and temporal disconnection remains a critical challenge in AC [36, 77].

Loss of Continuity and Shared Context: Reconstructing the state of a collaborative activity after a period of absence is one of the core challenges of AC [52, 77]. What has been done, why certain decisions were made, or what remains to be completed must be made accessible across time. Missing context not only disrupts workflow but also undermines confidence in continuing or modifying previous contributions [36]. In engineering domains such as design, co-creation, guidance, or maintenance, this challenge is particularly pronounced due to the interdependence of spatial and procedural information [35, 97]. Immersive technologies offer potential to preserve context over time, for instance by embedding task artifacts and messages directly into the 3D scene [77, 99, 156]. However, such mechanisms are rarely designed with asynchronous use in mind, and continuity across sessions remains limited.

Spatial and Embodied Disconnection: Collaborative work that relies heavily on spatial cues, embodiment, and mutual visibility – as is often the case in engineering – can significantly benefit from XR [36, 156]. In asynchronous settings, participants collaborate with others who are not present at the same time. They cannot rely on cues like gaze, deictic gestures, or body language to direct attention or signal intent [77]. Identifying what others interacted with – or how – is often ambiguous [52]. As a result, spatial disorientation and misinterpretation of actions are common. Embodied representations such as avatars, proxies, or temporal replays in XR may help bridge this gap by simulating past activity or visualizing user presence [36, 52, 154]. Yet, existing implementations are often limited in fidelity or integration, making it difficult to fully leverage embodiment as a coordination aid in asynchronous workflows.

Limitations of Supportive Tools: Most current XR systems are designed primarily for synchronous collaboration (as later presented in **Chapter 3**), while most non-immersive AC CSCW tools do not support physical interaction or spatial representations [35, 154, 156]. As a result, core capabilities for asynchronous workflows – such as temporal navigation, persistent annotations, interaction history, or spatial versioning – are rarely available [36, 172]. Without such features, collaborators are left without the ability to trace others' actions, revisit task progress, or coordinate their work contributions across time [159]. The absence of asynchronous design principles in current tools not only impairs usability but also constrains the types of collaboration that immersive systems could otherwise support [77]. At the same time, XR offers substantial potential to support AC in domains where current approaches remain limited [38].

2.1.2 The 3C Model

The foundational structure of collaborative work is often described through the 3C Model - *Communication*, *Coordination*, and *Cooperation*, as illustrated in Figure 2.2. This conceptual triad was introduced by Ellis et al. [46] and later formalized by Leimeister [95] in the context of Collaboration Engineering. The model continues to serve as a reference model for designing and evaluating collaborative systems. To address the challenge posed by **RQ1**, this thesis will employ the 3C model as a guiding framework for the design and evaluation of ACCT.

> **RQ1:** Can XR be used to improve ACCT?

The following sections discuss each dimension related to the challenges and opportunities of AC in XR, introducing the corresponding sub-questions related to **RQ1**.

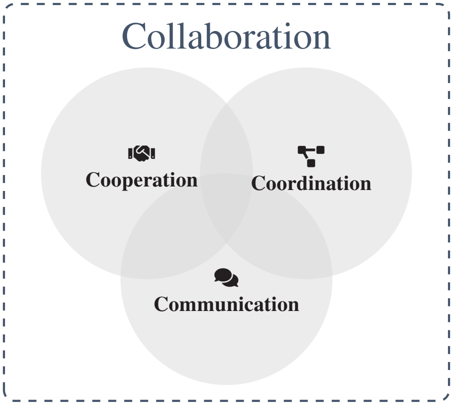


Figure 2.2: The 3C model of collaborative work, illustrating communication, coordination, and cooperation as foundational dimensions of collaboration

Communication: Communication refers to the process of exchanging information between individuals or groups in order to reach a common understanding [95]. In collaborative settings, this includes verbal interactions, non-verbal cues (such as gestures, gaze, or body language), and symbolic communication via shared digital artifacts [48]. In immersive environments, communication can be enriched by spatial and embodied modalities to allow multimodal communication [36]. However, in AC the loss of immediate feedback due to non-present collaborators is representing a significant challenge [77, 154]. In asynchronous settings, where real-time dialogue is not possible, the absence of direct feedback makes it necessary to externalize intent and context. XR technologies provide new possibilities for asynchronous communication, such as persistent spatio-temporal information and user representation using proxies (➜ Section 2.2.1) compensating the loss of immediate feedback [38, 66].

These forms of mediated, multimodal expression aim to bridge the temporal gap, but are still rarely implemented in current systems [48]. As such, research question **RQ 1.1** addresses how immersive technologies can effectively support and improve asynchronous communication:

> ❓ **RQ1.1:** Can avatar capture and playback in XR be used to improve asynchronous communication in CT?

Cooperation: Cooperation refers to how collaborators interact with shared content and each other within a shared workspace to contribute to a joint outcome [95]. In immersive asynchronous systems, this includes joint interaction with virtual objects as well as indirect interactions with other collaborators across time. Due to temporal fragmentation, direct cooperation is no longer possible [127]. Instead, cooperative actions must remain persistent and interpretable for others who enter the environment at a later time [34, 66]. This requires systems to provide mechanisms that support the visibility, accessibility, and contextual understanding of prior user actions.

To enable cooperation across time – *mediate* or *indirect* cooperation – immersive systems must approximate direct interaction through indirect mechanisms that compensate for the absence of co-presence and immediate interaction [52]. Research question **RQ 1.2** explores how immersive technologies can better support and improve asynchronous cooperation through immersive technologies:

> ❓ **RQ1.2:** Can interactive question–answer mechanisms in XR be used to improve asynchronous cooperation in CT?

Coordination: Coordination is the communication-mediated temporal and structural alignment of collaborative activities [95]. In asynchronous immersive systems, this includes how shared workspaces are navigated, how task progress is marked, and how the temporal sequence of activities is made visible [38, 77]. Unlike in synchronous collaboration, where timing and turn-taking are negotiated in real time, asynchronous coordination must rely on explicit mechanisms. These include spatio-temporal change management, such as progress indicators, and versioning, features that allow collaborators to track and manage contributions over time [172].

For text-based information, sophisticated solutions, such as Git[1], already exist and work well. For immersive changes, such as changes to 3D objects and scenes, these tools are insufficient or require further data processing, such as conversion to a text based file format [42, 172]. In immersive contexts, work coordination mechanisms are often not considered [139]. Yet, they are beneficial to enabling smooth handovers, avoiding conflicts, and ensuring workflow continuity [38, 77]. Supporting coordination effectively in AC environments thus requires systems to represent and manage spatio-temporal structures and changes consistently. Research question **RQ 1.3** explores how such coordination requirements can be enabled and improved in XR:

> **RQ1.3:** Can step-by-step organization of recordings and immersive version control in XR be used to improve asynchronous coordination in CT?

Together, these three aspects form the foundation of collaborative work. The preceding discussion shows that each dimension poses distinct challenges when extended into asynchronous immersive contexts.

[1] Git is a widely-used version control system for tracking changes in source code during software development: `https://git-scm.com/`.

Assumption: To enable effective asynchronous collaboration, CSCW systems must support asynchronicity across the 3 C's of collaborative work: communication, cooperation, and coordination.

2.2 Extended Reality

Collaborative work on CTs, as defined in Section 1.2, often exceeds the capabilities of traditional CSCW tools and "flat" media. These systems frequently fall short in supporting the spatial, embodied, and multimodal dimensions necessary for effective collaboration in real-world engineering environments. As discussed earlier, XR opens up new possibilities for more natural, immersive, and workspace-embedded interaction. It expands the design space for collaborative systems by addressing known limitations in continuity, representation, and spatial context.

XR refers to the spectrum of immersive technologies that blend physical and digital environments to varying degrees [4]. This includes AR, AV, and VR, all of which fall under the umbrella of XR. Originally, Milgram et al. [113] referred to the spectrum between the "real" and "virtual" as MR. These representations are commonly arranged along the Reality-Virtuality Continuum proposed by Milgram et al. [113], which provides a conceptual framework for understanding how digital content is integrated with real-world perception (see Figure 2.3).

Augmented Reality: AR is positioned near the real end of the Reality–Virtuality Continuum, as it minimally augments the physical environment with digital information. Unlike fully immersive systems, AR preserves a direct view of the real world while overlaying virtual content that enhances perception or task execution [113]. Typical examples include spatially anchored annotations, highlighting and drawing, or virtual but "flat" instructions that appear integrated into the user's surroundings.

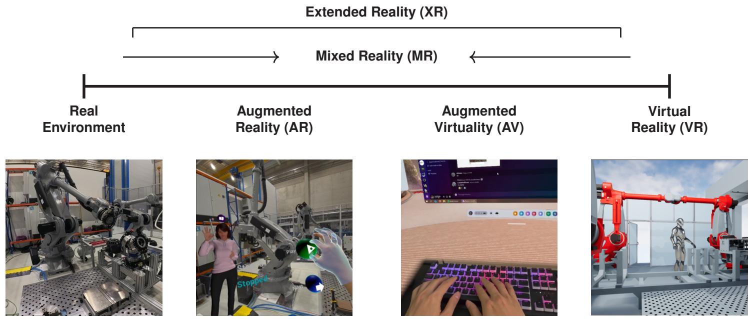


Figure 2.3: The Reality–Virtuality Continuum, adapted from Milgram et al. [113], with example imagery for AR, AV, and VR, and XR as the umbrella term.

Augmented Virtuality: Further along the continuum lies AV, which primarily consists of virtual environments enriched with live inputs or reconstructions from the physical world [48, 113]. In contrast to AR, where reality is enhanced by virtual elements, AV involves virtual spaces that are augmented by real-world data such as video feeds, sensor streams, or spatially integrated models of physical environments [38]. This allows the virtual environment to reflect or respond to real-world conditions in real time.

Virtual Reality: At the far end of the continuum, VR immerses the user entirely in a digitally generated environment, fully occluding direct perception of the physical world [38, 113]. The user is placed in a synthetic space, with all visual, auditory, and sometimes tactile input simulated by the system. VR is currently the most widely adopted immersive technology, particularly in entertainment, simulation, and training contexts [38, 48]. For example, in gaming applications, users may explore fictional worlds, interact with virtual characters, or complete spatial tasks entirely within a virtual setting, disconnected from the real environment.

Mixed Reality: A comparatively confusing concept within immersive environments is MR. Systematic reviews and expert interviews reveal that MR is defined inconsistently, with at least six overlapping definitions reported in the literature [145, 147]. Consequently, the boundaries between MR, AR, and AV are often blurred, leading to conceptual confusion and uneven terminology [110].

For the purpose of this thesis, MR is defined as the region between the real and fully virtual ends of the Reality–Virtuality Continuum – a blended setting in which digital and physical elements coexist and interact in real time [38, 48]. An MR application may, for example, overlay a virtual structure that aligns precisely with the geometry of a physical room, or endow real objects with interactive virtual properties. Unlike traditional AR, which typically displays flat or floating metadata anchored in the physical scene, and unlike VR, which wholly replaces real-world perception, MR emphasises seamless fusion, enabling direct, bidirectional interaction between real and virtual content.

Particularly in collaborative settings, MR offers two key affordances identified in the literature: *seamlessness* and *reality enhancement* [12]. Seamlessness refers to the uninterrupted integration of real and virtual content, while reality enhancement describes the augmentation of human perception and understanding through digital overlays. These affordances are essential to support intuitive and context-aware collaboration in immersive work environments.

2.2.1 Key Concepts for XR Collaboration

To understand how XR can facilitate ACCT and how it supports **RQ2**, it is essential to examine its fundamental concepts.

> ❓ **RQ2:** How does ACCT perform with XR compared to non-immersive representations?

This section introduces the key aspects that shape user experience and inform system design of immersive collaborative environments.

Immersion: Immersion describes the degree to which a system envelops the user in a virtual or augmented environment, creating a sense of presence and detachment from the outside world [94, 144]. It amplifies the embodied and spatial aspects of virtual environments. Higher levels of immersion are associated with stronger presence, which plays a crucial role in how users perceive, navigate, and interact with immersive environments.

Presence: Presence refers to the subjective feeling of "being there" within a mediated environment. While immersion describes the objective qualities of a system (e.g. display resolution, tracking accuracy), presence captures the user's subjective experience of spatial and social involvement. Because presence is critical for user engagement and comprehension in immersive settings – especially in the absence of real-time co-presence – it serves as a central construct for evaluating asynchronous XR collaboration [38, 139, 154]. Furthermore, presence was found to have a positive correlation with task performance and thus increases the quality of collaboration [128, 129, 164]. This study therefore examines presence and its impact on the effectiveness of ACCT, as explored in **RQ 2.2**:

> ❓ **RQ2.2:** How does XR ACCT perform in terms of presence compared to non-immersive representations?

Presence is commonly described through multiple dimensions. In this work, two particularly relevant aspects are emphasized in relation to asynchronous collaboration in XR – spatial presence and social presence – as illustrated in Figure 2.4.

Spatial presence is the sense of being physically located in the virtual or augmented space. This aspect is especially important when asynchronous users must interpret or respond to spatially situated work. Workspace and task awareness improve engagement and understanding, and positively affect task – and thus collaboration – performance and quality [38, 139, 172]. Social presence

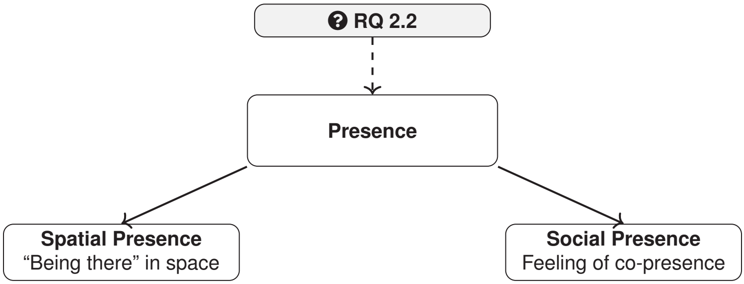


Figure 2.4: Presence and its subdimensions in XR collaboration with associated research questions

refers to the feeling of being co-present with others –even when represented virtually or asynchronously. Maintaining this sense of social connection across time is a key challenge in delayed collaboration. The positive association between co-presence and task awareness has been linked to better communication and user performance in XR collaboration [38, 139, 154]. Consequently, because collaboration quality hinges on recovering shared spatial context and collaboration across time depends on restoring social cues, this thesis examines both, spatial and social presence as the primary pathway by which XR may benefit ACCT.

Embodiment: Embodiment refers to the degree to which users feel that their body is represented and can act meaningfully within an XR environment. This includes virtual body ownership, alignment between real and virtual movements, and the ability to use one's body for interaction [85]. Embodiment enhances realism, expressiveness, and engagement – especially in collaborative scenarios where gesture, posture, and spatial positioning carry non-verbal communication cues [64]. It is a strong way to provide both consequential communication and conversational resources in XR CSCWs [139].

Interaction: Interaction in XR environments spans multiple modalities, including gaze, gesture, voice, touch, and controller input [19, 94]. Effective interaction design enables users to manipulate virtual content, navigate immersive spaces, and communicate intentions. For asynchronous collaboration, interaction mechanisms must also support recording, playback, and interpretation of prior actions [36, 151].

Proxy: Proxies refer to representations of non-present users in asynchronous XR settings [45]. These may take the form of avatars, ghost trails, recordings, or interactive cues that simulate or replay past activity [67]. Proxies are essential to compensate for the absence of real-time co-presence by conveying intent, actions, and attention of collaborators across time. They contribute to social presence and continuity by allowing others to see, follow, or respond to asynchronous contributions.

Registration: Registration is the process of accurately aligning virtual content with the physical world [7]. It ensures that digital elements appear stable and fixed in space relative to their real-world context, which is essential for maintaining spatial continuity and usability across time and users [24]. Reliable registration depends on robust localization – the system's ability to track its own position and orientation within the physical environment. In collaborative and asynchronous XR, registration underpins many functions, including shared annotations, scene consistency, and accurate replays of past interactions [89].

2.2.2 XR Technology Considerations

XR systems are designed to simulate environments that users can experience through multiple sensory channels. These immersive systems function as "simulators" with a human in the loop, creating artificial perceptual stimuli that approximate how humans experience reality. The more accurately these signals mimic reality, the higher the system's level of immersion, which then impacts how engaged and present the user feels. Figure 2.5 illustrates examples of XR devices used to deliver multi-sensory feedback.

(a) Haptic vest **(b)** Exoskeleton glove **(c)** Omnidirectional floor **(d)** Olfactory HMD

Figure 2.5: XR devices: (a) haptic vest with a VR HMD, (b) exoskeleton glove for force feedback, (c) Disney's HoloTile omnidirectional floor, and (d) olfactory headset

Visual Stimulation: Visual immersion is typically achieved through digital displays and projections that replace or augment the user's Field of View (FoV) [94]. Common setups include handheld devices with flat screens, projection-based environments (e.g. CAVE[2] or SAR[3] systems), and Head Mounted Displays (HMDs) for AR, MR, or VR [14, 37]. HMDs offer the most portable and immersive visual simulation, combining stereoscopic rendering with head tracking to maintain spatial coherence. These visual modalities are critical for spatial understanding, object manipulation, and orientation in immersive collaborative environments.

Auditive Stimulation: Auditory cues support spatial orientation, event awareness, and interpersonal communication in XR. Spatialized audio with head tracking allows sound sources to be localized relative to the listener, which aids distance and direction judgments and supports turn-taking in collaboration [163]. Audio cues (e.g. earcons or auditory icons) can provide notifications without adding visual clutter, and microphone as well as signal processing choices affect comprehensibility and privacy in co-located and distributed use [21, 62].

[2] CAVE stands for Cave Automatic Virtual Environment, an immersive virtual reality environment where projectors are directed to multiple walls of a room-sized cube to surround the user with large-scale stereoscopic visuals.

[3] SAR stands for Spatial Augmented Reality, a technology that uses projectors to overlay digital images onto physical objects in the real world.

Haptic Stimulation: Haptic simulation reproduces forces and torques at the joints and end-effectors, letting users perceive resistance, weight, or inertia during interaction [148]. Typical devices include exoskeleton gloves that apply joint torques or fingertip forces (Fig. 2.5b), as well as grounded multi-joint systems – akin to robotic arms – that can render 6 Degree of Freedom (DoF) forces/torques (e.g. devices by Haption). In collaborative XR, such kinesthetic cues enhance realism during joint manipulation tasks and support implicit coordination through physically meaningful constraints.

Tactile Stimulation: Tactile feedback targets the skin to reproduce surface-level sensations such as vibration, pressure, shear, texture, or temperature [39, 88]. Wearable arrays embedded in gloves or garments stimulate contact, edges, or material qualities, while full-body vests or suits deliver distributed cues for events like impact or proximity (Fig. 2.5a). Many contemporary devices combine both modalities – for example, force-feedback gloves augmented with fingertip vibrotactile actuators – but we distinguish them here for clarity of mechanism and effect.

Olfactory Stimulation: Olfactory devices deliver odorants to simulate smells associated with environments or materials. Effects are typically subtle but can increase a sense of presence and support memory encoding [5, 71]. Practical deployment remains limited due to latency, lingering odors, and repeatability constraints, so usage in collaborative engineering contexts is rare (see Fig. 2.5d for an example device).

Locomotion: To further deepen immersion, XR systems may incorporate full-body tracking or physical locomotion platforms. These allow users to navigate virtual spaces naturally, using walking, leaning, or turning motions. Movement simulation is essential for spatial coordination in collaborative tasks that rely on viewpoint control or physical gestures. Locomotion can be achieved naturally by walking within the physical environment, provided sufficient space is available. However, when physical space is limited, software-based techniques can be employed to alter the perceived virtual environment – for example, through redirected walking – allowing users to explore larger virtual areas without noticing spatial manipulation [132, 149]. In addition, various hardware solutions exist to support locomotion in constrained spaces, including omnidirectional treadmills and locomotion platforms that sense user movement and translate it into virtual displacement while keeping the user physically stationary (see Fig. 2.5c) [111]. More recently, wearable solutions such as motorized shoes have emerged, which simulate walking by subtly shifting the user back to their original position, preserving the sensation of natural locomotion [146, 169].

Software Platforms: Most immersive applications are developed using common real-time 3D engines such as Unity 3D, Unreal Engine, or Godot. These platforms offer rendering, physics, and interaction modules, but often require adaptation and custom implementation to meet engineering-specific requirements. In engineering research contexts, there are also specialized frameworks such as PolyVR[4] and COVISE[5] provide domain-specific modules for instance immersive simulation, engineering data formats and streams and scientific visualization.

XR technology choices should align with task demands and social requirements. Projection systems support co-located sharing in a common space; handhelds are portable but tie up the hands and offer minimal immersion. Optical see-through HMDs preserve real-world context but limit FoV, whereas passthrough HMDs maximize immersion while potentially reducing face-to-face cues.

[4] PolyVR is an open-source virtual reality framework used for immersive simulations and visualizations, primarily in engineering and research contexts: `https://www.imi.kit.edu/46_2274.php`

[5] COVISE (Collaborative Visualization and Simulation Environment) is a software platform designed for simulations and visualizations, often used in scientific and engineering research for collaborative work and domain-specific applications: `https://www.hlrs.de/de/loesungen/rechentypen/visualisierung/covise`

2.2.3 Technical Suitability for Engineering Use Cases

This thesis targets XR applications for room-scale, physically grounded engineering tasks conducted asynchronously. To support these scenarios, the XR system must fulfill a specific set of technical requirements that ensure usability, flexibility, and immersion in real-world environments. These are summarized in Table 2.2.

Table 2.2: XR technology requirements for asynchronous collaboration in practical engineering contexts

Requirement	Description
In-situ Augmentation	Virtual content must be spatially anchored in the real-world environment to maintain spatial and task continuity [77, 97].
Hands-free Interaction	Users must be able to interact with physical tools and objects without holding or operating the XR device manually [38, 73].
Physical Locomotion	Support for natural walking and movement through large-scale physical spaces, such as shop floors or workstations [35, 120, 139].
Inside-out Tracking	Reliable spatial tracking without external sensors to ensure consistent alignment of virtual content across sessions [35, 66, 139].
Realistic Avatars	Human-like avatars to support social presence, identity, and interpersonal cues during asynchronous interactions [154, 159].
Extensible Software	Support for open, modifiable platforms to allow domain-specific adaptations and integration with engineering workflows [3].
Portable and Standalone	Head-mounted displays that operate without tethering to external PCs, enabling in-field deployment and flexible usage [35, 73].

To meet these requirements, this work employs the Meta Quest 2, Quest Pro, Quest 3, and Microsoft HoloLens. Their on-board inside-out tracking provides robust body-motion capture with no external sensors, and their standalone modes allow seamless field deployment while still supporting tethered development when needed. On the software side, Unreal Engine (UE) was selected for its advanced virtual-character system and open ecosystem, which facilitates rapid prototyping and the integration of project-specific extensions.

2.2.4 Conceptual Challenges and Affordances for Asynchronous XR Collaboration

Asynchronous immersive collaboration imposes unique demands on XR systems that differ from those in synchronous contexts. Prior work has identified several XR-specific challenges relevant to asynchronous use:

Temporal Coordination and Persistence: Since collaborators are not co-present, contributions must persist in the shared environment across time [66]. This includes both object states and actions. Temporal navigation and replay functionalities are essential for enabling participants to reconstruct what others have done [34]. Change awareness, progress markers, and embedded task histories can help bridge temporal gaps between collaborators [36].

Task Anchoring and Spatial Continuity: Immersive collaboration often relies on spatial anchoring – placing tasks, annotations, or objects within the 3D workspace [97]. Maintaining spatial continuity across time and sessions is critical to ensure collaborators can locate, interpret, and act upon shared content [52]. Misalignment or reset of environments can lead to confusion and inefficiency [77].

Embodiment and User Representation: In synchronous XR, embodiment enhances social presence and coordination through avatars, gaze, and gestures [159]. In asynchronous settings, these elements must be reimagined. Concepts like proxies, ghost avatars, or embodied replays have been proposed to simulate presence, signal intent, or visualize actions of temporally absent users (➔ **Chapter 4**, Section 4.3). These mechanisms can support understanding and interaction continuity across time, though they are still underexplored [139].

2.2.5 Technical Constraints and System Capabilities for Asynchronous XR

Building on the conceptual challenges above, practical constraints and system-level requirements for asynchronous workflows are outlined in the following.

Despite the potential, XR technologies also face technical and practical constraints in asynchronous workflows. Limitations include high resource requirements, synchronization challenges, and lack of standardized mechanisms for version control or temporal scene reconstruction [156, 172]. Most XR systems remain optimized for synchronous use, and only few offer explicit support for time-shifted collaboration [38, 48, 97, 139].

> **Design Implication:** Asynchronous collaboration in XR requires systems that provide persistent spatial and temporal context, preserve embodiment across time, and support explicit coordination mechanisms that go beyond traditional synchronous paradigms.

Core Capabilities Needed for Asynchronous XR: To address these limitations, XR systems must go beyond traditional real-time affordances and support capabilities tailored to asynchronous workflows. Key requirements include:

- **Temporal Persistence and State Management:** Allowing contributions (e.g. actions, annotations, objects) to remain accessible and meaningful across sessions.
- **Recording and Replay:** Capturing embodied interactions (e.g. gestures, object manipulation) for future reconstruction and review.
- **Versioning and Change Tracking:** Supporting parallel edits, tracking modifications over time, and resolving conflicts across temporally disjoint users.

- **Scene Serialization and Synchronization:** Ensuring consistent, transferable scene states across devices and time.

These system-level capabilities are critical enablers for time-shifted collaboration and are explicitly addressed in the ACIE framework introduced in **Chapter 4** and evaluated in **Chapter 6**.

2.2.6 Evaluation Construct

Evaluation in CSCW and collaborative XR research typically employs mixed methods that combine quantitative performance metrics (e.g. task time, accuracy, error rate) with qualitative insights (e.g. user feedback, questionnaires). While many studies develop custom measurement instruments tailored to specific systems, several standardized constructs have emerged as reliable indicators of system effectiveness and user experience.

Among these, usability and cognitive workload are particularly prominent in the XR collaboration literature and are often identified as key performance indicators, since they directly influence collaboration quality and the overall user experience [16, 38, 43]. Usability has consistently been highlighted as a critical success factor for adoption and collaboration, shaping efficiency, ergonomics, and user acceptance [108, 123]. Cognitive workload is closely tied to performance, error rates, and user well-being, with excessive workload shown to undermine collaboration outcomes in asynchronous XR tasks [38, 40, 49, 80]. Together, these constructs provide complementary perspectives: usability captures how well the system supports interaction and adoption, while workload reflects the cognitive costs imposed on users in complex, distributed tasks. For this reason, they are employed in this work to address **RQ2**:

> **RQ2:** How does ACCT perform with XR compared to non-immersive representations?

Usability, as defined by the ISO 9241-11[6] standard, captures effectiveness, efficiency, and satisfaction – core aspects of any collaborative tool [78]. Cognitive workload, often assessed using the NASA TLX, reflects the mental demands imposed by asynchronous XR tasks [70]. These two measures were chosen because they comprehensively represent the critical XR concepts discussed earlier – such as immersion, embodiment, and interaction – from a user-centered evaluation perspective.

ISO 9241-11 Usability Standard

In CSCW and HCI research, usability remains a central construct for evaluating the quality and effectiveness of collaborative systems – including XR-based systems. According to the ISO 9241-11 standard, usability is defined as:

> *"The extent to which a system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."* [78]

This definition aligns well with the core demands of asynchronous XR collaboration, where users must complete spatial, embodied, and communicative tasks across time. As such, usability provides a robust lens for assessing how well immersive systems support collaborative work – particularly when users are temporally decoupled.

> ❓ **RQ2.1:** How does XR ACCT perform in terms of usability (effectiveness, efficiency, satisfaction) compared to non-immersive representations?

[6] ISO 9241-11 is a part of the ISO 9241 standard that defines usability requirements and evaluations: `https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en`

The ISO standard breaks down usability into three components: Efficiency, Effectiveness, and user Satisfaction in relation to the system, as illustrated in Figure 2.6.

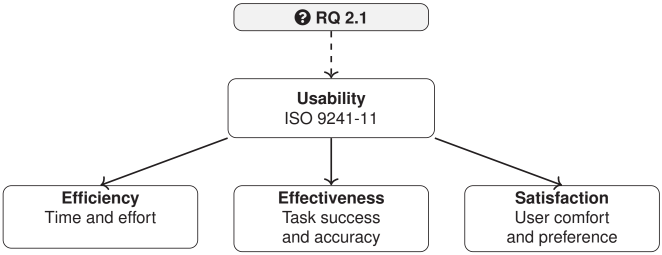


Figure 2.6: Decomposition of usability according to ISO 9241-11

Efficiency: Efficiency measures the resources expended in relation to the accuracy and completeness of task achievement. This includes the time, cognitive effort, and physical interaction required to perform tasks within the system. In asynchronous XR environments, efficient interaction design minimizes friction caused by temporal disconnection, enabling users to navigate, understand, and contribute with minimal effort. Efficient systems streamline workflows and reduce the burden of reorienting to prior context or navigating spatially distributed tasks. Common metrics for evaluating efficiency include task duration, number of steps taken, and physical navigation distance. This work explores how XR systems enable efficient task execution in asynchronous collaboration.

Effectiveness: Effectiveness refers to the accuracy and completeness with which users achieve specific goals within a system. In the context of XR-supported asynchronous collaboration, this includes how well users are able to complete tasks such as interpreting spatial instructions, modifying shared content, or following interaction sequences left by previous collaborators. High effectiveness implies that the system enables users to perform the intended task without errors or ambiguity, ensuring a reliable transfer of knowledge or actions across time. Common metrics for assessing effectiveness include task completion rate, error rate, and the accuracy of actions or outcomes. This work investigates how well XR systems support accurate and complete task execution in asynchronous collaboration.

Satisfaction: Satisfaction captures the users' subjective comfort, acceptability, and preference while using the system. It reflects how enjoyable or frustrating the interaction feels, and whether users perceive the system as supportive of their needs and expectations. In XR, satisfaction is influenced by factors such as spatial coherence, visual clarity, embodiment quality, and the ability to intuitively control or respond to content. Especially in asynchronous settings, satisfaction can hinge on how clearly prior contributions are represented and how well users feel connected to remote collaborators or the shared workspace. Common metrics for satisfaction include standardized questionnaires (e.g. SUS [22], UEQ [93]), Likert-scale ratings of user experience, and qualitative feedback on perceived ease of use, engagement, and frustration.

Cognitive Workload

Cognitive workload refers to the mental effort required to complete a task and manage task demands within a given system. In the context of asynchronous XR collaboration, cognitive workload is a critical factor that influences user performance, task engagement, and error rates – especially in complex, spatially distributed environments that require interpretation of others' actions across time.

A widely used method to assess cognitive workload is the NASA Task Load Index (NASA TLX) [70]. It is a standardized, multidimensional questionnaire developed by NASA to evaluate perceived workload in human-machine interactions. The NASA TLX captures six dimensions of task demand:

- **Mental Demand:** the amount of mental and perceptual activity required (e.g. thinking, deciding, calculating).
- **Physical Demand:** the amount of physical effort required to complete the task.
- **Temporal Demand:** the time pressure felt during task performance.
- **Performance:** the user's perceived success in completing the task.
- **Effort:** how hard the user had to work to achieve the level of performance.
- **Frustration:** the level of stress, annoyance, or discouragement experienced during the task.

The NASA TLX provides a comprehensive view of perceived workload by combining subjective evaluations of each subscale. In XR-based asynchronous collaboration, where users must interpret spatial cues, navigate complex environments, and manage delayed communication, these subscales reveal important insights into system usability and cognitive strain.

2.3 Summary and Outlook

This chapter established the theoretical and conceptual foundations for designing asynchronous collaboration in XR environments. Drawing on models from CSCW, XR, and user-centered evaluation, it introduced the analytical tools and constructs that underpin the remainder of this dissertation. In particular:

- **CSCW and the 3C Model** provided a foundational structure for analyzing asynchronous collaboration in terms of communication, coordination, and cooperation – directly informing the decomposition of **RQ1**.
- **Asynchronous collaboration** was framed within Johansen's time–space matrix as a distinct quadrant, highlighting the unique challenges of decoupled presence, temporal continuity, and spatial context.

- **XR Technologies and Affordances** were discussed with respect to how immersive systems can support asynchronous workflows – including spatial anchoring, persistence, embodiment, and technical constraints.
- **Presence and Embodiment** were identified as key experiential constructs in XR that shape how users interpret, navigate, and connect across time – supporting the development of asynchronous interaction models.
- **Evaluation Constructs**, including usability (ISO 9241-11) and cognitive workload (NASA-TLX), provide a systematic framework for assessing asynchronous XR systems, and directly inform **RQ2**.

Together, these perspectives form the conceptual basis for defining the Asynchronous Collaboration in Immersive Environmentsk (ACIE) framework and evaluating its performance in asynchronous engineering contexts.

In Simpler Terms

This chapter laid out the key ideas needed to understand what makes asynchronous collaboration in XR unique and challenging. It explained how people collaborate across time, what XR can (and can not) do to support that, and how to evaluate such systems from a user-centered perspective. These insights serve as a compass for the rest of the work.

How to Proceed

The conceptual groundwork laid out here directly informs the following chapters:

➋ **Chapter 3** reviews the current state of XR-based asynchronous collaboration systems – identifying limitations in how existing tools address the theoretical dimensions discussed here.

➤ **Chapter 4** builds on these foundations to introduce the ACIE framework, a novel concept aimed at addressing the specific challenges identified in both theory and practice.

3 State of the Art

This chapter positions the thesis within the broader research landscape by examining prior work on asynchronous collaboration in complex domains supported by immersive technologies. The analysis builds on the theoretical foundations established in ❯ **Chapter2** and identifies limitations and open questions that motivate the proposed approach. In doing so, it informs the design space of the ACIE Framework and refines the research questions introduced in ❯ **Chapter1**.

To capture foundational developments together with recent advances, a two–stage review design was employed (Figure 3.1). In Stage A, prior review articles on XR collaboration were synthesized to establish terminology, delimit the scope of asynchronous collaboration (AC), and surface initial gaps (❯ Section 3.1). Stage B was informed by Stage A and was conducted as a PRISMA-guided Structured Literature Review (SLR) targeting primary studies that explicitly address AC in XR (❯ Section 3.2), complemented by snowballing to include highly relevant works not retrieved by database queries. The resulting corpora were merged and coded along the 3C model and complementary dimensions, yielding a unified synthesis (❯ Section 3.3) that informs the subsequent gap analysis (❯ Section 3.4).

3.1 Findings from Prior Reviews

A set of key review articles was analyzed to assess the current research landscape on collaboration in XR environments, with a particular focus on asynchronous

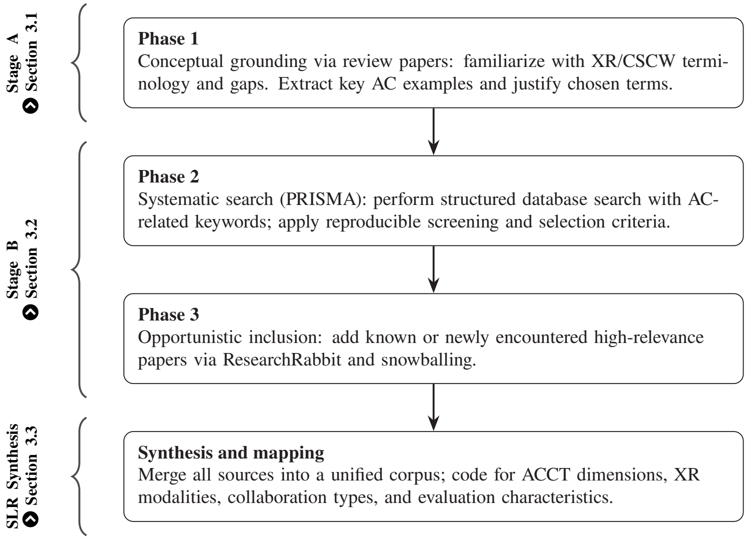


Figure 3.1: Literature-search workflow: from conceptual grounding to systematic search, opportunistic expansion, and unified synthesis. Stages are detailed in Section 3.1, Section 3.2, and Section 3.3.

collaboration in XR. The literature-search workflow followed a structured multi-phase approach, summarized in Figure 3.1.

A total of $n = 21$ review papers were initially identified on the broader topic *"XR collaboration"* searching the *Scopus* and *Web of Science (WoS)* digital libraries. Among those, $n = 5$ were duplicates, $n = 9$ did not address AC and $n = 1$ explicitly excluded AC from its scope. This left $n = 6$ review papers for detailed synthesis which are listed in Table 3.1. The selected reviews span research from 1995 to 2019 and collectively identify $n = 28$ papers focused on AC in XR.

The reviews primarily address collaboration in MR environments, often encompassing both AR and VR. Specifically, four reviews address MR collaboration more broadly [38, 48, 97, 126], while two reviews focus exclusively on AR

Table 3.1: Overview of review papers on XR collaboration, including the number of articles addressing asynchronous collaboration (AC)

Review	**XR dimension**	**Time range**	**Reviewed papers**	**AC papers**
Ens et al. [48]	MR	1995–2018	110	4 (5%)
de Belen et al. [38]	MR	2013–2018	259	*NA*
Pidel and Ackermann [126]	MR	*NA*	*NA*	10
Letter et al. [97]	MR	1998–2021	80	6 (8%)
Sereno et al. [139]	AR	1996–2019	65	4 (6%)
Feng et al. [53]	AR	2012–2022	133	7 (5%)

NA: Information not reported in the source article.

[53, 139]. The databases accessed across the reviews included *WoS*, *Springer Link*, *ACM Digital Library*, *IEEE Xplore*, *Elsevier ScienceDirect*, *ASME Digital Collection*, *Wiley Online Library*, and *Scopus*.

All six reviews unanimously highlight that asynchronous collaboration in XR is an underexplored yet highly promising research area, with growing scholarly attention [53]. Asynchronous collaboration currently represents a small fraction (5% – 8%) of the reviewed studies. Notably, Letter et al. [97] emphasize that the research gap is not primarily technical – it is not a matter of unavailable technologies or interface limitations – but rather reflects a lack of conceptual and empirical work specifically investigating AC in XR. They further highlight that although asynchronous collaboration in XR remains underexplored, several existing technical approaches could support it but not explicitly named as *"asynchronous collaboration"*. Table 3.2 summarizes a set of existing technical approaches that support asynchronous collaboration in XR, as identified by Letter et al. [97] and related works. These methods demonstrate that foundational mechanisms for deferred interaction already exist and could be adapted or extended within immersive systems.

Table 3.2: Existing technical approaches supporting asynchronous collaboration in XR (adapted from Letter et al. [97])

Reference	Approach
Boboc et al. [16]	Documentation of physical work steps
Feld and Weyers [51]	Trace-based steps in TUI environments
Ludwig et al. [101]	Tangible interfaces with digital access
Kim et al. [86]	Spatial capture of MR telepresence systems
Jones et al. [83]	Asynchronous audio/video conferencing practices
Perret and Vander Poorten [124]	Use of commonly available physical proxies

Across the identified reviews, a consistent set of benefits and opportunities associated with asynchronous collaboration in XR emerges. These can be categorized into three major themes:

1. **Creation and Retention of Information:** A primary advantage of AC in XR is the ability to create and retain digital information over time, allowing collaborators to engage asynchronously. Users can leave annotations, spatial tags or full recordings that are persistently accessible, facilitating knowledge retention and continuous project documentation [38, 48, 53, 126, 139]. This is particularly beneficial in domains like AECO, industrial maintenance (shift work), and crime scene investigation, where different experts contribute at different times [38]. Moreover, systems could benefit from incorporating mechanisms akin to GitHub-style change tracking for visualizing asynchronous updates to virtual environments [126].

2. **Collaboration on 3D and Spatial Information:** Asynchronous XR systems uniquely support collaboration anchored in 3D space. Through place-based annotations, geolocalized tagging, and shared 3D artifacts, collaborators can engage with spatial information even when they are not co-present [38, 48, 53, 126, 139]. Sereno et al. [139] notably emphasize the use of physical 3D tokens to facilitate quick and accurate navigation

to spatial annotations, highlighting the richness of embodied, spatially-grounded asynchronous interactions.

3. **Continuous Work Despite Absence of Collaborators:** A defining characteristic of asynchronous collaboration in XR is enabling workflow continuity across time. Rather than requiring simultaneous presence, collaborators can contribute sequentially, each building on prior inputs [38, 48, 53, 126]. This model is particularly useful for distributed or shift-based teams and supports longer-term, iterative project work. Ens et al. [48] further argue that future systems should not treat synchronous and asynchronous collaboration as exclusive concepts, but should instead support smooth transitions between them.

Overall, asynchronous collaboration in XR remains a nascent but increasingly relevant field. The reviewed literature shows a broad consensus that AC in XR offers distinct benefits for information creation and retention, spatial collaboration and workflow flexibility – but that existing systems only begin to explore these possibilities. Emerging challenges include the need for better visualization of asynchronous changes, capture of non-verbal communication cues in deferred collaboration and design methods for fluid transitions between synchronous and asynchronous modes [53, 97]. Addressing these gaps presents a promising frontier for advancing collaborative XR technologies, to which this work contributes.

3.2 Structured Literature Review

While prior reviews cover literature published between 1995 and 2019, this structured review update extends the scope to include relevant publications from 2020 to 2025.

Based on these insights, a search strategy was developed for a systematic literature review, following principles adapted from the PRISMA framework[122].

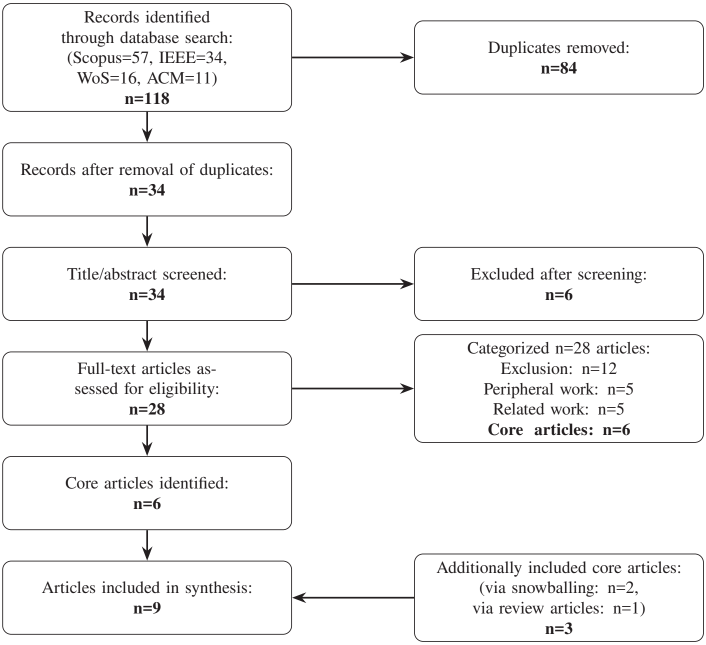


Figure 3.2: PRISMA-based flow diagram of the structured literature review process for identifying relevant articles on asynchronous collaboration in XR [122]

The Boolean search expression was designed to capture relevant studies explicitly addressing asynchronous collaboration in XR contexts. The search was applied to the *title*, *abstract* and *author keywords* fields, and included filters for English-language, peer-reviewed journal and conference publications from 2020 to 2025. Keywords and composition are listed in ➜ **Appendix A.1**.

The following academic databases were queried: IEEE Xplore, ACM Digital Library, Web of Science and Scopus.

> **Search Expression:** (“asynchronous collaboration”) AND (“XR” OR “VR” OR “AR” OR “MR” OR “Extended Reality” OR “Virtual Reality” OR “Augmented Reality” OR “Mixed Reality”)

This search yielded a total of $n = 118$ publications, which serve as the basis for the systematic review presented in this work. The distribution of the search results across databases is shown in Table 3.3, while Figure 3.2 depicts the resulting PRISMA flow diagram. In addition to the structured database search, a second source of literature was incorporated, based on the previously analyzed review articles. These reviews yielded $n = 28$ potentially relevant papers. After applying the same screening criteria used in the structured search –excluding single-user systems ($n = 5$), non-immersive or 2D-only implementations ($n = 12$) and one conceptual paper ($n = 1$) – a total of $n = 10$ review-derived papers were retained and added to the article pool.

Table 3.3: Number of papers retrieved per database using the defined search expression

Database	**Number of Results**
Web of Science (WoS)	16
ACM Digital Library	11
IEEE Xplore	34
Scopus	57
Total	**118**

After 84 duplicates were removed, the initial screening resulted in 28 included articles. Articles that were off-topic, lacked empirical results or presented only conceptual discussions were excluded, as well as articles that were preliminary work of other included articles.

All screened articles were categorized into four groups – Excluded, Peripheral, Related, and Core – that are listed in Table 3.4. This classification reflects each work’s alignment with the analytical framework guiding this study, which emphasizes embodied interaction, spatial and cognitive task complexity, asynchronous communication, coordination tools and cooperative mechanisms.

Table 3.4: Categorization of screened articles by origin (review articles, structured literature review, and snowballing), grouped into excluded, peripheral, related, and core categories

Articles Source	**Excluded**	**Peripheral**	**Related**	**Core**
AC Review Articles	[9, 14, 31, 54, 57, 68, 140, 142, 155]	[17, 25, 58, 77, 84, 102, 103, 114, 125, 130, 134, 141, 167]	[66, 76, 90, 115, 166]	[156]
$\sum_{reviews}$	9	13	5	1
Structured Literature Review	[6, 10, 17, 31, 32, 33, 41, 106, 107, 117, 118, 152]	[29, 59, 61, 81, 157]	[23, 96, 104, 133, 150]	[34, 52, 63, 154, 171, 172]
$\sum_{SLR}$	12	5	5	6
Snowballing	–	–	[120]	[36, 159]
$\sum_{snowballing}$	0	0	1	2
$\sum_{total}$	**21**	**18**	**11**	**9**

Excluded works failed to meet basic relevance criteria. These include single-user systems, non-immersive or 2D-only implementations, and conceptual proposals without implementation or evaluation. While some addressed adjacent themes such as spatial annotation, they lacked focus on temporally decoupled, embodied multi-user collaboration.

Peripheral works engage with the idea of asynchronous collaboration in XR but fall short in essential dimensions. Common limitations include minimal embodiment, static content, low-immersion AR settings or one-way interaction mechanisms. Many also lack empirical grounding. Though broadly relevant, they were excluded from the synthesis due to insufficient support for interactive or immersive depth.

Related works align more closely with the review's goals, often exploring individual aspects like version control or shared scene navigation. However, they are typically narrow in scope or passive in interaction, lacking features such as coordinated workflows or reactive communication. These systems offer partial support for asynchronous XR, but not enough to qualify as core contributions.

Core contributions most directly address the research questions of this thesis. These systems combine asynchronous communication with spatial interaction and coordination mechanisms, and crucially, are implemented and validated. They demonstrate mature support for embodied, multi-user collaboration across time, forming the empirical foundation of the synthesis.

This structured categorization ensures that only systems with significant, demonstrable support for asynchronous, embodied collaboration in XR are considered in the synthesis. In addition to the structured database search, further candidate papers were identified through prior review articles. A total of $n = 28$ potentially relevant works were extracted. After applying the same screening criteria used in the SLR – excluding single-user systems ($n = 5$), non-immersive or 2D-only implementations ($n = 12$) and one conceptual paper ($n = 1$)-$n = 10$ review – derived papers were retained. These were integrated into the overall categorization and synthesis presented in this chapter.

Each system was evaluated across six categories central to asynchronous collaboration in XR, using a 0–3 scale as detailed in Table A.1 (❯ **Appendix A**). These dimensions reflect the extent to which a system supports immersive, coordinated, and cooperative interaction across time:

- **Embodiment:** refers to how effectively a system supports embodied communication, such as avatar presence, gesture expression or full-body representation. Most systems only partially support embodiment, often limited to head and hand tracking or ghost replays. Higher scores reflect more expressive, full-body capture or situated avatar interaction.

- **Task Complexity:** captures the cognitive, spatial, or physical sophistication of supported tasks. While many systems address moderate complexity, such as design review or procedural walkthroughs, few cover all dimensions – especially physical or tool-based interaction – in a structured manner.
- **Communication:** focuses on the degree of asynchronous interaction supported between users. Most systems offer unidirectional communication through playback of recorded messages or demonstrations. Higher scores reflect reactive, bidirectional or multi-modal communication – such as voice, gesture, or annotations – that support dialogic interaction over time.
- **Coordination:** refers to how a system helps users manage task flow across sessions. This includes timeline-based navigation, branching structures, checkpoints, and task versioning. While timelines are common, few systems support structured task management or transitions between synchronous and asynchronous states.
- **Cooperation:** measures how users can build on one another's work asynchronously. Many systems only support sequential handovers, with limited mutual awareness or interdependent task evolution. Systems scoring higher facilitate shared editing, response chaining or the persistence of interactive contributions across users and time.

Together, these dimensions highlight key functional areas for enabling rich, embodied asynchronous collaboration in XR-and also the gaps that current systems often fail to bridge.

In addition to the core systems discussed above, several further papers were identified through backward and forward snowballing. Among these, Holoportation [120] is categorized as related, as it presents a technically advanced platform for real-time volumetric telepresence and mentions potential for playback-based experiences. However, it does not implement or evaluate asynchronous collaboration (AC) features such as message persistence, user response mechanisms or interaction across time, which precludes its classification as core. Conversely, AgainTogether [159] is included in the core set due to its explicit support for asynchronous spatial communication, combining volumetric capture with embodied replays. It addresses several elements of the 3C model and demonstrates temporal decoupling between users, making it directly relevant to the study of AC in XR.

3.3 SLR Synthesis

A six–dimension schema was applied, grounded in the 3C model (communication, coordination, cooperation) and complemented by embodiment, task complexity, and XR continuum placement. The synthesis is presented by dimension; the coverage of core systems is summarized in Table 3.5.

Table 3.5: Conceptual mapping of representative XR approaches for ACCT across five dimensions: embodiment, task complexity, communication, coordination, and cooperation

Related Work	Embodiment	Task Complexity	Communication	Coordination	Cooperation	Score
Guided Tour [63]	●○○	●●○	●○○	●○○	○○○	5
Time Travellers [34]	●○○	●●●	●●○	●○○	●●○	9
MAVRC [36]	●○○	●●●	●●●	●○○	●○○	9
ReliveInVR [159]	●●●	●●○	●○○	●●○	●●○	10
XR-LIVE [154]	●●○	●●●	●○○	●●○	●●○	10
VRGit [172]	●●○	●●●	●○○	●●●	●●○	11
Virtual Triplets [171]	●●○	●●●	●●○	●●○	●●○	11
Loki [156]	●●●	●●●	●●○	●○○	●●○	11
Async Reality [52]	●●●	●●●	●●●	●●○	●●○	13

Note: Each criterion is scored on a scale from 0 to 3 and visually encoded using circle symbols. An empty circle (○) indicates no or negligible support, while up to three filled circles (●) reflect increasing levels of functionality or completeness.

3.3.1 Embodiment

Embodiment across reviewed XR asynchronous collaboration systems varies widely in fidelity and expressiveness. Many systems rely on minimal representations, often limited to head and hand tracking or ghost avatars, as seen in Time Travellers and MAVRC. These offer spatial awareness but lack expressive bodily interaction. XR-LIVE, Guided Tour, and Virtual Triplets employ simplified avatars or virtual guides, supporting presence but not interactional nuance.

Higher fidelity embodiment is realized in AsyncReality and ReliveInVR through full-body volumetric capture, enabling users to replay rich, situated performances. Loki similarly supports embodied recordings, including gestures and movement, making messages more expressive and situated. However, only AsyncReality directly anchors the embodiment in the physical context with temporal replay triggers. VRGit, despite excelling in task-state management, offers only basic avatar presence, illustrating a trade-off between interaction control and expressive embodiment.

This landscape highlights an ongoing gap: few systems enable expressive, multimodal embodiment that supports asynchronous social cues, spatial referencing, or embodied feedback. Advancing this dimension will be key to fostering natural, immersive asynchronous interactions.

3.3.2 Task Complexity

Most systems support spatial and cognitive tasks of moderate complexity, such as navigation, design review, or procedural instruction. Time Travellers, Guided Tour, and XR-LIVE focus on step-based or annotated task walkthroughs, often with predetermined flows. Virtual Triplets and MAVRC allow for more open-ended exploration or message placement but do not include dynamic task branching or adaptation.

Systems like VRGit and AsyncReality demonstrate higher complexity by enabling branching contributions (VRGit) or contextual replay tied to user behavior (AsyncReality), supporting reflective or iterative workflows. AsyncReality introduces a unique causality graph linking tasks to physical space and temporal dependencies, raising the complexity by linking actions to situated contexts.

Overall, physical interaction remains rare – most tasks rely on spatial cognition and narrative reasoning rather than object manipulation or real-time technical operation. This suggests an opportunity to integrate asynchronous workflows into more hands-on or tool-supported XR domains.

3.3.3 Communication

Communication is the most commonly supported aspect of asynchronous XR collaboration, as nearly all reviewed systems offer some form of recorded message or demonstration for later playback. This typically takes the form of unidirectional communication from one user to another – such as experts recording guidance or instructors presenting lectures – which others can view asynchronously. For example, Guided Tour and XR-LIVE provide recorded demonstrations or avatar-based lectures that users can revisit at any time, often with spatially situated audio and visual cues. These systems create a persistent communication channel but do not allow recipients to respond or contribute within the same medium.

Some systems move beyond passive consumption by allowing users to contribute their own recordings or scene modifications. MAVRC expands on this model by supporting multi-modal annotations – voice, hand gestures, and text – anchored to specific viewpoints in 3D space. This enables richer expression and context-aware messaging, although interaction remains unidirectional. Time Travellers supports annotations added during playback, offering a simple form of user-to-user messaging through spatial tags. AsyncReality introduces embodied recordings triggered by spatial context, adding a reactive dimension to communication. Although still lacking direct bidirectional exchange, these embodied messages are not only spatially anchored, but also function as context-specific responses to prior actions. Loki further enriches the medium by capturing full-body volumetric recordings combined with audio, enabling expressive and immersive messages. However, it lacks persistent threads or mechanisms to organize and navigate asynchronous conversations.

Virtual Triplets diverges from the broadcast model by supporting reactive question-answer interactions with a virtual agent. While not peer-to-peer, this introduces dynamic responsiveness into the communication flow. Yet across the reviewed systems, true bidirectional or threaded communication between asynchronous collaborators remains rare. Most tools do not allow for conversational loops, feedback, or message linking, limiting asynchronous communication to isolated contributions rather than collaborative dialogue.

In sum, while communication is the most consistently implemented 3C element, it often remains static, linear, and one-sided. Future systems should evolve toward more dialogic, multi-modal, and reciprocal communication structures that better support sustained asynchronous interaction.

3.3.4 Coordination

Coordination mechanisms vary significantly across the reviewed systems, with only a few offering structured temporal or task navigation tools. XR-LIVE stands out by providing auto-pausing playback and structured steps, simulating instructor pacing. Similarly, Time Travellers offers a timeline interface with playback control but lacks deeper features like checkpoints or branching. AsyncReality introduces a novel causality graph, enabling context-sensitive playback of recorded user actions based on logical dependencies – providing a temporal coordination layer distinct from traditional timelines.

The most sophisticated coordination support is seen in VRGit, which enables version branching, merging, and task reuse, echoing Git-style workflows for 3D content. Conversely, Guided Tour and Virtual Triplets provide only linear playback or step-wise flow with no control over task progression or state history. MAVRC allows users to revisit task stages through spatially anchored multimodal annotations. While these annotations provide implicit guidance and preserved viewpoints, MAVRC lacks explicit task progression tools such as checkpoints or version tracking. Loki similarly supports playback of spatial interactions without structured navigation or coordination layers.

ReliveInVR similarly supports timeline-based control of playback and allows users to asynchronously view and respond to recorded actions placed in the environment. While it does not provide explicit branching, checkpoints, or workflow management, the ability to layer temporally decoupled responses and spatially organize them contributes to a moderate degree of coordination support.

Overall, coordination remains underdeveloped in most systems, with only VRGit offering full asynchronous navigation, branching and workflow management.

3.3.5 Cooperation

Cooperation – understood here as the interaction and mutual influence between asynchronous users across time – is the least supported dimension in existing XR systems. Rather than enabling collaborators to react to, adapt or build upon each other's contributions in a meaningful way, most systems treat asynchronous engagement as a one-way or solo experience.

Guided Tour and XR-LIVE offer no cooperative mechanisms; users only consume pre-recorded content without contributing anything that might be received or responded to by others. Time Travellers and AsyncReality make progress by enabling task continuation – users can record actions that others later see and respond to – yet these responses do not feed back into a collaborative loop. There are no feedback channels, no shared checkpoints, and no awareness of a counterpart's presence or intent, which limits true cooperative interaction.

Virtual Triplets introduces a more reactive environment through agent-based dialogue, but this is limited to human-agent interaction rather than peer-to-peer cooperation. Loki supports rich embodied message capture that could serve as a cooperative gesture, but the lack of persistent task state, context-aware reactions or reply mechanisms keeps the cooperation one-sided. Loki and MAVRC allow users to leave contextualized messages and replays, but their contribution model is limited to sequential input without interaction chaining or collaborative construction.

MAVRC's annotations enable later users to understand prior input, yet the system lacks mechanisms for mutual adaptation, shared editing or structured cooperative workflows. As such, cooperation remains limited to temporal awareness rather than active co-construction. VRGit remains the only system to support structured, interdependent editing over time.

ReliveInVR contributes to cooperative asynchronous workflows by enabling users to observe and build upon each other's embodied messages. While there is no shared editing space or explicit branching structure, the system allows users to reply with new embodied recordings, creating a spatial and temporal chain of contributions. This supports a form of reactive cooperation through layered interactions in the shared environment, albeit without formal co-editing or feedback loops.

VRGit comes closest to enabling asynchronous cooperation by visualizing branching contributions and allowing users to merge or build upon others' edits. While it does not offer rich dialogic interaction, it facilitates mutual awareness and interdependent activity over time – a key step toward more structured cooperation.

In sum, most XR systems support only sequential handovers of activity. They lack the cooperative scaffolding – such as reply mechanisms, feedback loops or shared awareness – that would allow users to dynamically influence one another's work. Building systems that support reactive, reciprocal asynchronous collaboration remains a critical challenge and opportunity.

3.3.6 XR Continuum Analysis

Most reviewed systems support AC in immersive VR environments [36, 63, 63, 154, 159, 171, 172], while only a few extend these capabilities into MR contexts [34, 52, 156]. All reviewed systems span different points on the XR continuum, yet their degree of immersion and integration with the physical world varies significantly. Guided Tour, Virtual Triplets, and VRGit are VR-only systems that provide rich, immersive environments, but lack interaction with physical surroundings. XR-LIVE simulates educational environments within VR and supports spatial task flows but remains confined to the virtual domain without physical-world anchoring.

Time Travellers bridges VR and MR using HoloLens, enabling spatial anchoring and some real-world visibility; however, the interaction still centers on virtual annotations and playback. AsyncReality stands out for embedding asynchronous interaction directly into physical workspaces via MR overlays and spatially-triggered embodied recordings. It is the only system reviewed that tightly couples virtual communication with real-world task environments, addressing the needs of field-based or hands-on domains.

Loki demonstrates broad support across the XR continuum: it enables volumetric capture and playback in both VR and MR contexts, supporting hybrid scenarios where users can alternate between immersive and real-world settings. Its flexibility in deployment makes it particularly well-suited for bridging physical and virtual collaboration spaces.

This diversity reveals a key research opportunity: while immersive VR experiences are common, few systems support blended or context-aware XR scenarios that bridge physical and virtual actions in asynchronous workflows. Expanding support for MR use cases and hybrid transitions remains an open challenge in designing flexible, embodied asynchronous collaboration tools.

3.3.7 System Analysis

Existing research on asynchronous collaboration (AC) in extended reality (XR) environments is characterized by fragmented coverage across key dimensions of interaction and workflow management. Each system typically addresses distinct use-cases and provides partial coverage of the 3C model – communication, coordination, and cooperation – as summarized in Table 3.5.

For instance, Giovannelli et al. [63] proposed Guided Tour, a system that emphasizes spatial exploration and passive learning through recorded demonstrations, yet lacks mechanisms for interactive engagement, bidirectional communication, or cooperative task contribution. Cho et al. [34] introduced Time Travellers, which incorporates basic avatar embodiment and timeline-based playback for sequential task review. While it allows users to annotate recordings, the system does not support reactive messaging, synchronous playback control or mutual content modification.

Virtual Triplets [171] presents a more sophisticated interaction model through a virtual agent that engages users in reactive question–answer dialogue, coupled with full-scene embodiment and high task complexity. However, its communication remains system-mediated, and there is no support for co-editing or branching collaborative workflows.

VRGit [172] stands out in terms of coordination and cooperation, offering Git-like branching and merging for collaborative 3D design. Nevertheless, the system lacks asynchronous communication capabilities, such as messaging or dialog, and provides only basic embodiment. XR-LIVE [154] delivers a structured educational experience with spatial checkpoints and guided task flow, but communication is strictly unidirectional, and no peer interaction or feedback mechanism is present.

AsyncReality [52] integrates volumetric capture and real-world context through spatially-triggered recordings, enabling users to experience embodied playback. Yet, communication remains one-way, and there is no mechanism for joint content creation or reactive response. Loki [156] also uses volumetric capture and offers a more interactive communication model by supporting both live and recorded spatial interactions, combining audio and embodied actions. While it lacks persistent messaging or true asynchronous dialogue, its hybrid design supports partial temporal continuity and role asymmetry.

MAVRC [36] introduces a VR-based environment for asynchronous communication in collaborative spatial tasks, using multi-modal annotations including audio, text, and virtual gestures. It supports contextual replay within a shared virtual scene, providing users with a preserved perspective of the recorded content. Though embodiment is limited to head and hand presence, and cooperation mechanisms are basic, MAVRC excels in communication expressiveness and temporal task guidance through multi-modal cues.

ReliveInVR further contributes to the field by supporting spatially situated, asynchronous embodied messaging between users. It enables collaborators to respond to each other's recordings within the same immersive environment, supporting communication, coordination, and a limited form of cooperation. Despite lacking branching or task-state tracking, its support for contextual, reactive interaction represents a meaningful step toward dialogic asynchronous collaboration.

These systems collectively show that richer embodied or audio-based communication is possible in AC, yet truly dialogic and co-constructive interaction across sessions remains rare.

3.4 Research Gaps

Despite growing interest in asynchronous collaboration (AC) in XR, existing systems exhibit persistent limitations across the core dimensions of interaction – communication, coordination, and cooperation – and fall short of delivering a unified, embodied collaborative framework. While recent additions such as Loki show progress in hybrid embodiment and educational replay, most existing solutions still emphasize only one or two aspects of the 3C model, without offering a fully integrated framework that supports all three in tandem.

Fragmented Feature Support: Functionality remains fragmented and tailored to narrow use cases. Systems like Guided Tour and XR-LIVE provide rich one-way communication via expert demonstrations but offer no means for peers to respond, contribute or co-create. Others – such as Time Travellers, and AsyncReality – support limited reactive or spatial communication, such as annotations or playback of embodied recordings, yet lack feedback loops or dialogic structures. Loki advances this space with volumetric replay and spatial messaging, but lacks coordination mechanisms and persistent task or message tracking. VRGit excels in coordination and cooperation through version control, but does not support embodied or multi-modal communication. Virtual Triplets introduces agent-based Q&A, but remains confined to scripted instructional flows, lacking peer-driven cooperation.

Limited Bidirectional Communication: Most XR systems support only unidirectional or system-mediated messaging. Even in more advanced systems, such as Loki, asynchronous messages lack threading, response cycles, or message provenance. Rich, multi-modal, dialogic interaction across time – encompassing gesture, voice, spatial referencing, and replyable message history – remains essentially absent across current systems.

Constrained Coordination Models: Timeline interfaces dominate the current coordination landscape, with few systems supporting structured navigation or task state awareness. VRGit remains the only example with non-linear branching and merging, while others rely on linear playback or implicit handovers. Loki offers basic timeline playback, but no structuring mechanisms. No system supports transitions between synchronous and asynchronous modes – a major gap for hybrid workflows.

Underdeveloped Cooperation: Cooperation – understood as meaningful interaction between asynchronous users – is still rare. Most systems only allow sequential handovers without true feedback loops or shared spaces. While VR-Git supports persistent artifacts and branching, it lacks communicative integration. Time Travellers, AsyncReality, and Loki allow users to view or respond to prior actions, but do not support mutual influence, adaptive task states or shared progress management. Real cooperative editing, response chaining, or joint progress tracking is effectively absent.

Lack of Awareness and Provenance: Users in most systems remain unaware of who contributed what, when or why. Few tools provide any visual or semantic trace of interaction history. Even in Loki or Time Travellers, where users view prior activities, there is no way to annotate, thread or contextualize them. Only VRGit offers a structured task evolution model, but it is not linked to embodied or spatially anchored actions.

Minimal Integration with Real-World Context: Most systems operate exclusively in virtual domains. Loki supports VR and MR context, but does not deeply embed interaction in physical workflows. Only AsyncReality anchors embodied communication directly in the user's environment. Bridging embodied XR communication with physical tools and real-world workflows remains largely unexplored – despite its critical role in practical, domain-specific applications such as engineering, fieldwork or healthcare.

While some progress is evident – especially in systems like Loki (for hybrid embodiment) and VRGit (for structured task evolution) – no existing solution provides fully integrated, multi-modal asynchronous collaboration with dialogic communication, context-aware coordination, and mutual cooperation. Most systems still rely on static handovers, one-way playback and passive observation, offering little support for truly collaborative, temporally distributed workflows. Future systems must address this fragmentation by supporting persistent, embodied and co-constructive collaboration – across time, space and modality.

3.5 Summary and Outlook

This chapter reviewed the current state of asynchronous collaboration in XR, combining insights from prior review literature and a structured literature review. Through a two-stage process – synthesizing existing XR collaboration reviews (1995–2019) and conducting a PRISMA-based structured review (2020–2025) – nine core systems were identified and evaluated across six key dimensions: embodiment, task complexity, content richness, communication, coordination and cooperation.

The analysis revealed several critical limitations in the current landscape:

- **Fragmented Feature Support**: Most systems focus on isolated aspects – communication, coordination or cooperation – rather than addressing them in an integrated, holistic manner.
- **Limited Bidirectional Communication**: Many solutions rely on one-way messaging (e.g. instructional replay), lacking support for dialogic interaction, feedback loops or follow-up mechanisms.
- **Constrained Coordination Models**: Task flows are typically linear and fixed. Only VRGit introduces branching and merging (akin to version control), but without immersive integration.
- **Underdeveloped Cooperation**: Existing workflows mostly follow a sequential handover model, with minimal support for meaningful interaction or co-construction across time.

- **Lack of Awareness and Provenance**: Few systems capture and communicate who did what, when, or why. Embodied traceability and contextual referencing are often absent.
- **Minimal Real-World Anchoring**: Many systems operate solely in virtual contexts and do not incorporate physical workflows or real-world spatial alignment.

While some platforms – such as VRGit, Loki, and AsyncReality – propose partial solutions, none of the reviewed systems provide a unified framework capable of supporting embodied, multi-modal and interactive asynchronous collaboration across the full XR spectrum.

In Simpler Terms

This chapter showed that while there are promising XR tools for working across time, most of them do only part of the job. Some let users leave behind instructions, but do not allow others to respond easily. Others help track versions of work, but do not show what people actually did in 3D space. What is missing is a complete system – one that lets collaborators leave understandable, spatially anchored messages and gives others the tools to respond, adapt, and build on that work as if they were present.

How to Proceed

The limitations identified in this chapter define the conceptual gap addressed in the next part of this thesis:

➜ **Chapter 4** introduces the ACIE framework – a modular concept for asynchronous collaboration in XR – developed to address the gaps identified in the literature and to support complex engineering workflows across time and space.

❯ **Chapter 5** describes how the ACIE framework was implemented across XR platforms, with system components realized for asynchronous communication, coordination and cooperation.

4 Conceptual Design for Asynchronous Collaboration in XR

Based on the theoretical foundations established in ❯ **Chapter 2**, this chapter introduces a concept for AC in XR environments. The proposed framework integrates methods that enable the three C's of collaboration – communication, cooperation, and coordination – to occur asynchronously, supporting seamless interaction across time.

In complex tasks with strong spatial and physical components – such as those encountered in engineering – XR technologies offer a powerful technical foundation for enabling AC. As discussed in ❯ **Chapter 3**, recent research has begun to explore XR's potential for asynchronous collaboration, yet most approaches address only isolated aspects: some support asynchronous messaging without coordination; others implement versioning without embodied interaction. Comprehensive, integrative solutions remain rare.

To address this gap, this chapter presents ACIE – a holistic framework for asynchronous collaboration in immersive environments such as XR. It builds upon and extends prior work to meet the demands of spatial and procedural engineering tasks.

The remainder of this chapter is structured as follows: ❯ Section 4.1 introduces the overall ACIE framework. ❯ Section 4.2, ❯ Section 4.3, and ❯ Section 4.4 detail the three core modules: asynchronous communication, cooperation, and coordination, respectively. Finally, Section 4.5 discusses how the framework supports asymmetric collaboration across XR modalities.

4.1 ACIE Framework

In this work, it is investigated how XR technologies can support asynchronous collaboration in complex engineering tasks – such as remote maintenance, assembly training, or procedural reviews. These tasks are inherently spatial, procedural, and embodied – dimensions that are insufficiently supported by traditional media such as text, static images or linear video.

By embedding digital content into physical environments, XR enables novel forms of visualization and interaction with both past and ongoing processes. Beyond improving spatial perception, XR affords persistent, embodied representations of collaboration – allowing users to observe predecessors, communicate with successors and contribute asynchronously. This positions XR as a uniquely powerful enabler of temporally decoupled collaboration.

To operationalize this potential, the ACIE framework introduces three interrelated modules that support asynchronous communication, cooperation and coordination in immersive, physically anchored environments as depicted in Figure 4.1. These modules address research questions **RQ1.1**, **RQ1.2**, **RQ1.3** which are summarized in Figure 4.2.

Asynchronous communication (**RQ1.1**) requires spatially and temporally embedded information to be exchanged effectively – without necessitating co-presence.

To support this, ACIE introduces *Immersive Messages*, which capture users' spatial actions and voice as multi-modal recordings, and *Proxy Playback*, which replays these messages in situ using embodied avatars. The *Immersive Communication Protocol (ICP)* defines a structured data format for motion and context information, ensuring consistent representation and replay across systems.

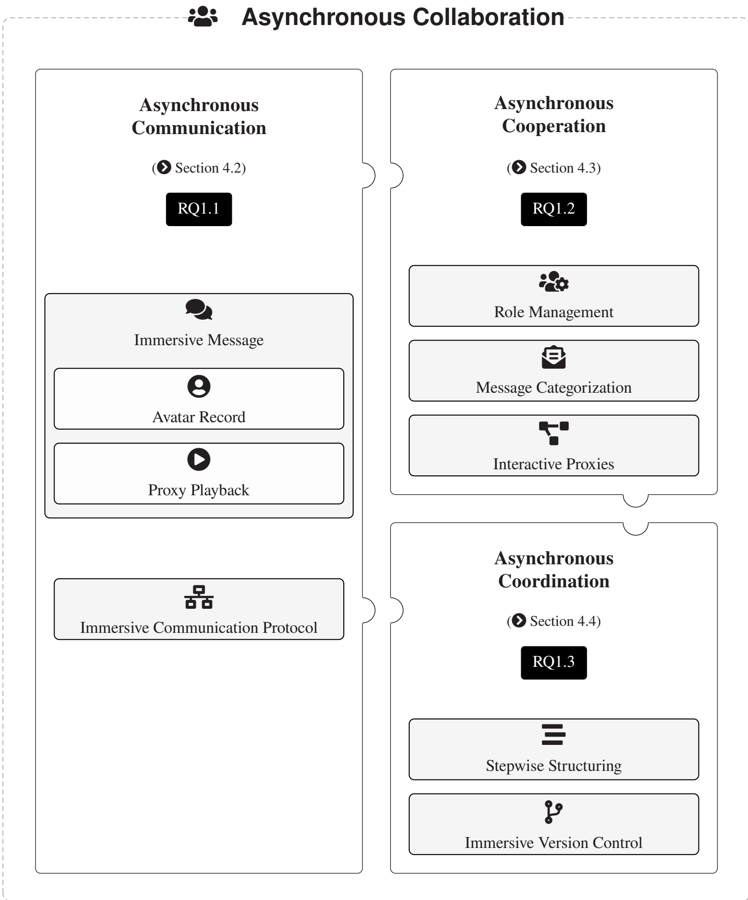


Figure 4.1: Conceptual model of asynchronous collaboration structured into communication, cooperation, and coordination mechanisms, with related research questions and corresponding sections.

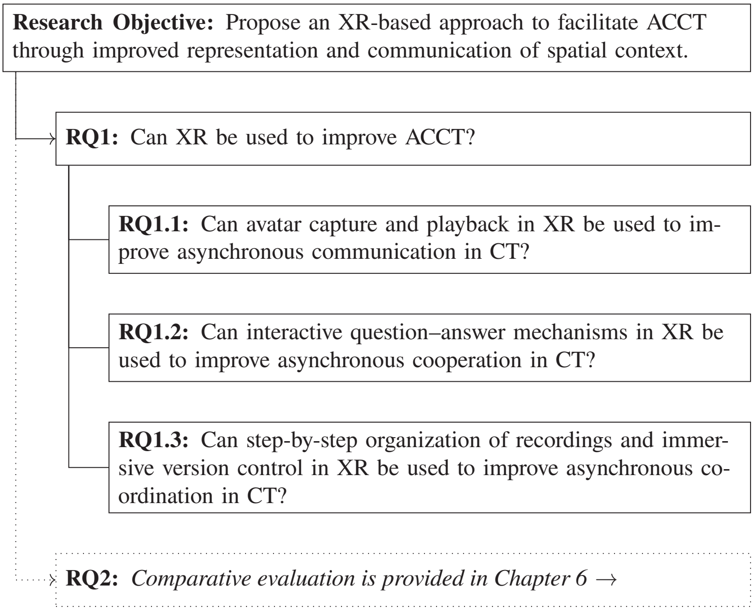


Figure 4.2: Decomposition of **RQ1** into sub-questions (communication, cooperation, coordination).

These mechanisms establish a persistent, embodied communication channel across time.

Asynchronous cooperation (**RQ1.2**) requires mechanisms for building upon, adapting to and responding to others' contributions across time.

To this end, ACIE incorporates *Message Categorization* (e.g. directive vs. reactive content), *Role Management* (e.g. expert vs. novice perspectives) and *Interactive Proxies*, which extend passive avatars into embodied agents that can be queried or responded to interactively. These components support temporally decoupled yet responsive interaction between collaborators.

Asynchronous coordination (**RQ1.3**) involves managing the temporal and procedural integration of contributions within a shared spatial context.

ACIE addresses this through *Stepwise Structuring* of Immersive Messages, which enables segmented replay and navigation, and through an *Immersive Version Control System (iVCS)*. Inspired by Git, the iVCS records, exchanges, and merges 3D workspace states across collaborators – supporting versioning, conflict resolution, and task alignment in immersive workflows.

Analysis of related work has shown that existing systems tend to address only isolated aspects of asynchronous collaboration. Some focus on immersive messaging without structuring or interaction [36, 159], while others emphasize spatial versioning without embodied communication [172]. None of the reviewed solutions provide a comprehensive or integrative framework.

The ACIE concept directly addresses this gap by unifying immersive communication, interactive cooperation and spatial coordination within a single framework[1].

The following sections elaborate on the three modules of ACIE: asynchronous communication (Section 4.2), cooperation (Section 4.3) and coordination (Section 4.4).

4.2 Asynchronous Communication

Effective asynchronous collaboration requires more than overlapping availability – it necessitates a robust infrastructure for transmitting information and work progress across time. This section introduces two central components that support immersive asynchronous communication: *Immersive Messages* and the *Immersive Communication Protocol (ICP)*. Together, these elements enable seamless, temporally distributed communication in support of complex, spatially embedded tasks.

[1] This concept, presented under the title **EduAvatars**, was recognized with the Best Concept Award 2024 by the German Institut für Virtual Reality (DIVR): https://divr.de/award/rueckblick-xr-science-award-2024/

4.2.1 Immersive Messages

Immersive Messages form the core communication mechanism in ACIE. They enable temporally displaced interaction by capturing and transmitting spatially embedded, embodied activity through XR-based recording techniques. Two primary mechanisms define this process: *Avatar Recording* and *Proxy Playback*. Figure 4.3 illustrates this asynchronous communication flow, beginning with the recording of spatial activity and voice, followed by the serialization and transmission of the Immersive Message, and culminating in the contextual replay of the contributor's behavior through a proxy avatar in the recipient's environment.

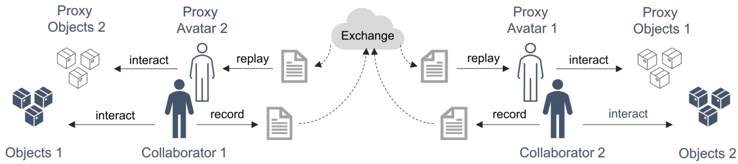


Figure 4.3: Workflow of an Immersive Message Exchange in ACIE: (1) Avatar Recording captures a collaborator's spatial motion and voice; (2) the data are serialized into an Immersive Message file and exchanged between the collaborators; (3) on reception, Proxy Playback re-embodies the sender in the recipient's scene, enabling spatially aligned, temporarily shifted collaboration.

An Immersive Message is created by capturing a collaborator's spatial tracking data relative to the task environment – including body movement, verbal utterances, non-verbal gestures, and object interactions – and structuring this data into replayable formats. This process, referred to here as *Avatar Recording*, is conceptually similar to motion capture but leverages native XR device sensors instead of external tracking systems or specialized suits. The recorded data is stored in serialized files and distributed to collaborators for asynchronous playback.

To deliver the message, Immersive Message files are imported into the recipient's immersive environment and replayed through *Proxy Avatars*. These avatars embody the original contributor's recorded behavior and allow recipients to observe the message in context while retaining full interactivity with their surroundings. This playback mechanism – termed *Proxy Playback* – supports intuitive, spatially aligned communication across sessions.

Immersive Messages thus enable collaborators to share instructions, feedback and observations across time and space. The following subsections describe the two underlying mechanisms in detail.

Avatar Recording

In immersive environments, users are represented by avatars that convey embodiment and interaction capabilities. To facilitate asynchronous communication, these actions must be recorded and structured for replay. In ACIE, this is achieved via Avatar Recording – a multi-modal process capturing both spatial motion and voice.

The recording consists of two synchronized streams:

- **Motion Stream (Frame-Based Capture):** Continuous recording of head and hand poses (6DoF), gestures and object interactions. This data is serialized into timestamped frames, forming a temporally structured motion sequence for precise spatial replay. Early work such as V-Mail [75] and MASSIVE-3 [65] demonstrated this approach, and recent extensions have incorporated embodied task execution and richer semantics [52, 156].
- **Voice Stream (Audio Recording):** Parallel to motion, speech is captured as an audio stream to convey intent, instruction or commentary. Synchronized voice enhances comprehension, providing semantic context alongside visual behavior. This principle is supported by systems such as that by Chow et al. [36], which demonstrated the value of integrating multi-modal feedback.

Proxy Playback

Proxy Playback is the mechanism by which Immersive Messages are delivered asynchronously. It enables collaborators to experience the actions and voice of a previous contributor through a proxy avatar within the immersive workspace.

A proxy avatar serves as a temporally displaced representation of a collaborator, visualizing their recorded actions – including body movements, object interactions and speech – in the shared spatial environment. Playback is fully embedded within the recipient's environment, allowing them to interpret prior contributions from their own perspective and engage with ongoing task elements.

Unlike traditional asynchronous media such as video or annotated text, proxy playback is spatially and contextually anchored. This allows users to grasp not just what was done, but how and where it occurred. The recipient remains in control of the environment and can pause, reposition or revisit recorded content as needed.

> **Ghost Metaphor:** Proxy avatars may be visualized with semi-transparent or ethereal effects to signal their asynchronous, non-live nature. This visualization supports intuitive awareness of temporal displacement and was successfully used in prior XR training systems [170].

By delivering spatially embedded, multi-modal communication, Proxy Playback forms the immersive layer for asynchronous collaboration. Its extension into interactive and responsive messaging is elaborated in Section 4.3.

While previous work (Section 3.2) has demonstrated components of motion capture and replay, ACIE integrates them into a cohesive communication model – directly linking them to asynchronous cooperation and coordination (Sections 4.3 and 4.4).

4.2.2 Immersive Communication Protocol

To ensure consistent asynchronous communication, ACIE introduces the *Immersive Communication Protocol (ICP)* – a specification for encoding and interpreting embodied interactions across time. ICP defines how motion and interaction data are structured, stored and replayed in a platform-independent format.

The protocol, as depicted in Figure 4.4, is responsible for serializing spatial behaviors, such as gesture, movement and object manipulation, into structured message objects. While voice data is recorded separately in standard formats (e.g. WAV), ICP ensures synchronized playback by linking audio and motion streams.

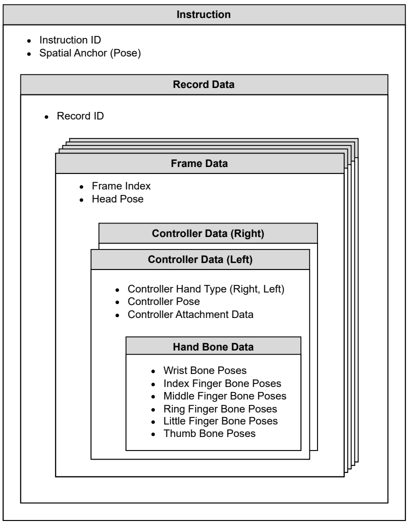


Figure 4.4: Record data structure (ICP implementation).

Structured Motion Recording

ICP structures user activity into a sequence of time-indexed frames. Each frame captures a snapshot of the user's state and interactions at a specific time. An overview is provided in Table 4.1.

Table 4.1: Key components of a frame in the Immersive Communication Protocol.

Component	Description
Frame Index	Sequential identifier denoting the frame's position in time.
Head Pose	6 DoF head position and orientation, indicating gaze and viewpoint.
Controller Data	Pose and orientation of each controller; includes object attachment metadata.
Hand Bone Data	Joint-level tracking of fingers and wrists, enabling expressive replay.
Object Interaction	Interaction state with referenced objects, including IDs and transformation data.

These components provide the structural foundation for immersive messaging – enabling temporally accurate playback, referencing of interaction events and integration with cooperative or coordinated workflows. The separation of motion and voice data ensures modularity while allowing synchronized, multi-modal replay.

This structured messaging architecture supports asynchronous communication in ACIE, addressing the following sub-question:

> ❓ **RQ1.1:** Can avatar capture and playback in XR be used to improve asynchronous communication in CT?

4.3 Asynchronous Cooperation

While asynchronous communication allows collaborators to exchange information across time, effective cooperation in complex tasks requires more than messaging. It involves engaging with the contributions of non-present collaborators, adapting to their inputs and building upon their work toward a shared goal. Unlike synchronous cooperation, where participants interact in real time and can respond immediately, asynchronous cooperation must occur without co-presence or instant feedback. This temporal separation makes the coordination of interactive contributions particularly challenging.

To address these challenges, ACIE employs three interrelated mechanisms:

- **Role Management** defines interaction privileges and responsibilities based on participant roles;
- **Message Categorization** structures collaboration into directive and reactive exchanges;
- **Interactive Proxies** extend message replay with embodied, reactive interaction interfaces.

Together, these components compensate for the lack of co-presence by enabling deferred responsiveness, turn-based engagement and contextual feedback.

4.3.1 Role Management

In cooperative workflows, roles define the nature and boundaries of participant interactions. While many systems apply uniform roles, complex engineering scenarios often exhibit asymmetrical knowledge distribution. A typical configuration involves an *Expert*, who provides guidance or content, and a *Novice*, who receives and acts on that input [63, 154, 156].

This asymmetry informs permissions and responsibilities: experts generate, edit and curate instructional material, while novices execute instructions and provide feedback. Table 4.2 summarizes the role-based access patterns in typical ACIE scenarios.

Table 4.2: Role-Based permissions in asynchronous immersive cooperation.

Permission	Expert	Novice
View instructions	✓	✓
Create instructions	✓	–
Modify instructions	✓	–
Delete instructions	✓	–
Report questions/issues	–	✓
View questions/issues	✓	–
Respond to questions/issues	✓	–
View responses	✓	✓

These roles help structure cooperation by aligning participant capabilities with the flow of task execution and support needs.

4.3.2 Message Categorization

To facilitate structured cooperation across time, messages in ACIE are categorized by intent and interaction pattern. Two primary classes are distinguished:

- **Directive Messages** typically sent by experts, these provide unidirectional guidance such as instructions or demonstrations. They initiate collaboration but do not anticipate immediate feedback.
- **Reactive Messages**, sent in response to directives, often by novices. These include questions, issue reports, confirmations and other replies that form a distributed dialogue.

This categorization enables temporally structured, bidirectional cooperation. For example, a novice may execute a directive message and subsequently send a question or issue report. The expert may then respond with an answer or solution, which continues the loop.

Table 4.3 lists common message types and their interaction patterns.

Table 4.3: Message types in asynchronous immersive cooperation.

Message Type	Direction	Sender	Receiver
Instruction	Unidirectional	Expert	Novice
Demonstration	Unidirectional	Expert	Novice
Question	Reactive	Novice	Expert
Issue Report	Reactive	Novice	Expert
Answer	Reactive	Expert	Novice
Solution	Reactive	Expert	Novice
Feedback	Reactive	Expert / Novice	Expert / Novice
Confirmation	Reactive	Any	Original Sender

This typology supports turn-based collaboration, allowing participants to engage in asynchronous problem-solving while maintaining message traceability and contextual coherence.

4.3.3 Interactive Proxies

To further approximate the interactivity of synchronous collaboration, ACIE introduces *Interactive Proxies* – enhanced versions of proxy avatars that serve as embodied interfaces for deferred interaction.

In addition to replaying past actions, interactive proxies support contextual user input and response delivery. They allow collaborators to:

- **Initiate new messages** – by interacting with a proxy avatar, users can ask questions, report issues or submit feedback;

- **Access responses** – proxies can act as representatives for receiving answers or follow-up content.

This interaction model enables participants to communicate indirectly but meaningfully through embodied representations. For instance, a novice might interact with an expert's proxy avatar to report a problem. Later, the expert can engage with the novice's proxy to review the issue and reply accordingly.

Proxies may incorporate features such as full-body avatars, gaze direction, facial expressions or voice playback (➜ **Chapter 5**), enhancing the sense of presence and supporting intuitive engagement.

Conceptually, Interactive Proxies enable asynchronous co-presence – allowing users to interact with temporally displaced collaborators in a shared spatial environment. They form a bridge between recorded behavior and responsive dialogue, extending static replay into dynamic cooperation.

4.3.4 Asynchronous Interaction

Figure 4.5 illustrates the asynchronous interaction loop between expert and novice collaborators. This loop implements asynchronous cooperation through a structured exchange of directive and reactive messages.

The process begins when the expert creates a directive message (e.g. an instruction or demonstration). The novice receives this message and applies it in their task context. If difficulties arise, such as unclear instructions or encountered issues, the novice creates a reactive message (e.g. question or issue report) by interacting with the expert's proxy avatar. When the expert next enters the system, they review the reactive message via the novice's proxy and respond, either by clarifying (response) or adapting the original instruction (instruction update).

This back-and-forth interaction enables problem-solving and coordination despite temporal separation, while approximating direct interaction between the

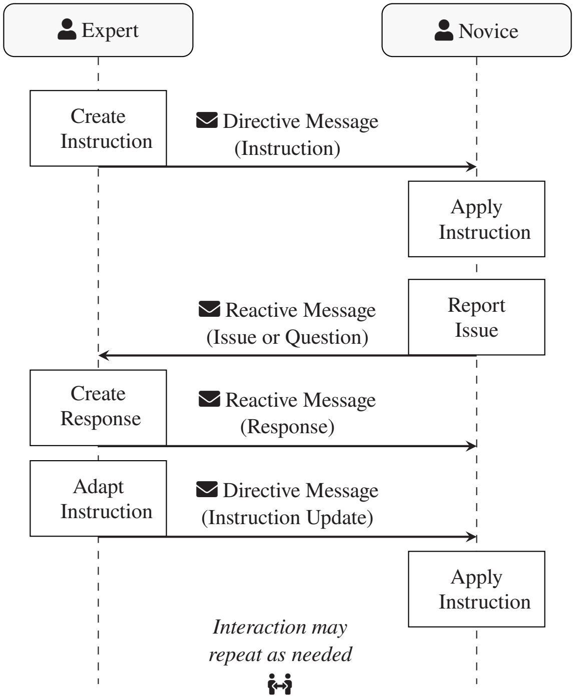


Figure 4.5: Enhanced asynchronous sequence diagram with explicit message types and report interaction step.

collaborators through their proxies. Thus, ACIE compensates for the lack of co-presence and fosters interactive collaboration across time.

This interaction loop addresses **RQ1.2** by providing a mechanism through which users engage in question-and-answer processes asynchronously within spatial XR environments. Specifying the following RQ for further investigation:

> **RQ1.2:** Can interactive question–answer mechanisms in XR be used to improve asynchronous cooperation in CT?

The proxy-mediated exchange of structured messages supports mutual adaptation, task clarification, and contextualized feedback, key aspects of successful cooperation.

Although the diagram is presented in an asymmetric, expert-novice configuration, the underlying structure generalizes to symmetric collaboration as well. In this case, all collaborators can: create and share directive messages (e.g. documentation or instructions) and respond to one another with reactive messages (e.g. providing feedback). This supports a broad range of use cases, including peer workflows, co-creation scenarios and distributed documentation.

4.4 Asynchronous Coordination

Asynchronous collaboration in XR environments presents unique coordination challenges due to the decoupling of time and co-presence between contributors. Coordination in this context must address not only communication but also the structured delivery, management, and integration of spatial contributions and immersive messages created asynchronously.

4.4.1 Message Coordination

A central challenge in asynchronous coordination is managing messages that are temporally and contextually displaced. Effective collaboration requires not only the transmission of Immersive Messages, but also mechanisms to structure, order and reference them in ways that preserve temporal continuity and task coherence.

While synchronous coordination benefits from real-time awareness, co-presence and immediate feedback, asynchronous collaboration must function in the absence of these cues. Contributors are often compelled to provide comprehensive information within each message, anticipating possible responses or clarifications. As a result, asynchronous communication tends to become denser, encompassing multiple task steps, decisions and potential contingencies within a single message.

This temporal dispersion increases the cognitive and organizational demands of coordination. Without structural support, collaborators may lose context, overlook dependencies, or duplicate effort. To mitigate these risks, it is essential to make visible the temporal sequence, task progression, and interaction dependencies – enabling collaborators to understand not only what was done, but also when and why.

A common method for sharing immersive contributions involves recording entire sessions and replaying them asynchronously through proxy avatars. While this offers embodied communication, it introduces coordination difficulties: referencing specific moments within long recordings is cumbersome, especially when relying on abstract indicators such as timestamps.

To address this, ACIE introduces the concept of *Stepwise Structuring*. Rather than treating an immersive message as a monolithic recording, users segment it into discrete steps aligned with the logical structure of the task – either manually or through semi-automated capture per step. Each step represents a modular sub-message, allowing collaborators to:

- Replay specific segments on demand,
- Navigate non-linearly between steps (e.g. forward, backward, repeat),
- Clarify individual actions or instructions,
- Integrate new contributions in a logically structured workflow.

Stepwise structuring improves flexibility and navigability. Segments can be independently replaced or updated, and reactive messages can be anchored to specific steps, enhancing traceability and interaction clarity.

Moreover, this structure enables advanced forms of coordination: feedback can be contextualized per task segment, review markers and status indicators can be embedded within the message timeline, and collaboration histories can be constructed through branching and merging paths – allowing multiple contributors to operate in parallel and later converge their input.

Prior work has shown that stepwise instruction improves collaborative task performance by reducing information overload and enhancing recall [154]. In ACIE, immersive messages are thus reimagined as interactive, navigable timelines – not merely playback files, but structured coordination artifacts.

Stepwise Structuring constitutes a foundational contribution to asynchronous XR collaboration. It supports orientation across time, fosters responsive interaction, and enables modular integration of contributions across collaborators.

To explore this contribution empirically, the following sub-question is addressed in subsequent chapters:

> ❓ **RQ1.3:** Can step-by-step organization of recordings and immersive version control in XR be used to improve asynchronous coordination in CT?

4.4.2 Work Coordination

While message coordination ensures temporal alignment and referenceability of communication, asynchronous collaboration also demands methods for managing the evolution of shared spatial work environments. *Work Coordination*, in this context, refers to the systematic handling of contributions, modifications and accumulated progress over time within an immersive workspace.

This challenge is particularly evident in XR, where changes are not textual or symbolic, but spatial, embodied and interactive. Contributors manipulate objects, alter positions and perform actions embedded directly in the 3D environment. Capturing, exchanging and integrating these spatial states asynchronously requires a fundamentally different coordination logic than that used in traditional collaborative systems.

Version Control Systems (VCS) offer a mature model for managing contributions and tracking progress in software development. However, their application to immersive environments remains limited. Prior work has explored VCS-like systems for 3D models and physical assemblies, allowing users to inspect changes or visualize construction histories over time [96, 118]. While these approaches demonstrate the value of tracking spatial evolution, they do not fully address the complexities of asynchronous, immersive collaboration.

ℹ The iVCS adapts core version control concepts to immersive spatial workflows:

Git Command	**Git Function**	**iVCS Function**
Commit	Record changes to the repository	Capture 3D workspace snapshot
Push / Pull	Transfer commits to/from a remote repository	Exchange versions asynchronously
Branch	Create a new line of development	Create independent spatial work paths
Merge	Integrate changes from another branch	Resolve and integrate 3D contributions

A more recent system by Zhang et al. [172] introduces a truly immersive VCS for VR, supporting spatial contribution recording, persistent history tracking as well as branching and merging of collaboration paths. Building on this foundation, ACIE proposes an *immersive Version Control System (iVCS)* specifically tailored for asynchronous collaboration on complex tasks in XR.

Inspired by Git[2], the iVCS enables contributors to record, exchange and organize spatial contributions directly within the XR environment – without relying on external tools that would disrupt immersion. To support asynchronous work coordination, it manages three core operations: registering new contributions, exchanging them between collaborators and organizing the shared progress. These operations correspond to the core Git concepts of commit, push/pull and branching/merging.

New Contributions

Collaborators can capture the current state of the immersive workspace as a versioned snapshot. Each version encodes object positions, transformations and interaction states, representing a distinct point in the task's evolution. Unlike symbolic commit operations in software, this process is integrated into the spatial workflow – allowing users to commit their progress without breaking immersion.

Exchanging Contributions

Since asynchronous collaborators work at different times and locations, the iVCS enables them to exchange contributions as packaged state updates. These include version lineage (e.g. predecessor-successor relationships), supporting the reconstruction of how each contribution builds on prior work. Push and pull operations are seamlessly embedded within the immersive environment.

[2] Git is a distributed version control system used in collaborative software development: https://git-scm.com/

Organizing Contributions

As collaboration progresses, divergent contributions may arise. The iVCS allows contributors to branch their work into parallel histories and later merge them into a unified version. Conflict detection is critical: spatial conflicts occur when collaborators make incompatible changes to the same object. These must be resolved interactively and visually, leveraging XR's affordances for inspecting 3D states and maintaining semantic consistency in the shared scene.

Therefore, the iVCS constitutes a second foundational mechanism for asynchronous coordination in XR. Whereas stepwise message structuring supports communication flow, the iVCS ensures that spatial work contributions remain persistent, versioned and resolvable across time and collaborators.

The investigation of how this concept contributes to asynchronous coordination in complex tasks is included in **RQ1.3**.

4.5 Asymmetric Collaboration

In virtual engineering, collaboration often spans the full XR continuum – from fully virtual environments (VR) to physical spaces enhanced with digital content (AR and MR). Depending on the task stage, user role or physical location, collaborators may engage with different modalities of the same workspace. Supporting such asymmetry is essential for real-world deployment, where tasks must transition fluidly between physical and virtual contexts.

ACIE addresses this requirement by ensuring that asynchronous collaboration functions consistently across XR modalities. Core components – including Immersive Messages, Proxy Playback and the Immersive Communication Protocol (ICP) – are designed to operate agnostically across VR, AR, and MR. This enables collaborators to create, exchange and replay contributions regardless of their current modality.

However, this flexibility introduces spatial challenges. To maintain coherence between modalities, the system must ensure accurate alignment of virtual contributions with physical environments. Robust **registration and localization** – especially in MR – are therefore integral to ACIE's implementation (➋ **Chapter 5**) and ensure that spatial contributions remain consistent across contexts.

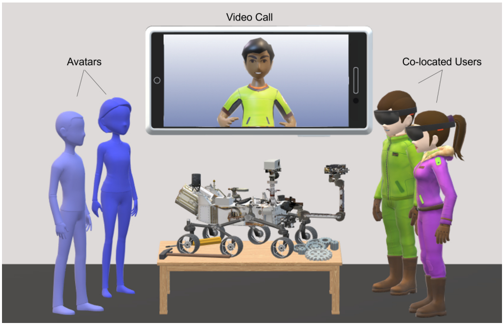


Figure 4.6: Asymmetric collaboration across XR modalities.

4.6 Summary and Outlook

This chapter introduced the ACIE framework – a holistic concept for supporting asynchronous collaboration on complex engineering tasks in XR. Building on the theoretical foundations of ➋ **Chapter 2** and the gaps identified in ➋ **Chapter 3**, the framework systematically addresses the three pillars of asynchronous collaboration (communication, cooperation, and coordination) through dedicated XR concepts.

- **Asynchronous Communication**: Enabling *Immersive Messages* and *Proxy Playback*, supported technically by the *Immersive Communication Protocol*. It allows spatially and temporally embedded actions to be reliably captured and replayed across collaborators and devices. These concepts are further addressed in **RQ1.1**.

- **Asynchronous Cooperation**: Enabling structured message types (directive vs. reactive), role-based interaction models (e.g. expert vs. novice) and *Interactive Proxies* for embodied, turn-based responses across time. These features support asynchronous mutual adaptation and are further explored in **RQ1.2**.
- **Asynchronous Coordination**: Addressed through two mechanisms – *Stepwise Structuring* of immersive content and the *Immersive Version Control System (iVCS)* – enabling organization, branching and integration of spatial contributions over time using Git-inspired versioning adapted for the use in XR. Both concepts are addressed in **RQ1.3**.
- **Asymmetric Collaboration**: While not in the focus of ACIE research, asymmetric collaboration was addressed as a practical necessity for engineering applications. Through registration and localization, ACIE ensures spatial alignment of contributions across modalities (VR, AR, MR). These aspects are essential for real-world deployment.

Together, these concepts form a modular yet interrelated foundation for asynchronous collaboration in XR.

In Simpler Terms

This chapter outlined how engineers can collaborate effectively in XR environments, even when working at different times and from different locations.

The ACIE framework enables users to leave behind detailed, embodied records of their actions – including gestures, object manipulations and voice instructions – that can later be revisited by others in the same spatial context. These asynchronous contributions function much like a temporal footprint in the shared workspace, enabling others to observe, respond to and build upon previous work, as if the original collaborator were present.

To manage this process, ACIE draws on familiar paradigms such as messaging systems and version control – reimagined for spatial and embodied interaction. As a result, every contribution becomes part of a structured, persistent collaboration history that supports clarity, continuity and coordination over time.

How to Proceed

The conceptual foundation laid out in this chapter serves as the basis for the following sections:

- **Chapter 5** presents the technical implementation of the ACIE framework, detailing how the proposed concepts were realized and integrated.
- **Chapter 6** addresses **RQ2**, focusing on the evaluation of the framework in applied engineering scenarios.

ACIE addresses the core components of asynchronous collaboration in XR. The next chapter presents its technical implementation and prepares for its evaluation in applied engineering contexts.

5 Implementation

In this chapter, the implementation of the ACIE framework, conceptually introduced in ❯ **Chapter 4**, is described. While a complete technical documentation of all components and iterations is beyond the scope of this dissertation, selected insights into key architectural decisions, representative prototypes and implementation phases are presented. The process follows the principles of Design Science Research (DSR), in which conceptualization and implementation are conducted as iterative cycles. Accordingly, the system was not realized in a single step but through repeated loops of implementation, evaluation, refinement and re-evaluation. This iterative character is reflected in the multiple prototypes and user studies detailed in the sections that follow.

5.1 System Overview

The implemented system architecture[1], illustrated in Figure 5.1, supports asynchronous collaboration in XR through a modular and extensible pipeline. At its core, the system is orchestrated by the `Replay Manager`, which coordinates the capture, storage and playback of spatial activity data. This includes skeletal avatar poses, full-body motion and user interactions with virtual objects – enabling rich reconstruction of asynchronous user actions.

[1] This implementation under the acronym SKI^2LL-AI ranked fourth among 30 nominated projects in the KIT NEULAND Innovation Contest 2025: https://kit-neuland.de/en/innovation-contest

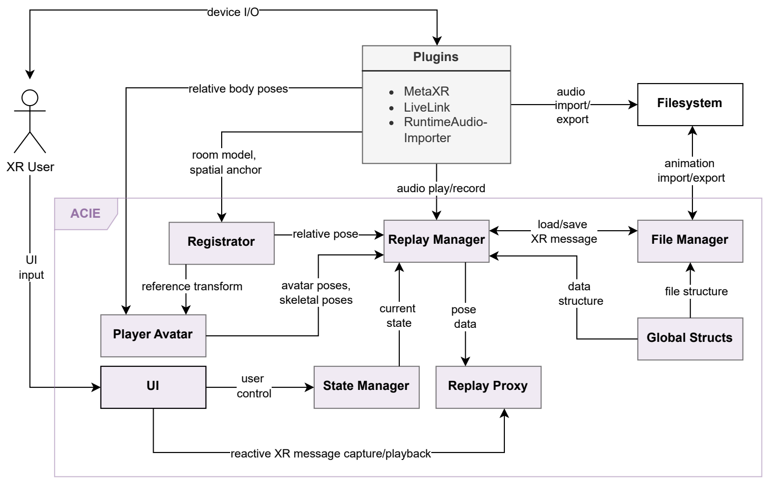


Figure 5.1: System architecture overview of the ACIE framework, illustrating key modules for XR interaction, asynchronous messaging, and file-based coordination. Arrows indicate data and control flow between runtime and support components.

The prototype was developed in Unreal Engine (UE) 5.3.2[2], targeting standalone XR headsets such as the HoloLens[3] and Meta Quest[4]. This platform focus allows for untethered deployment in real-world environments, including production sites and shop floors.

To capture and align motion data, the system integrates both the `MetaXR` plugin[5] (v1.97.0) and UE's native `LiveLink` system. The `MetaXR` plugin enables full-body pose estimation via Meta's Body Movement SDK and supports passthrough MR, room modelling and spatial anchoring. Spatial anchors are

2 UE (UE) is a high-performance, real-time development environment supporting modular C++ and visual scripting, and native XR integration: https://www.unrealengine.com
3 Microsoft HoloLens Augmented Reality Head Mounted Display: https://learn.microsoft.com/de-de/hololens/
4 Meta Quest Mixed Reality Head Mounted Display: https://www.meta.com/de/quest/
5 MetaXR Plugin: https://developers.meta.com/horizon/downloads/package/unreal-engine-5-integration

used to register all recorded interactions relative to the physical environment. `LiveLink` translates real-time tracking input into skeletal poses compatible with UE's animation blueprint system. These poses are applied to the `Player Avatar` during live operation and to the `Replay Proxy` during asynchronous playback.

The `Registrator` component computes the relative transformation between the player and the environment using `MetaXR` spatial anchor data, ensuring consistent alignment and spatial referencing across devices and sessions. The `Replay Manager` records pose sequences and associated metadata using globally defined structs such as `Frame`, `Pose` and `Record`, which serve as the backbone of the immersive message format. Structured data is passed to the `File Manager`, serializing and exporting these immersive messages to the `Filesystem`.

Conversely, the `File Manager` also handles the import and parsing of immersive messages from storage. These are passed to the `Replay Manager`, where the `Replay Proxy` is animated and transformed in each update cycle based on the reconstructed motion data.

System transitions between operational states (e.g. capture, playback, pause) are managed by the `State Manager`, which communicates with both the `Replay Manager` and the UI layer. User input from virtual menus (e.g. step navigation, question playback or answer recording) is forwarded to the relevant system components to trigger the corresponding actions.

Finally, the system manages voice input and audio data via the `Runtime Audio Importer` plugin[6] (v1.0 for UE 5.3+) and UE's native audio capture and playback systems. This enables immersive messages to convey not only gestural and spatial content, but also spoken instructions, facilitating multi-modal asynchronous communication. Audio and motion playback are synchronized by the `Replay Manager` based on time-aligned sequencing.

[6] Runtime Audio Importer: https://solutions.georgy.dev/runtime-audio-importer

5.2 Asynchronous Communication

This section outlines the technical implementation of asynchronous communication using `XR Messages`, which operationalize the concept of Immersive Messages introduced in **Chapter 4**. These messages enable temporally decoupled task guidance and feedback by encoding spatial, gestural and verbal cues into replayable data.

5.2.1 XR Messages

As shown in Figure 5.2, the ACIE framework captures spatial, gestural, and verbal data through modular components to enable asynchronous, immersive communication. These `XR Messages` consist of motion data, gestural input, and synchronized audio, allowing one user to leave immersive instructions that can be replayed asynchronously by another user via a proxy avatar.

Player Avatar

The fidelity and structure of the user avatar determine the granularity and richness of the captured data. In earlier versions of ACIE, simplified upper-body avatars were used, consisting only of head, torso and hands. Accordingly, only the head and hand poses needed to be captured and replayed, alongside metadata such as hand states (e.g. open, grab) Mayer et al. [5].

As the system evolved, so did the avatar complexity, and with it, the sophistication of data required for realistic proxy representation. A major advancement was the introduction of hand tracking, enabling high-fidelity capture of finger motions and detailed hand gestures.[7] Mayer et al. [8]. This advancement significantly enhances non-verbal communication fidelity-crucial in practical

[7] MetaXR hand tracking provides 25 tracked joints per hand, supporting natural gestures and tool interactions; see Appendix A, Figure A.3.

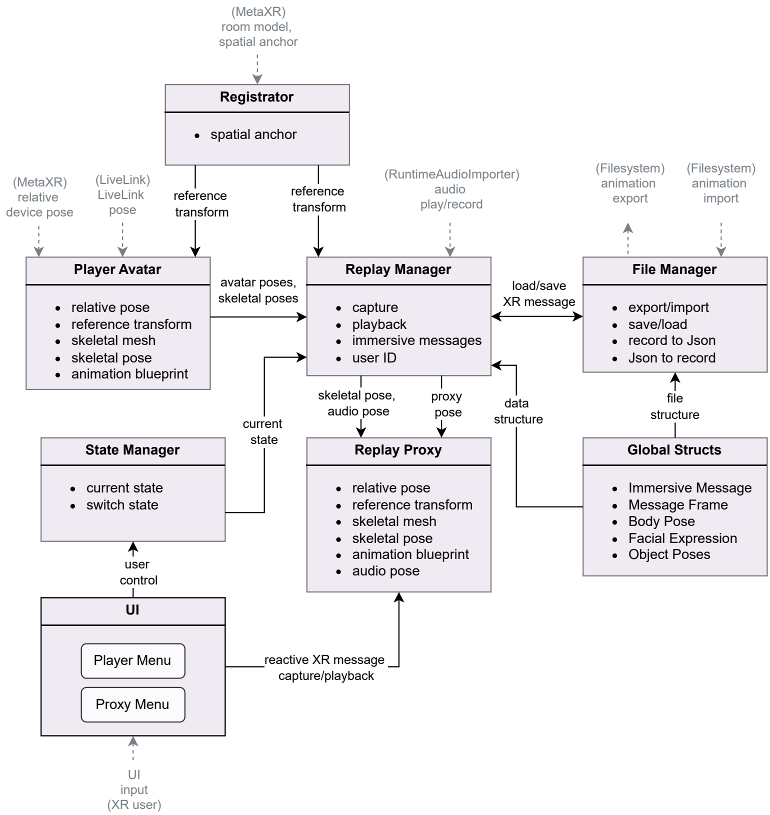
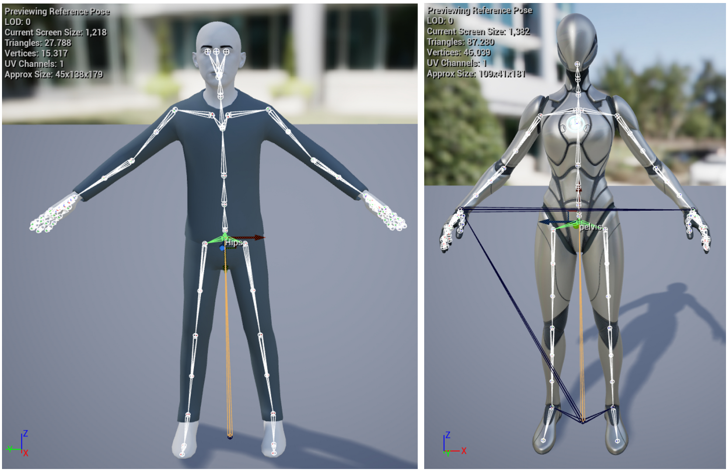


Figure 5.2: Detailed overview of the ACIE components involved in avatar capture and proxy playback

tasks, where gestures and handling actions are often difficult to convey purely through verbal description.

Another substantial improvement was the adoption of full-body avatars, which likewise required capturing and replaying full-body movements on proxy representations 📕 Kastner et al. [2]. These avatars improve the perception of



(a) MetaXR skeletal mesh structure **(b)** UE Quinn skeletal mesh structure

Figure 5.3: Comparison of skeletal mesh structures of MetaXR and UE avatars, used for retargeting in avatar motion replay. Detailed joint structures are provided in Appendix A, Figures A.1 and A.2

co-presence and embodiment, reportedly enhancing user experience in collaborative XR scenarios.

In the current implementation, the default UE mannequin avatar *Quinn* is used. Figure 5.3 illustrates the differences between the skeletal structure retrieved via `MetaXR` API and that of the UE avatar. These include coordinate systems, joint alignments (e.g. pelvis region) and naming conventions. Since `MetaXR` provides raw joint data at runtime, a retargeting asset is employed to map this data onto the UE skeleton. Corresponding joint names and mappings are detailed in ➜ **Appendix A**, Table A.2, Table A.3 and Table A.4.

During runtime, skeletal tracking data is streamed via `MetaXR` and `LiveLink` into the avatar system. This data is processed in the avatar's animation blueprint, using inverse kinematics and retargeting logic to reconstruct and animate the user's full-body motion in real time. The resulting animation reflects the user's embodied actions and forms the basis for immersive message capture.

Avatar Capture

The avatar motions, specifically the skeletal joint transformations computed in each update frame, are captured by the `Replay Manager`. The capture sequence, shown in Figure 5.4, is triggered by the `Replay State Machine`, which initiates the recording phase.

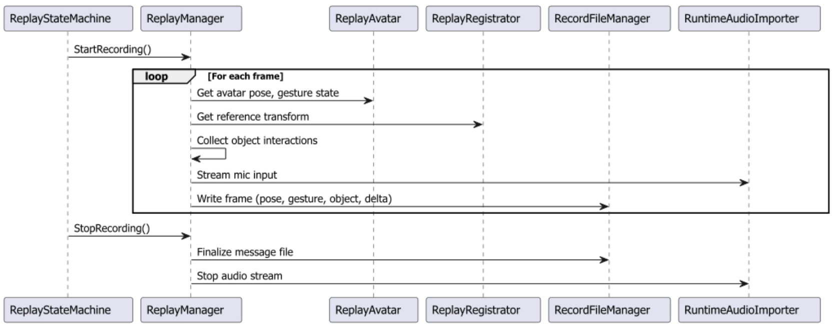


Figure 5.4: Sequence diagram of the avatar recording process

Upon activation, a new message data structure is created (❯ Section 5.2.2, Figure 5.6). In every frame, the current relative avatar pose, skeletal joint transforms (referenced to the spatial anchor) and manipulated object transforms are appended. The necessary reference transform is retrieved from the `Registrator`. Scene understanding and spatial anchoring is further described in ❯ Section 5.5. When the system exits the recording state, the XR Message is serialized into a JSON file and saved together with the associated audio recording via `File Manager` to files on the local `File System`. The pseudocode of this procedure is depicted in Algorithm 1.

Proxy Playback

Once an *XR Message* has been recorded, it can be replayed using the system's *Proxy Playback* functionality. During this process, a proxy avatar reenacts the

Algorithm 1 Record Immersive Message

Initialize message M
Start audio-stream recording
$t_0 \leftarrow$ current time
$referenceTransform \leftarrow$ spatial anchor transform
while recording state **do**
    $\Delta t \leftarrow$ current time $- t_0$
    $relativePose \leftarrow$ player transform in scene
    $skeletalPose \leftarrow$ joint transforms applied with $referenceTransform$
    $objectsPose \leftarrow \emptyset$
    for each manipulated object o **do**
        $p_o \leftarrow$ transform of o applied with $referenceTransform$
        Add (o, p_o) to $objectsPose$
    end for
    $frame \leftarrow (\Delta t,\ relativePose,\ skeletalPose,\ objectsPose)$
    Append $frame$ to M
    $t_0 \leftarrow$ current time
end while
Stop audio recording and save as WAV
Serialize M to JSON
Save JSON and WAV to `Filesystem`

previously captured motion and object interactions, effectively delivering the immersive message to another user in a shared XR environment. It is crucial that the environment state remains unaffected, supporting non-invasive guidance, asynchronous collaboration or side-by-side task execution. The playback loop is detailed in Algorithm 2.

Algorithm 2 Replay Immersive Message (Proxy Playback)

Load frame sequence $F = [f_1, f_2, \ldots, f_n]$ from JSON
Start audio playback and sync at t_0
$t \leftarrow 0$
for $i \leftarrow 1$ to n **do**
    Wait for Δt_i
    Apply pose f_i.*skeletalPose* to `Replay Proxy`
    Apply pose f_i.*objectsPose* to object proxies
    $t \leftarrow t + \Delta t_i$
end for

Playback is managed by the `Replay Manager`, loading the recorded message file from the `File System` via the `File Manager`, as shown in the component overview (Figure 5.2). The internal execution sequence is visualized in Figure 5.5.

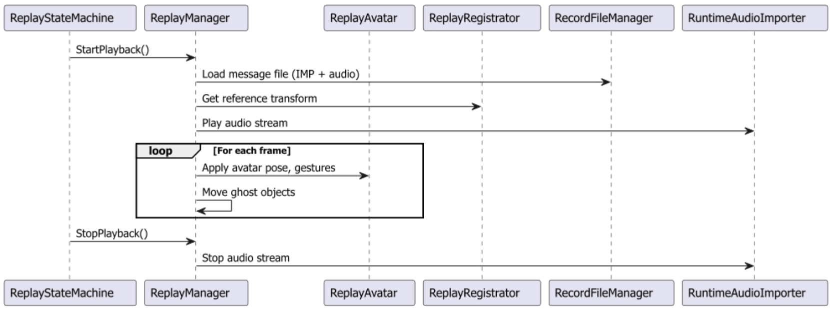


Figure 5.5: Sequence diagram of the proxy avatar playback process

Here, the `Replay State Machine` triggers playback by signalling the `Replay Manager`, which then loads the message, retrieves the reference transform from the `Registrator` and starts audio playback. In each frame of the update loop, the skeletal pose and object transforms are applied to their respective `Replay Proxies`. Each message consists of a sequence of time-stamped frames containing motion and proxy data. Once playback is complete or cancelled by the user, the audio and proxy playback is stopped and the system returns to idle.

This mode enables immersive message delivery as a temporally synchronized, embodied experience-preserving spatial references, supporting verbal and non-verbal communication and allowing asynchronous collaboration without requiring the sender's live presence.

5.2.2 XR Messaging Protocol

To enable consistent recording, interpretation and transmission of `XR Messages`, the system relies on a well-defined message protocol. This protocol specifies the structure and semantics of all data elements that make up an immersive message, including avatar poses, object interactions, timing and optional audio references.

By enforcing this structured format, the system ensures that Immersive Messages can be serialized, stored and later interpreted correctly across different devices and contexts. It also facilitates interoperability and debugging, and enables future extensions such as AI-based interpretation, partial replays or remote message visualization.

Each message is encoded as an `Immersive Message` data structure, as illustrated in Figure 5.6. It includes metadata fields such as message ID, author and message type (e.g. step, question, answer), as well as a sequence of time-stamped frames. Each frame contains a skeletal pose and a list of manipulated objects, representing the user's spatial actions at a specific moment.

The skeletal pose itself is further decomposed into joint-level data structures to support full-body motion capture. Object interactions include spatial transformations and metadata such as the path to the associated mesh. Time-stamps enable precise synchronization between motion data and speech playback.

This message protocol forms the backbone of the system's file-based communication and enables asynchronous exchange of embodied behavior in XR.

5.2.3 Evaluation of Asynchronous Communication

To assess the utility and feasibility of the asynchronous communication system, a sequence of user studies was conducted across different prototypes and development stages. Each study targeted specific aspects of the system's effectiveness in immersive collaboration scenarios and led to iterative improvements within the DSR design cycle. Figure 5.7 presents the decomposition of **RQ1** into the more specific subquestion **RQ1.1** and its corresponding hypothesis **H1.1** *"Avatar capture and playback in XR exhibit acceptable usability for asynchronous communication in CT, according to usability criteria."*.

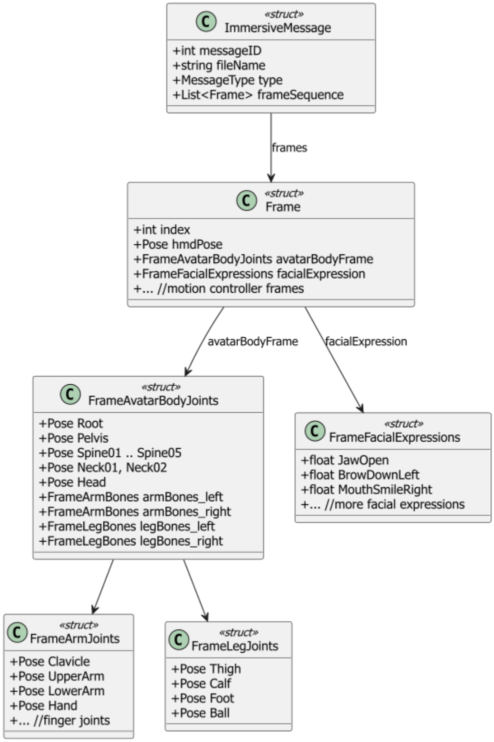


Figure 5.6: Data structures used in the immersive message protocol

To validate the hypothesis, several experiments were carried out to assess the practical use of the concepts implemented in ACIE prototypes. Standardized metrics such as the System Usability Scale (SUS) and NASA TLX – widely used in HCI and CSCW – were employed to evaluate usability and cognitive load. These studies enabled iterative validation of system functionality and informed framework refinements for the final evaluation.

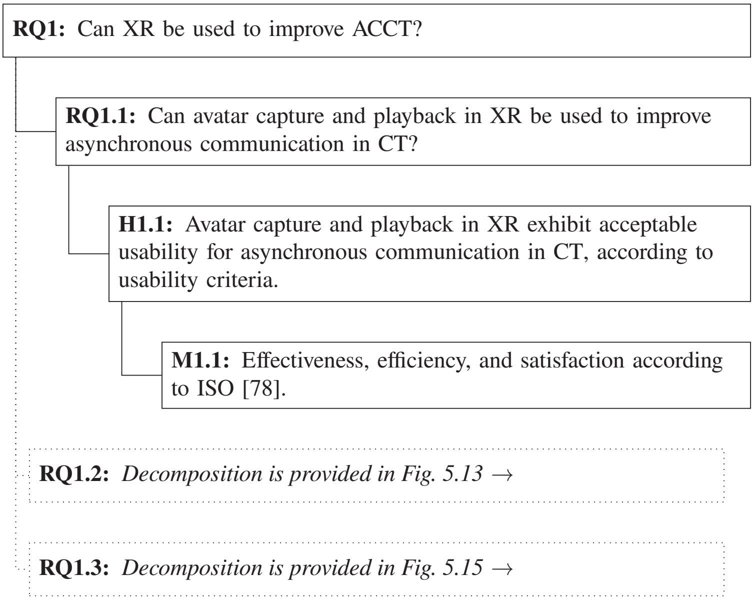


Figure 5.7: Decomposition of **RQ1.1** with corresponding hypothesis and measurements.

Literature: This concept's implementation and evaluation appeared in Mayer et al. [5] within a VR context, were discussed in Mayer et al. [8] for AR, and were analyzed in Kastner et al. [2] with MR.

Initial validation was conducted in VR using a prototype with a simplified half-body ghost avatar Mayer et al. [5]. In a between-subject study ($N = 10$), participants completed a 10-step assembly task using virtual parts (e.g. ball bearings, screws, circlips, rods, gears) with an Oculus Quest 2 HMD. Guidance was provided either by replayed translucent "ghost"-avatar instructions or by a PDF manual shown in VR – step navigation was available in both conditions. Usability was high in both groups (SUS: Avatar $M = 86.0$, $SD = 7.83$; Manual $M = 78.5$, $SD = 12.57$), with no significant difference (Mann-Whitney $U = 6.5$, $p = .245$). NASA-TLX totals were lower for the avatar condition

($M = 35.67$, $SD = 13.37$) than for the manual condition ($M = 45.50$, $SD = 13.99$), but the difference was not significant ($t(10) = 1.814$, $p = .121$). An After Scenario Questionnaire (ASQ) post-test showed similarly high satisfaction (Avatar $M = 4.40$, $SD = 0.37$; Manual $M = 4.60$, $SD = 0.37$), with no significant difference (Mann-Whitney $U = 8.5$, $p = .403$). Together, these results indicate comparable usability and efficiency, with trends toward lower mental demand, effort and temporal pressure in the avatar condition (see Fig. 5.8).

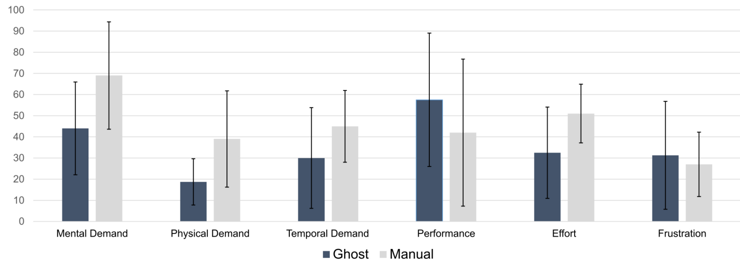


Figure 5.8: NASA-TLX workload ratings from the SalentoXR study, showing reduced cognitive demand on most subscales for the ghost avatar condition

The next iteration, HoloHands Mayer et al. [8], introduced multi-modal instructions using hand tracking and voice input. In a within-subject study ($N = 22$), participants performed two block-assembly tasks with 57 Jenga pieces each under both conditions: AR HoloHands on HoloLens 2 and a PDF manual on a desktop display. Accuracy was higher with the PDF condition (Wilcoxon $Z = 2.69$, $p = .01$, $r = .41$; $M_{\text{HoloHands}} = 45.82$, $M_{\text{PDF}} = 52.95$). Usability ratings were good in both conditions (SUS AR $M = 75.34$, $SD = 16.94$; SUS PDF $M = 84.09$, $SD = 12.04$) and did not differ significantly at the total-score level (paired $t(21) = -1.89$, $p = .072$; Wilcoxon $T = 77.5$, $p = .121$). NASA-TLX totals likewise showed no significant difference, while item-level effects indicated higher perceived cumbersomeness and need for support for AR (SUS4 $Z = 2.91$, $p < .01$, $r = .44$; SUS8 $Z = 2.54$, $p = .01$,

$r = .38$), lower perceived performance on TLX ($Z = 3.04$, $p < .01$, $r = .46$), and lower ASQ time satisfaction for AR ($Z = 2.30$, $p = .02$, $r = .35$).

Finally, the full-body avatar prototype was evaluated in a real-world pump-station setup 📕 Kastner et al. [2]. Participants ($N = 11$) executed a scripted maintenance procedure (inspection, valve and level checks, and a measurement run increasing pump frequency from 8 to 32 Hz in 2 Hz steps) while following avatar proxy instructions on a standalone MR headset. All participants completed the task without errors. Usability (SUS) differed descriptively by experience (experienced $M = 75.0$, $SD = 12.4$; inexperienced $M = 66.4$, $SD = 16.1$), but the group difference was not significant (Welch $t(7.88) = 0.99$, $p = .353$). Workload (NASA-TLX) was comparable between groups (inexperienced $M = 3.16$, $SD = 1.05$; experienced $M = 3.57$, $SD = 1.41$); the difference was not significant (two-sample $t(9) = 0.55$, $p = .595$).

Table 5.1: Usability (SUS) of Avatar Capture and Playback across evaluated implementations.

Study	Condition A $M \pm SD$	Condition B $M \pm SD$	p
Mayer et al. [5] (VR)	Replay Avatar 86.0 ± 7.83	PDF Manual 78.5 ± 12.57	.245
Mayer et al. [8] (AR)	AR HoloHands 75.34 ± 16.94	PDF Manual 84.09 ± 12.04	.121
Kastner et al. [2] (MR)	Experienced 75.0 ± 12.4	Inexperienced 66.4 ± 16.1	.353

These results are summarized in Table 5.1 and provide formative support for hypothesis **H1.1**. Across studies, usability was good to high and all tasks were completed. Where inferential tests were conducted, differences on total SUS and overall NASA-TLX were not statistically significant; however, descriptive tendencies toward lower mental demand, effort and temporal pressure with avatar replay, together with positive user ratings, indicate that immersive message replay is feasible and well accepted for asynchronous communication.

Summary: This section evaluated the feasibility of asynchronous communication using avatar-based proxy playback across VR, AR and MR contexts. Results from successive prototypes showed comparable workload and good usability; together, these findings provide formative support for **H1.1** and indicate that immersive message replay can support communication and task guidance in asynchronous collaboration.

5.3 Asynchronous Cooperation

Asynchronous collaboration in immersive environments introduces unique challenges due to the decoupling of time and co-presence. Prior research on AR-based remote assistance has shown that shared hand gestures and annotations can support spatial tasks effectively [160, 161]. Building on this, the ACIE framework expands asynchronous cooperation by introducing embodied proxy avatars that not only replay captured user behavior, but also enable meaningful interaction with non-present collaborators.

5.3.1 Proxy Avatars

Proxy avatars serve as interactive stand-ins for remote users, embodying their prior actions, queries or responses. Depending on the context, different avatar types were implemented throughout the framework development to suit task and embodiment needs. Figure 5.9 shows the three implemented proxy-avatar types – half-body, full-body, and Metahuman – in real environments.

Half-Body Proxy Avatar

An early, simplified version consisting of head, torso and hands. In applications using controllers, virtual hands represent the controllers with three animation

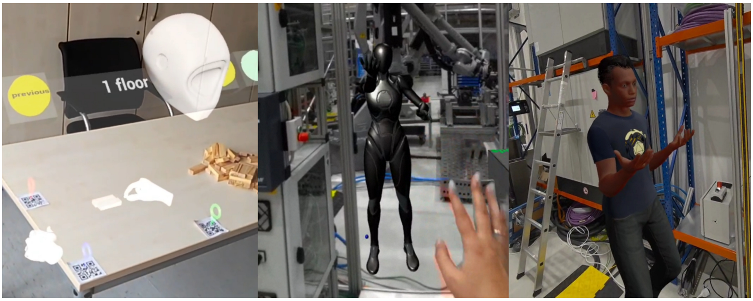


Figure 5.9: Examples of proxy avatars implemented in the framework. From left to right: a half-body proxy (head and hands) in see-through AR; a full-body proxy avatar in passthrough MR; and a Metahuman proxy avatar in passthrough MR.

states: open, near-grab, and grab. In stationary scenarios – such as workbench assembly tasks – users primarily focus on hand and object interactions. In these settings, the half-body proxies proved sufficient to convey hand actions effectively.

Technically, only the head and hand transformations are recorded. In controller-based applications, virtual hand states are captured. In hand tracking scenarios, the system logs all hand joint data to enable smooth motion and gesture replay.

Full-Body Proxy Avatar

For non-stationary, room-scale scenarios, full-body avatars were implemented using the abstract default characters from UE, as illustrated in Figure 5.3b. These avatars enhance the perceived co-presence of non-present collaborators and help users interpret spatial navigation. Compared to floating half-body proxies, grounded full-body avatars more clearly communicate location and movement-especially important in AR and MR contexts, where small registration mismatches can occur between virtual and physical spaces.

Full-body motion is estimated using the `MetaXR` plugin's inverse kinematics model. The resulting skeletal animation is retargeted to the standard UE mannequin via Unreal's retargeting tools. This retargeting increases modularity, as the UE skeletal structure is widely supported within the ecosystem.

During playback, the recorded skeletal transformations are applied via an animation blueprint, driving the avatar's full-body motion by mapping the data to corresponding joints.

Metahuman Proxy Avatars

Since full-body motion is mapped to the UE skeleton, it is possible to swap the avatar mesh with any compatible character model – including photorealistic Metahumans[8]. This allows a high degree of personalization.

Personalized avatars can be created by capturing a user's facial geometry using photogrammetry (e.g. via Polycam[9]) and mapping it onto a Metahuman character using UE's Metahuman Creator. Figure 5.10 shows an example with an integrated scanned face.

The resulting Metahuman can then be imported into UE and used as the skeletal mesh in the ACIE `Replay Proxy` component.

Ghost Proxy

To distinguish proxy avatars from live collaborators and to preserve spatial clarity, ghost metaphors were employed. Both avatars and proxy objects are rendered semi-transparently, visually separating them from avatars 📘 Mayer et al. [5]. This enhances perceptual clarity, supports temporal awareness and reduces visual clutter during task execution, particularly useful in asynchronous

8 UE's Metahuman: https://www.unrealengine.com/en-US/metahuman
9 3D scanned using Polycam: https://polycam.io/

(a) 3D surface model **(b)** Metahuman with scanned face

Figure 5.10: Creation of a personalized Metahuman avatar using photogrammetry

contexts where proxies replay past actions. Figure 5.11 depicts different ghost proxies in VR and MR.

(a) Ghost proxy in VR **(b)** Ghost proxy in MR

Figure 5.11: Ghost avatars and object proxies for non-intrusive replay

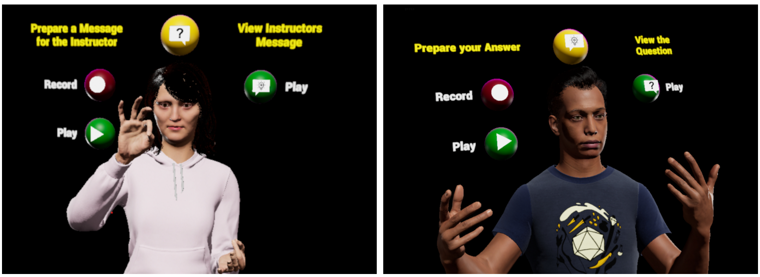
Object Proxies

In VR scenarios – and occasionally in AR or MR – users interact with virtual objects. When this interaction is recorded, the system logs the object's model reference, pose and transformation sequence. During playback, the proxy object is reconstructed using the recorded model path and animated using the logged transformation data. Figure 5.11a shows ghosted object proxies (in blue) alongside an avatar. Unlike proxy avatars, object proxies are non-interactive, they serve purely for contextual replay of user actions.

5.3.2 Role-Aware Proxy Interaction

Proxy avatars not only replay behaviors, they act as communication anchors. Depending on the user's role (e.g. novice or expert), interaction menus embedded in the proxies allow for recording and playback of reactive messages, contextualized to specific task steps (➔ Section 5.4).

Figure 5.12 illustrates the role-based menu logic: novices interact with expert proxies to ask questions, while experts interact with novice proxies to record and deliver responses. Each menu offers preview, record and playback functionality, reinforcing the temporal link between collaborators.



(a) Novice interacting with expert proxy **(b)** Expert interacting with novice proxy

Figure 5.12: Role-specific proxy menus for asynchronous interaction. The novice records and previews a message on the expert's proxy (e.g. a question) and later plays back the expert's response. The expert reviews the original issue and records an answer on the novice's proxy.

5.3.3 Evaluation of Proxy-Based Asynchronous Cooperation

To assess the impact of embodied proxy avatars on asynchronous cooperation, a controlled within-subject study was conducted comparing three different guidance modalities in an industrial setting. Figure 5.13 presents the decomposition of research question **RQ1** into its specific subquestion **RQ1.2**, the corresponding hypothesis **H1.2** and measurement criteria **M1.2**.

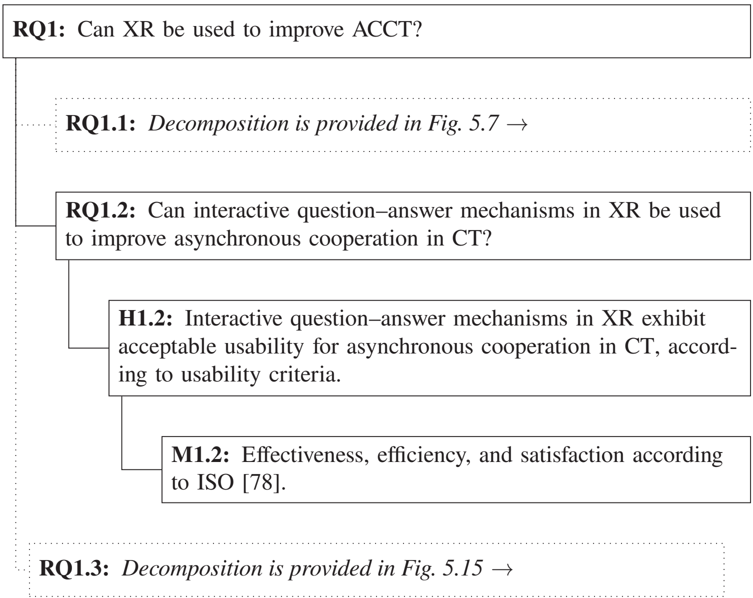


Figure 5.13: Decomposition of **RQ1.2** with corresponding hypothesis and measurements.

Participants completed an industrial commissioning task that required identifying system errors and using the proxy interaction system to communicate asynchronously with an expert. Each participant experienced all three instruction modalities: (1) a 2D Video on a smartphone (baseline), (2) AR video playback embedded in the physical environment using Microsoft HoloLens and

(3) an avatar proxy playback using the Meta Quest Pro. Figure 5.14 depicts all modalities in the experiment setting.

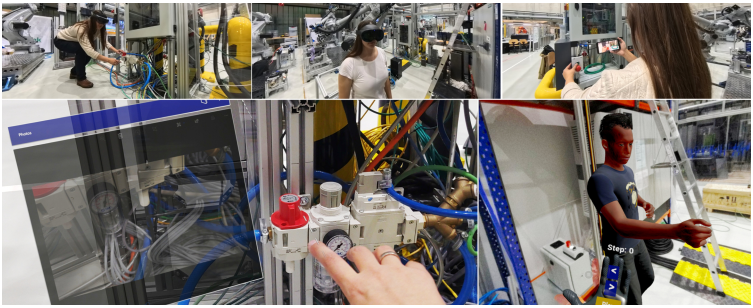


Figure 5.14: Industrial task guidance using three different asynchronous instruction modalities: (1) 2D video on smartphone (top right), (2) AR video embedded in physical space via HoloLens (top left and bottom left), and (3) immersive avatar-based playback in VR using Meta Quest Pro (top center and bottom right)

Participants used the role-based proxy interaction menus to record, receive and playback asynchronous voice-and-motion messages in context. Between iterations, they filled out questionnaires while the expert prepared their responses. Afterward, participants could return to the task and address the reported issues using the expert's answers.

Cooperation was measured using a custom Likert scale $(1-5)$, reflecting perceived helpfulness, engagement and coordination effectiveness of the interaction. Usability was assessed using the SUS. All participants were able to successfully complete the task in each condition, confirming that the asynchronous Q&A system functioned reliably and supported effective information exchange.

Quantitative results showed modality effects on both usability and cooperation. For SUS, the smartphone was highest ($M = 90.8$, $SD = 13.8$), followed by the avatar ($M = 63.3$, $SD = 7.6$) and HoloLens ($M = 55.0$, $SD = 9.0$). A repeated-measures ANOVA indicated a significant difference, $F(2, 4) = 20.8$, $p < .01$, $\eta_p^2 = .91$; pairwise t-tests showed smartphone $>$ HoloLens

($t(2) = -4.51$, $p < .05$), while smartphone vs. avatar ($t(2) = -2.40$, $p = .14$) and avatar vs. HoloLens ($t(2) = 1.65$, $p = .24$) were not significant. For cooperation, means were highest for the avatar ($M = 4.6$, $SD = 0.51$), followed by the smartphone ($M = 4.1$, $SD = 1.02$) and HoloLens ($M = 2.9$, $SD = 0.84$). A repeated-measures ANOVA revealed a significant effect, $F(2, 4) = 32.91$, $p < .01$, $\eta_p^2 = .94$; pairwise tests indicated smartphone $>$ HoloLens ($t(2) = -11.00$, $p < .01$) and avatar $>$ HoloLens ($t(2) = 8.66$, $p < .05$), with no difference between avatar and smartphone ($t(2) = 1.51$, $p = .27$). Tables 5.2 summarizes the usability results.

Table 5.2: Usability (SUS) across modalities. Means are reported as $M \pm SD$.

Condition	**SUS** $M \pm SD$	p **(overall ANOVA)**
2D Video (Smartphone)	90.8 ± 13.8	
Avatar Proxy (Meta Quest)	63.3 ± 7.6	$< .01$
AR Video (HoloLens)	55.0 ± 9.0	

These findings provide formative support **H1.2**, indicating that interactive Q&A via embodied proxy avatars can improve ACCT. While traditional AR video-based methods were easier to use, the avatar proxy provided the clearest sense of co-presence and supported intuitive spatial reference, which contributed to higher perceived cooperation. The ability to record and respond to contextualized messages at specific task steps enabled meaningful temporal linking between participants, demonstrating the potential of immersive proxy-based communication in asynchronous industrial scenarios.

✔ **Summary:** This section formatively evaluated proxy-based asynchronous cooperation across three modalities. Avatar proxies showed the highest cooperation ratings with significant differences relative to AR video, whereas the smartphone baseline was descriptively highest on usability. Taken together, these results cautiously suggest that embodied proxies can strengthen cooperative perception, while familiar video retains advantages in ease of use.

5.4 Asynchronous Coordination

Asynchronous coordination in XR environments tackles the challenge of enabling temporally decoupled yet contextually coherent collaboration. Figure 5.15 presents the decomposition of research question **RQ1.3** into the corresponding hypothesis **H1.3** and measurement criteria **M1.3**.

This question guides the implementation of two technical components: a stepwise message organization system and the immersive Version Control System (iVCS).

5.4.1 Message Navigation and Referencing

Empirical studies suggest that stepwise instruction delivery improves collaborative task performance by reducing information overload and enhancing memory retention [154]. Structuring immersive messages into discrete task steps transforms asynchronous interaction from passive playback into a navigable and task-aligned communication flow. In ACIE, this structure plays a central role in asynchronous coordination, enabling contextual responses, step-specific queries and just-in-time guidance.

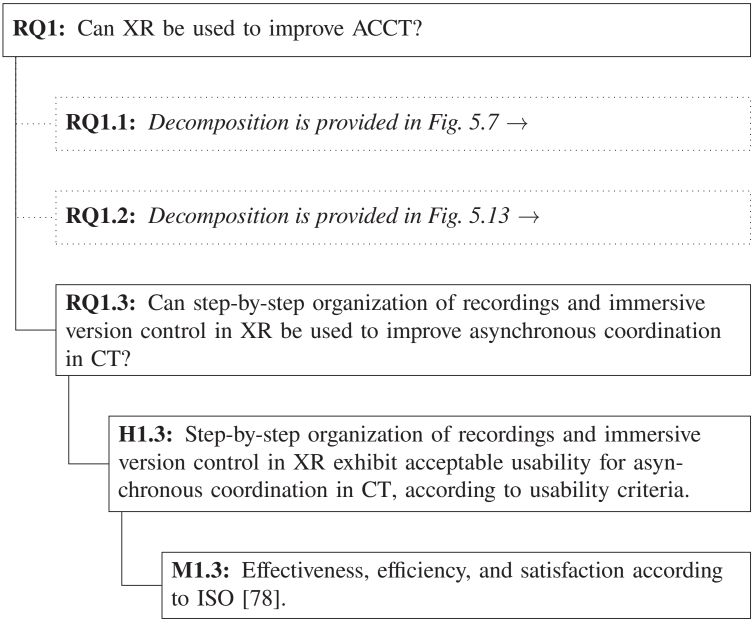


Figure 5.15: Decomposition of **RQ1.3** with corresponding hypothesis and measurements.

This concept is grounded in the overarching inquiry of how XR can support asynchronous coordination in cross-context tasks.

To implement step-aligned messaging (**RQ1.3**), the `State Manager` tracks the user's current instruction step and message state (e.g. idle, playing, recording). State changes are initiated via the `Player Hand Menu` and forwarded to the `Replay Manager` for processing. Figure 5.17 shows menu variations for both roles (novice and expert), as well as status changes during playback and recording.

The hand menu appears when the user performs a pinch gesture (thumb and index finger touch) and disappears when the gesture ends. It presents role-agnostic controls for navigating between steps (`Next`, `Back`) and for triggering playback (`Play`) or recording (`Record`), depending on the current role and message type.

The current step number and message status (e.g. stopped, playing, recording) are displayed at the center of the menu for immediate feedback.

To ensure consistent organization of immersive messages in the file system, a directory-based structure is employed. Messages are grouped by task step, user ID and message type – distinguishing between instructional content, queries and replies:

`[userID][stepID][MessageType].[extension]`

An example directory for the first three steps may look as follows:

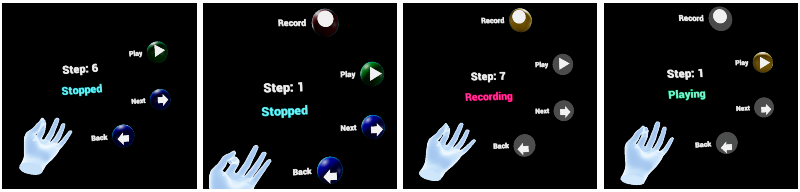
```
step_01/
  Alice/
    Alice_Instruction_01.json
    Alice_Instruction_01.wav
    Alice_Answer_01.json
    Alice_Answer_01.wav
  Bob/
    Bob_Question_01.json
    Bob_Question_01.wav

step_02/
  Alice/
    Alice_Instruction_02.json
    Alice_Instruction_02.wav

step_03/
  Alice/
    Alice_Instruction_03.json
    Alice_Instruction_03.wav
    Alice_Answer_03.json
    Alice_Answer_03.wav
  Bob/
    Bob_Question_03.json
    Bob_Question_03.wav
```

Figure 5.16: Directory structure of asynchronous collaboration session recordings.

This structure ensures that each immersive message is contextually linked to a specific step and user, allowing dynamic message resolution and synchronized playback within the collaborative timeline.



(a) Novice view **(b)** Expert view **(c)** Record step 7 **(d)** Play step 1

Figure 5.17: Hand menu variations depending on user role (novice or expert) and the current state of the replay system (e.g. recording, playing, stopped)

5.4.2 Immersive Version Control System (iVCS)

To support asynchronous coordination and temporal awareness in spatial collaboration, an Immersive Version Control System (ivcs) was designed and implemented using UE 4.27 and later UE 5 [91]. The iVCS system translates established software versioning principles into a spatial and embodied environment.

At its core, the system builds on UE's `SaveGame` component to persist complete scene states, capturing object positions, properties and associated user metadata. These saves act as commits within a branching version tree that reflects the evolving history of a collaborative task. Users can traverse this history spatially within the VR environment, reviewing and selecting past states as needed.

The following Git-like features were implemented in immersive form:

- **Commit (Save):** Each version of the scene is stored as a SaveGame object, saved to disk with a unique identifier and user annotation. This enables reliable rollback points and change tracking.
- **Branching:** Users can create alternative progress paths from any existing version node. Each branch maintains its own linear history, allowing collaborators to explore divergent solutions or revisions.

- **Checkout:** By selecting a node in the version tree, users can load and enter that scene state. Spatial continuity is preserved to prevent disorientation.
- **Merge (Prototype):** Although full spatial conflict resolution was not implemented, the system architecture allows for future development of merge tools. A conceptual UI for merging scene differences was explored, emphasizing visual comparison.
- **Visual Version Tree:** A 3D tree structure presents all versions and branches as spatial nodes. Each node includes a thumbnail preview and metadata (user, timestamp, description), aiding orientation and comparison.
- **Timeline Navigation:** Users can navigate version history using raycasting or teleportation, viewing earlier stages of collaboration as spatially anchored entities.

The version tree interface is integrated directly into the immersive workspace and supports embodied exploration through intuitive gestures and visual feedback. It enables collaborators to understand how a shared artifact evolved over time and to make informed decisions about revisiting, continuing or discarding prior versions. Figure 5.18 illustrates this concept, showing both a sample collaborative scene and its corresponding version tree layout in the immersive environment.

From a CSCW perspective, iVCS enhances shared temporal grounding and makes the collaborative process itself a visible and navigable object of coordination. Its immersive design aligns the mechanics of versioning with the embodied and spatial qualities of XR interaction, thereby bridging the gap between procedural tracking and experiential understanding in asynchronous collaboration.

(a) Collaborative VR workspace **(b)** Version tree of the scene history

Figure 5.18: Immersive Version Control System (iVCS): Users can spatially explore and manage scene states through an interactive version tree (b), anchored in a collaborative virtual environment (a)

5.4.3 Evaluation of Asynchronous Coordination

The usability of stepwise coordination was already co-evaluated in the studies presented in Sections 5.3 and 5.2. Since participants successfully completed all required tasks and usability results were satisfactory, this aspect of Hypothesis **H1.3** can be considered as formatively supported.

The focus in this section is on validating the immersive version control system (**iVCS**) for asynchronous work coordination. A VR prototype was developed with core version control functionalities: *add, commit, branch, merge, push, and pull* – mapped to immersive interactions. Participants ($N = 8$) engaged in a spatial layout task, as illustrated in Figure 5.19. During the task, they received text instructions in VR to create, revisit and merge versions of their layout of the virtual scene, simulating a typical revision workflow.

All participants successfully completed the required versioning operations and coordination tasks. Usability was rated as above average (SUS: $M = 71.9$, $SD = 13.1$); ASQ averaged $M = 6.56$, $SD = 0.53$. As this was a single-condition assessment, no between-condition significance tests were applicable. A majority (75%) preferred the immersive interface over traditional 2D timeline-based alternatives (e.g. Git). Participants specifically praised the spatial preview of revisions and the clarity of visual change differentiation.

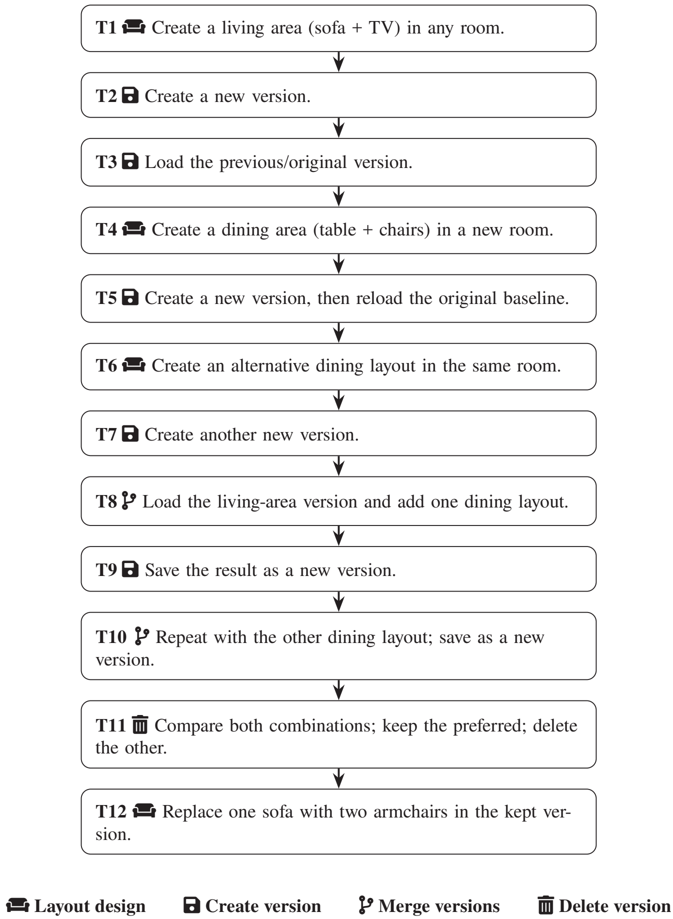


Figure 5.19: iVCS evaluation task: Tasks T1–T12 were performed by each participant using the Oculus Quest 1 HMD

These findings formatively support the iVCS part of **H1.3**, indicating that immersive version control supports immersive work coordination and change awareness. The spatial organization of changes and immersive navigation of version history align with CSCW principles of shared temporal context and activity tracking. The system thus contributes to asynchronous coordination by providing a collaborative workspace history.

> **Summary:** This section formatively evaluated immersive version control (iVCS) as a coordination mechanism for asynchronous spatial collaboration. Participants completed branching and merging tasks reliably; usability was above average (SUS) and post-scenario satisfaction was high (ASQ). Together with evidence from Section 5.4.1, these descriptive results provide cautious support for **H1.3**, indicating that spatial revision tracking and stepwise message organization can facilitate temporal coordination in XR.

5.5 Cross-Reality Collaboration

In industrial environments, collaboration often involves a blend of different realities and immersive tech. To accommodate diverse workflows, ACIE was expanded to integrate mixed-reality settings, embedding virtual elements naturally within physical spaces. This section outlines spatial consistency across sessions and collaborators, performance optimization for standalone headsets which are more convenient in industrial settings, and validation of the system in realistic environments.

5.5.1 Scene Understanding and Spatial Anchoring

Mixed reality was enabled through passthrough AR using the `Meta XR` plugin. This allows interactive virtual avatars or task guidance to be superimposed onto the user's real-world view. The headset's built-in cameras and sensors capture depth information that can be used to occlude virtual elements behind physical objects, although dynamic objects (e.g. people or newly placed equipment) remain more difficult to detect.

Scene understanding refers to the automatic detection and mapping of physical objects in the environment. In this work, Oculus XR Scene capabilities were employed to construct a Scene Model that captures information about walls, floors, and static machinery (e.g. robotic arms). Shared spatial anchors were subsequently used to synchronize the virtual and real environments for all collaborating users, ensuring a stable reference frame for augmented content such as avatar replays or instructional holograms.

In ACIE, spatial consistency across sessions is ensured using a scene anchor object called the `Registrator`. Since the `MetaXR` plugin does not provide direct access to the camera for image-based or marker-based registration, a workaround was implemented using the built-in Room Setup feature. During this setup, users scan the environment by walking around and facing all parts of the room. `MetaXR` then detects structural elements like walls and furniture, overlaying virtual mockups onto the corresponding physical objects. In this approach, a table serves as a stable physical reference point for aligning the MR scene and replayed Immersive Messages, ensuring environmental coherence across devices and over time.

The virtual table object, that is superimposing a physical table, is the root object (`Registrator`) for the virtual scene. All child objects attached to the root will be automatically aligned relative to the table. In each session, the MR system recognizes the position and rotation of the table and transforms the virtual content relatively to it.

5.5.2 Performance Limitations and Optimization

Since the deployment targeted standalone XR hardware (Meta Quest Pro and Meta Quest 3 devices), specific steps were needed to maintain interactive frame rates and smooth avatar playback. These optimizations were essential to mitigate stuttering or lag, especially given the hardware's limited GPU resources compared to a desktop system.

The following rendering and project settings were applied to optimize for standalone performance:

- **Reduced Geometry Detail:** Complex CAD models (e.g. robot arms) were preprocessed to remove hidden geometry and to decrease polygon counts.
- **LOD Tuning:** MetaHuman avatars were configured with reduced Level of Detail (LOD), disabled hair simulation and simpler shading.
- **Scene Quality Tweaks:** Rendering resolution and dynamic shadows were adjusted to achieve a stable 60 FPS whenever possible.

Additionaly, optimizations were implemented in the ACIE framework:

- **Replay Manager Optimization:**
  - **Delta-Time Logic:** Avatar motion and audio were updated based on elapsed time, rather than every frame.
  - **Batch Updates:** Full-body transforms were computed less frequently (e.g. every few frames), aiming for a balance between smooth playback and system overhead.
  - **Early Exits:** Internal checks were added to skip calculations if certain bones or segments had not changed, reducing CPU load.

- **Hierarchical LOD for Avatars:** Body segments outside the user's immediate field of view were rendered at a lower detail or omitted altogether.
- **Synchronization of Audio and Animation:** Additional logic prevented audio from desynchronizing when frame rates dropped, so that voice instructions remained intelligible.

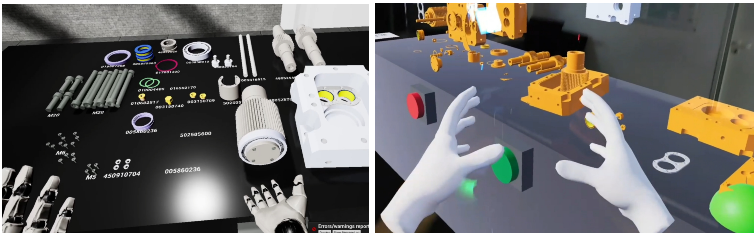
Overall, the PC-based VR build showed little performance strain, but the standalone device required careful trade-offs between visual fidelity and responsiveness. Frame rates around 60 FPS were targeted to maintain immersion, with the understanding that further dynamic optimization (e.g. adaptive resolution scaling) could be used if necessary.

5.5.3 Evaluation of Mixed Reality Modalities

To validate the feasibility and usability of ACIE in real-world mixed-reality deployments, the system was tested across both AR and VR modalities with a focus on asynchronous guidance. In a technical validation, assembly instructions including hand motion and part manipulation were recorded in VR and later imported and replayed on the HoloLens AR device [136]. Likewise, instructions recorded in AR could be imported and played back in VR on a Meta Quest 2 headset.

Figure 5.20 illustrates the shared assembly environment as experienced in both VR and AR.

To enable this interoperability, a mapping function was implemented that translates between controller-based hand animations in VR and hand-tracked gestures in AR. This ensured consistency of gesture representation across modalities and preserved the quality of instruction delivery.



(a) Cross-reality assembly application in VR **(b)** Cross-reality assembly application in AR

Figure 5.20: Cross-reality assembly application, allowing the recording of assembly instruction in VR and playback in AR and vice versa

Performance Evaluation on Standalone MR Headset

Furthermore, the system's runtime performance was tested on the Meta Quest Pro headset in a MR condition during avatar recording and playback. The objective was to maintain consistent interactive frame rates (60 Frames Per Second (FPS)) without sacrificing fidelity or spatial accuracy.

Figure 5.21 shows a representative trace of the system's runtime performance during a mixed-reality session, captured using the Meta Quest Performance Analyzer.

While a 60 FPS target was used as a practical benchmark, actual frame rates varied between 41 and 65 FPS during avatar playback in a MR environment. The most noticeable drops occurred during activation and deactivation of immersive message playback – specifically, when loading recordings from internal SD storage or exporting captured data to the local file system. Despite these short-term dips, the experience remained usable throughout, with no major disruptions in interaction or playback reported by users. However, the system did not fully meet the Meta Quest Pro's native refresh rate of 72 FPS, underlining the trade-off between visual fidelity and real-time responsiveness on standalone XR hardware. Future improvements could mitigate these drops by introducing preloading or buffering mechanisms to avoid file I/O bottlenecks during critical playback operations.

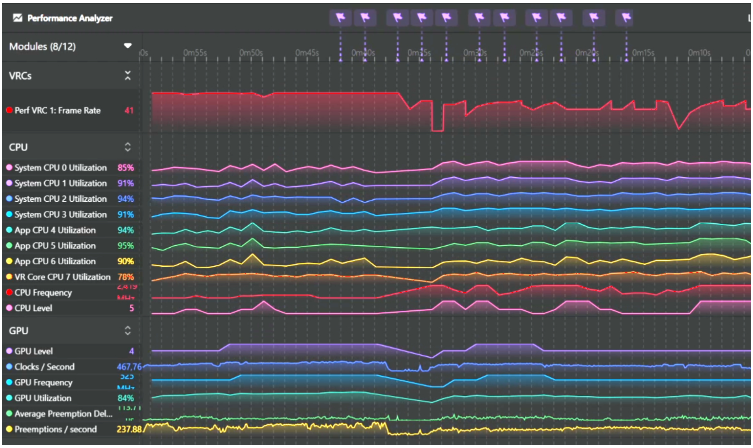


Figure 5.21: Performance profile of the ACIE system during mixed reality (MR) operation on a Meta Quest Pro. The top curve (in red) shows the frame rate, fluctuating between 41–65 FPS. The middle group of curves depicts CPU thread utilization across both system and application cores. The bottom curves represent GPU activity and scheduling performance.

5.6 Summary and Outlook

This chapter presented the ACIE framework and its implementation – a modular and extensible concept designed to support asynchronous collaboration on complex engineering tasks in XR. Grounded in the theoretical foundations from **Chapter 2** and the limitations identified in **Chapter 3**, the framework systematically addresses the three pillars of asynchronous collaboration: communication, cooperation and coordination, to address **RQ1** (see **Chapter 1**, Fig. 1.1).

Answer to RQ1.1: Asynchronous communication was enabled through the combination of Immersive Messages, Proxy Playback, and the Immersive Communication Protocol (ICP). These modules allowed users to record, transmit and replay embodied task instructions across time and devices. Formative results suggested that this approach is usable for immersive asynchronous collaboration across several realities – AR, VR and MR – formatively supporting hypothesis **H1.1**[10].

Answer to RQ1.2: Continuous cooperation was realized via role-aware interaction with Interactive Proxy Avatars, supporting structured Q&A exchanges and temporal linkage between collaborators. The system enabled spatial anchoring of questions and deferred responses, improving perceived cooperation and mutual understanding. These findings formatively support hypothesis **H1.2** after successful validation tests[10].

Answer to RQ1.3: Coordination mechanisms were implemented through stepwise task structuring and the Immersive Version Control System (iVCS). This system allowed users to explore, branch and merge spatial task states, supporting flexible revision workflows. Both concepts were usable across several studies, formatively confirming hypothesis **H1.3**[10].

Asymmetric Collaboration: Although not the central focus, the implementation considered spatial registration and modality alignment to ensure consistent cross-reality interaction between AR, VR, and MR users. These capabilities served as foundational enablers to support interoperability across XR modalities.

To assess the design hypotheses introduced in ❯ **Chapter 4**, each core module was evaluated individually during implementation. The results of these formative validations are summarized in Table 5.3.

[10] Formative evidence based on small samples and descriptive statistics only; no inferential statistics were performed.

Table 5.3: Validation of XR-related hypotheses in the Design Cycle based on implemented system modules and formative results

Hypothesis	Result	Evidence
H1.1.1	✔	Avatar capture and proxy playback were successfully implemented and tested to support embodied, asynchronous task guidance using the Immersive Messaging workflow.
H1.2.1	✔	Interactive proxies with role-specific functions enabled contextual Q&A and asynchronous cooperation, and were successfully validated in an expert–novice scenario.
H1.3.1	✔	Users were able to successfully navigate through recorded task sequences using the stepwise structure of recordings.
H1.3.2	✔	The iVCS component enabled branching, merging, and resuming of specific virtual scene states, supporting asynchronous coordination over time.

✔ **Formative support** (small samples; no inferential tests)
✖ **Not supported**

Answer to RQ1 (synthesis) Taken together, the formative implementation and module-level evaluations indicate that XR can be used to support asynchronous collaboration on complex tasks: immersive messaging with proxy playback, role-aware interactive proxies for Q&A, and an immersive version control system were usable across AR, VR and MR and supported communication, cooperation and coordination in deferred workflows.[11]

These results confirm that the core concepts of the ACIE framework are not only theoretically grounded but also technically feasible and usable in practice. The next chapter builds on this foundation to assess the integrated system in an applied industrial scenario.

[11] Formative evidence based on small samples and descriptive statistics only; no inferential statistics were performed.

In Simpler Terms

In essence, the ACIE framework allows engineers to leave behind contextualized, embodied traces of their work – including voice, movement and spatial annotations – which others can later explore, respond to and build upon. These contributions form a persistent timeline of collaboration that is spatially grounded, temporally flexible and intuitively navigable.

How to Proceed

This implementation chapter details the technical realization of the ACIE framework. Building on the conceptual foundations, the individual modules are developed, integrated, and formatively evaluated to ensure feasibility, usability, and domain alignment.

➔ **Chapter 6** applies the complete framework in a realistic robotic cell use case, demonstrating how ACIE supports asynchronous collaboration between expert and novice operators. The holistic evaluation in that chapter assesses system effectiveness in practice.

6 Evaluation

In line with the relevance cycle of the Design Science Research (DSR) methodology, this chapter presents the evaluation of the developed XR-based framework for asynchronous collaboration. The objective of this evaluation is to assess the system's effectiveness in a real-world engineering context and to determine its impact on key aspects of collaborative performance.

To this end, a comparative user study was conducted in a robotic production environment, contrasting the immersive XR-based approach with a conventional video-based workflow. The study focused on spatially procedural tasks within a frequently reconfigured manufacturing cell, offering a representative scenario for evaluating task continuity, procedural clarity and user experience in asynchronous collaboration.

The chapter is structured as follows: ❯ Section 6.1 introduces the practical use case and its industrial setting. ❯ Section 6.2 describes the evaluation design, including the experimental setup, participants and metrics. The results of the comparative study are presented in ❯ Section 6.3, supported by quantitative and qualitative findings. Finally, ❯ Section 6.4 summarizes the key outcomes and prepares the ground for the integrated discussion and conclusion in ❯ **Chapter 7**.

6.1 Practical Field Use Case

To evaluate the developed system in a realistic industrial context, the *Karlsruhe Research Factory for AI-integrated Production*[1] was selected as the practical testbed. Within this environment, a robotic manufacturing cell based on the concept of *Value Stream Kinematics (VSK)* was used. The VSK system, as depicted in Figure 6.1, is designed to support high flexibility in manufacturing by combining programmable industrial robots, modular tooling and intelligent planning software. It enables the execution of a wide range of production processes – including assembly, quality assurance and machining – using universal kinematic components such as jointed-arm robots.

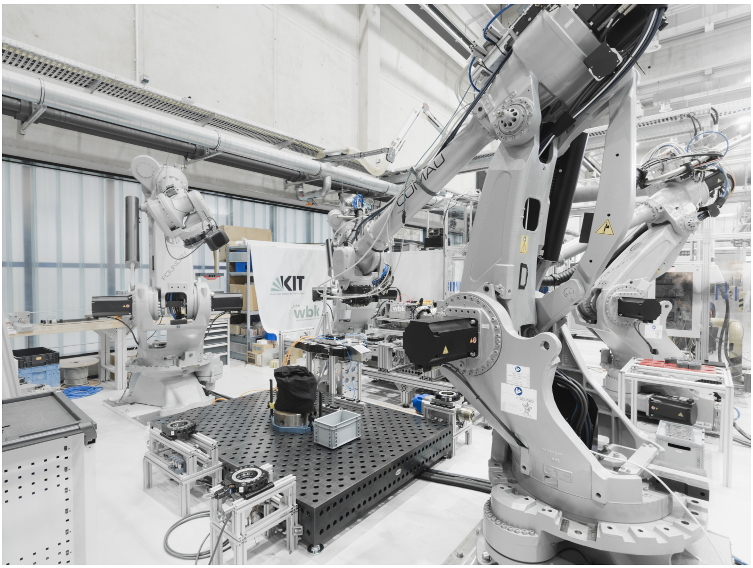


Figure 6.1: The Value Stream Kinematics (VSK) evaluation environment – a flexible production system at the Karlsruhe Research Factory

[1] `https://www.karlsruher-forschungsfabrik.de/en.html`

In the evaluated setup, two mirrored robotic units form the manufacturing cell. Each unit includes two large robot arms, a central matrix platform for tool and workpiece handling and adjacent areas for manual interaction. A laser barrier separates the two sides, enabling parallel or independent operation. Peripheral systems, such as control terminals, storage cabinets and computing stations, support the orchestration of both manual and automated tasks.

This environment is subject to continuous evolution. The VSK is frequently reconfigured to support new research projects and demonstrations, often with little or no updated documentation. This dynamic nature imposes cognitive and procedural challenges on the teams operating the cell. Engineers, researchers and student workers must frequently adapt to undocumented changes in layout, tools, components and procedures. Especially the safety-critical commissioning processes must be revisited and adapted with each reconfiguration – tasks that require detailed knowledge of the physical setup, process steps and current system state.

These conditions present an ideal application domain for asynchronous collaboration support. In practice, domain experts (e.g. researchers or supervisors) and less experienced users (e.g. student workers or new operators) rarely work on-site at the same time due to part-time schedules, academic duties or staff turnover. While conventional documentation exists, it often lags behind reality and is insufficient for onboarding or troubleshooting. As a result, operational efficiency suffers, and knowledge gaps persist across personnel rotations.

The developed XR-based asynchronous collaboration framework addresses these challenges by enabling spatially anchored, embodied communication across time. Using immersive technologies, experts can capture and convey process-relevant knowledge – such as reconfiguration steps or commissioning procedures – which can be consumed later by other users directly in context. This supports just-in-time learning, increases awareness of changes and reduces the need for synchronous instruction. The field evaluation investigates how this approach compares to traditional video-based documentation in terms of spatial understanding, procedural clarity, perceived co-presence and user engagement.

6.1.1 Commissioning and Safety Procedure

Before operating the VSK system, a standardized commissioning and safety procedure must be performed. Unlike a continuously running production facility, the research cell is often idle or in standby mode between experiments. Hence, the system must be explicitly initialized before each use, including powering up essential components, configuring physical elements and verifying system readiness and safety compliance. Some operations of the commissioning process in the VSK are depicted in Figure 6.2.

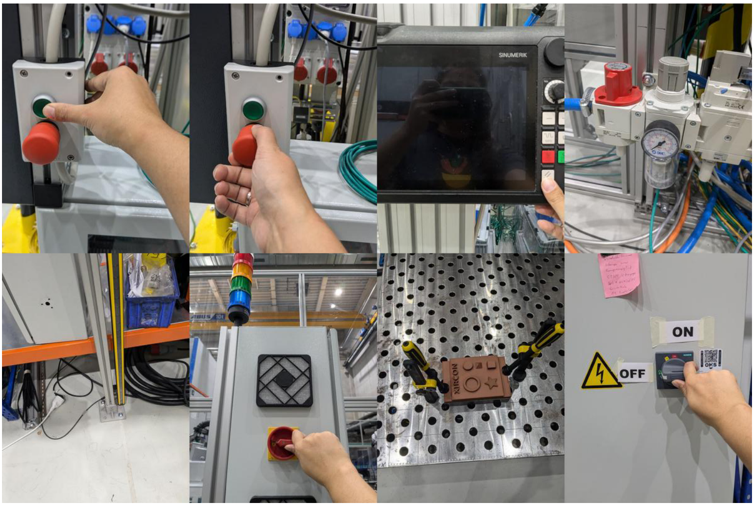


Figure 6.2: Examples of commissioning steps performed before operating the VSK system

The commissioning process consists of the following core steps:

1. **Safety entry:** Activate the external safety switch, then enter the robot cell.

2. **System power-up:** Turn on the robot cabinet and the external network cabinet. This initiates the system boot sequence.

3. **Matrix setup:** Physically configure the matrix platform by mounting the correct tool part (e.g. the blue part), aligning it precisely at the designated grid coordinates and securing it against unintentional movement. Loose items must be removed from the workspace.

4. **Pneumatic check:** Ensure that the pneumatic system is pressurized by confirming that the pressure display reads above zero.

5. **Laser barrier test:** Verify the correct operation of the safety laser barrier that separates the mirrored robot areas. The indicator lights should be green when unobstructed and turn red when triggered by an object.

6. **Control panel reset:** After confirming system readiness, operate the control panel: press the reset button, observe the green light signal and place the panel outside the cell.

7. **Safety door activation:** At the door, engage the internal safety switch and press the green start button, which should begin flashing to indicate successful activation.

8. **Robot drive enablement:** On the control panel interface, switch to the user mode and activate the drives via the *“Antriebe An/Aus”* command. An audible *“click”* confirms readiness. The system is now fully operational.

This process is essential to ensure safe, consistent operation of the VSK and must be adapted whenever changes to the system configuration occur. In practice, however, its execution is highly dependent on accurate, up-to-date knowledge of the physical setup and the specific components involved. As such, it served as a representative scenario for evaluating the proposed XR-based asynchronous collaboration system in the user study.

6.1.2 Challenges in the Existing Workflow

The commissioning and reconfiguration of the VSK cell regularly require knowledge transfer between experienced researchers and less experienced operators, such as student assistants. Any changes to the setup – such as newly integrated tools, modified safety procedures or technical adjustments – must be communicated clearly and verified prior to operating the system. These include safety-critical elements like verifying the laser barrier or pneumatic pressure as well as the correct configuration of the matrix platform. Typically, this handover takes place synchronously and on site: experts and operators meet at the factory to collaboratively review the changes and ensure the commissioning process is executed correctly.

However, this synchronous collaboration model presents several limitations:

- *Scheduling bottlenecks:* Assistants often need to wait for experts to become available before critical tasks can proceed.
- *On-site dependency:* The nature of the collaboration is highly spatial and physical, requiring stakeholders to be present in the factory.
- *Static documentation:* Existing documentation (e.g. PDF protocols) was often outdated or incomplete, particularly for safety-critical tasks. In fact, some procedures changed during the study itself, highlighting the need for more adaptable and up-to-date communication methods.

These operational challenges reflect a broader structural issue within engineering workflows. Currently, there is a lack of systematic methods for planning and coordinating complex procedures, such as production cell reconfiguration, in an asynchronous and spatially contextualized manner. While machines and sensors are increasingly represented in digital twins, human actions – especially those involving tacit knowledge or situational judgment – are rarely integrated into digital systems. As a result, valuable experiential knowledge is lost between personnel rotations, making onboarding and troubleshooting more difficult.

This work addresses that gap by exploring whether immersive, spatially embedded XR technologies can offer an effective medium for asynchronous collaboration in this context. The central focus is to examine whether spatial modalities – such as in-situ guidance, recorded procedures and embodied annotations – improve communication, coordination, and cooperation in comparison to conventional approaches like video documentation. The following chapter presents an empirical study that evaluates this approach by comparing an XR-based asynchronous workflow with a traditional video-based method in a real-world commissioning task.

6.1.3 Integrating Asynchronous Collaboration

To address the challenges of knowledge transfer and procedural variability in the reconfiguration of the VSK robotic cell, the ACIE framework was deployed and evaluated in a realistic industrial setting. The objective was to investigate whether immersive, spatially embedded methods of asynchronous collaboration could improve understanding and execution of tasks, compared to conventional approaches such as *2D Video* documentation.

In this use case, the ACIE system was applied to the VSK system's safety-critical commissioning procedure. Using mixed reality headsets, domain experts recorded spatial guidance and task walkthroughs in situ, which were later experienced by other users in the same physical context. The system supported asynchronous instruction via full-body avatar replays, embedded media messages and anchored object interactions – enabling distributed users to collaborate across time without being co-located. Figure 6.3 gives a CAD-based, top-down plan of the VSK cell used in the study.

To evaluate this approach systematically, a comparative user study was conducted. Participants completed identical commissioning tasks under two conditions: a baseline condition using traditional *2D Video* guidance and a test condition using the immersive ACIE system. The experimental evaluation focused on the two main perspectives, usability and presence.

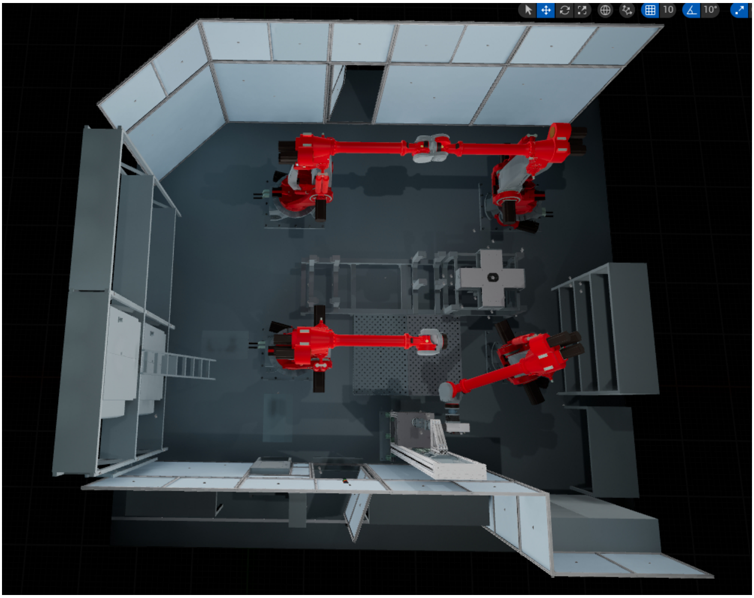


Figure 6.3: Top-down CAD view of the VSK cell

The structure of the research questions, hypotheses and measurement methods is illustrated in Figure 6.4 in a goal-question-metric (GQM) model to map overarching research interests into testable elements.

RQ2.1 presents the usability-focused decomposition. Here, usability is assessed along three dimensions: efficiency (e.g. time and completion), effectiveness (e.g. task success, error rate) and satisfaction (e.g. user experience ratings). Each of these aspects is measured using a combination of objective performance data and subjective feedback via standardized questionnaires such as SUS and NASA TLX. Therefore, this figure outlines the internal structure used to investigate how the ACIE system performs in practical task execution.

RQ2.2 in turn, addresses the second evaluation perspective: presence. This includes both spatial presence – the extent to which users feel immersed and

RQ2: How does ACCT perform with XR compared to non-immersive representations?

- **RQ2.1:** How does XR ACCT perform in terms of usability (effectiveness, efficiency, satisfaction) compared to non-immersive representations?
  - **H2.1:** In ACCT, XR increases effectiveness and user satisfaction relative to non-immersive representations, with a trade-off in efficiency.
    - **M2.1:** Effectiveness (task accuracy/completion), efficiency (completion time, error rate), satisfaction/usability (ASQ, SUS, UEQ); workload (NASA-TLX).
- **RQ2.2:** How does XR ACCT perform in terms of presence compared to non-immersive representations?
  - **H2.2:** Spatial and social presence are higher in XR ACCT than in non-immersive representations.
    - **M2.2:** Spatial presence (IPQ: REAL, SP, INV) and social presence/co-presence (NMSPI); collaboration ratings where applicable.

Figure 6.4: Hierarchical overview for **RQ2** (Usability and Presence): research questions, hypotheses, and associated measurements.

situated in the task environment – as well as social presence – the perceived connectedness with others, even in asynchronous collaboration. These dimensions are assessed through targeted instruments such as the IPQ (Igroup Presence Questionnaire) and NM-SPI (Networked Minds Social Presence Inventory), complemented by task-specific questions. This part of the evaluation examines whether immersive asynchronous collaboration fosters a stronger sense of embodied interaction compared to conventional methods.

This provides a structured map of how the ACIE framework is evaluated and allows a systematic investigation of whether and how immersive asynchronous methods can improve collaborative task execution in flexible, safety-critical production environments.

6.2 Study Design

To systematically evaluate the ACIE framework, a controlled experiment was conducted in a realistic industrial environment to investigate the effectiveness of asynchronous collaboration using immersive technologies. The goal was to compare immersive asynchronous collaboration with a conventional video-based approach under controlled conditions, but in a realistic environment. The following section outlines the experimental design, underlying rationale, and methodological choices that guided the study setup.

6.2.1 Experimental Conditions

Two asynchronous collaboration methods were compared in a between-subject design – *MR Avatar* and *2D Video*.

MR Avatar (Immersive): Participants used a Meta Quest 3 headset to follow spatially embedded guidance provided by a recorded avatar in mixed reality.

Figure 6.5: Study task showing three instruction steps (top) presented by the MR avatar and participants (bottom) following the avatar's guidance

This guidance was rendered directly into the robot cell environment. Figure 6.5 shows three example instruction steps (top row) and corresponding participants (bottom row) interacting with the MR avatar in-situ.

2D Video (Baseline): Participants used a handheld smartphone to follow conventional video-based instructions recorded from a first-person perspective. This condition served as a non-immersive baseline for asynchronous task guidance (see Figure 6.6).

6.2.2 Participants

Participants were recruited from the local university campus, primarily via mailing lists from student dormitories and academic departments. Most were students or visiting researchers from diverse engineering disciplines. As no prior experience with robotics, industrial systems or XR technology was required, this was explicitly communicated during recruitment, demographics questions

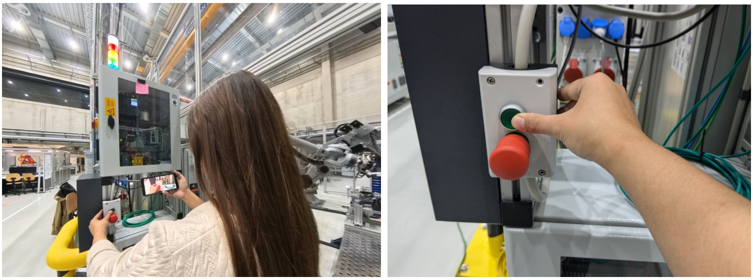


Figure 6.6: Participant following asynchronous task instructions in the 2D Video condition: using a handheld smartphone (left) and the video instruction showing the current step (right).

Table 6.1: Analyzed sample size and gender distribution by condition

Condition	N	**Male** n **(%)**	**Female** n **(%)**
Headset with Avatar	22	15 (68.2%)	7 (31.8%)
Smartphone Video	21	15 (71.4%)	5 (23.8%)
Total	**43**	**30 (69.8%)**	**12 (27.9%)**

asked are listed in **Appendix A**, Table A.6. Participation in the experiment was voluntary. Figure 6.7 shows the gender-by-condition split, and Table 6.1 summarizes counts.

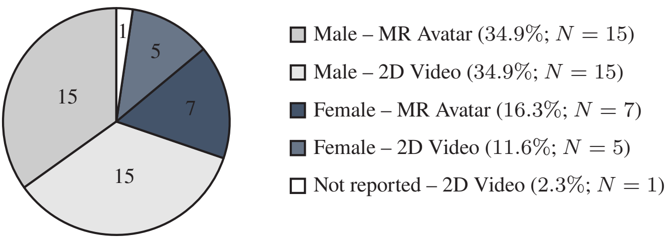


Figure 6.7: Participant demographic distribution by gender and condition

In total, $N = 46$ participants were recruited. Three participants (all male) were excluded from all analyses due to technical problems in the robot cell that required procedural changes. The analyzed sample therefore comprised

$N = 43$ participants (*MR Avatar* $N = 22$; *2D Video* $N = 21$). One participant in the *MR Avatar* condition discontinued after Part 1 due to cybersickness; data from Part 1 were retained, whereas Parts 2 and 3 were excluded. In addition, one participant did not understand the task in Part 2; no data were acquired for that part from this participant (excluded from Part 2 only). One participant did not report gender and was therefore omitted from the gender counts but retained in all other analyses. Figure 6.8 summarizes the mean (1–5) self-ratings for technology familiarity and usage, gaming, XR, and robotics experience, as well as comfort with new technology, broken down by condition.

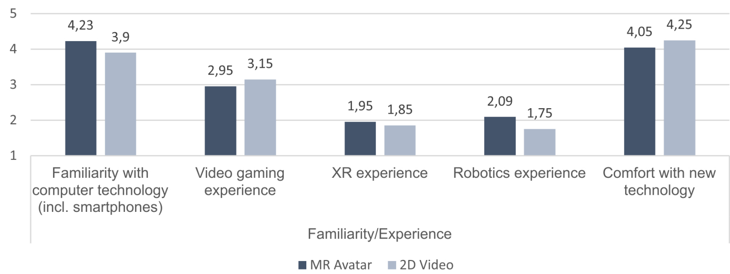
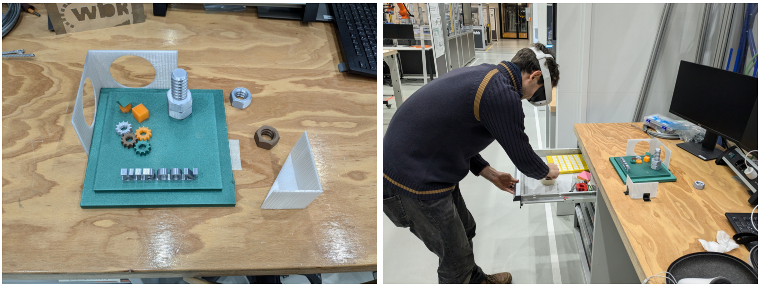


Figure 6.8: Technology familiarity, experience, and comfort level

6.2.3 Experimental Procedure

The study was structured into three main parts, each reflecting a distinct phase in the asynchronous collaboration workflow:

1. **System Training and Questionnaires:** Each participant first completed a pre-questionnaire (Q1), followed by a brief training task (T0) illustrated in Figure 6.9. This phase ensured that participants understood how to use the assigned technology, the nature of asynchronous collaboration in the experiment, and how to interact with the non-present collaborator through the system.



(a) 3D-printed assembly task used for training **(b)** Participant training with the MR Avatar

Figure 6.9: Training scenario used to familiarize participants with the experiment conditions and asynchronous workflow

2. **Novice Execution – Part I:** Participants performed a robot commissioning task using either the *MR Avatar* or the *2D Video* system (T1). Some steps intentionally included errors or incomplete instructions to trigger asynchronous problem reporting.

3. **Novice Execution – Part II:** After a short break, participants resumed the task for a second attempt, now with access to expert responses to the issues reported earlier (T2). During the break, participants completed a mid-experiment questionnaire (Q2). After finishing the task, they were asked to complete the post-task questionnaire (Q3).

4. **Expert Execution – Part III:** In the final round, participants switched roles and acted as experts. They reviewed five pre-recorded issue reports and were asked to record responses that addressed each problem (T3). A final questionnaire (Q4) concluded the experiment session.

The order of tasks and questionnaires is visualized in Figure 6.10, which illustrates the full experimental flow from participant onboarding to final evaluation.

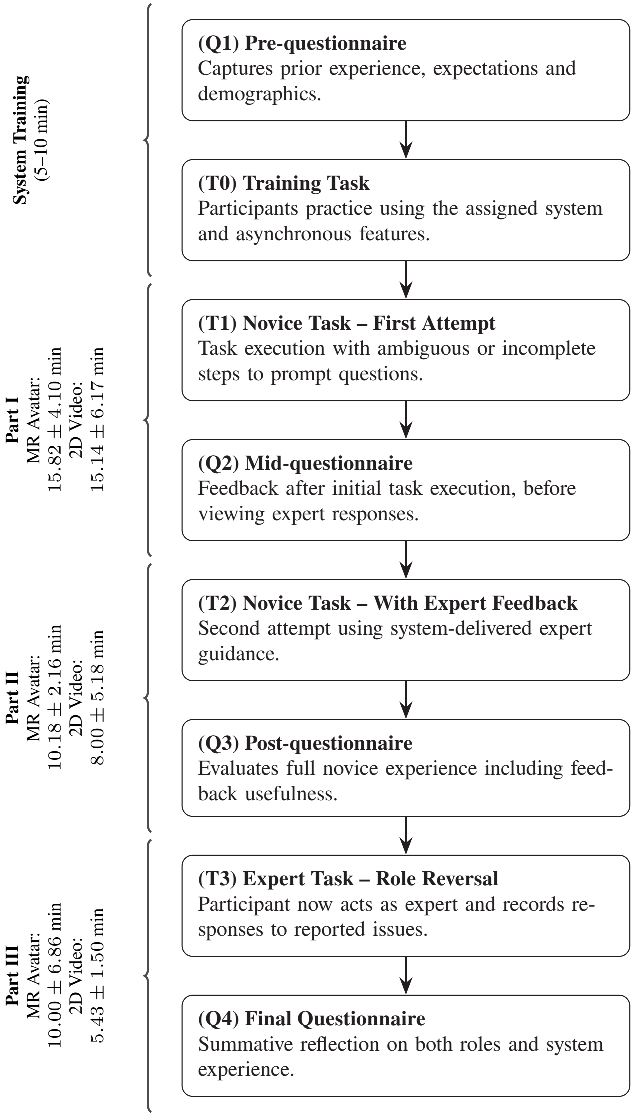


Figure 6.10: Overview of the experimental procedure, including the sequence of tasks (T0–T3) and questionnaires (Q1–Q4). Left brace labels include per-phase completion times ($M \pm SD$, minutes) for *MR Avatar* and *2D Video*; training was constrained to 5–10 minutes.

6.2.4 Task and System Setup

The tasks, as illustrated in Figure 6.11, were derived from the actual safety and commissioning procedures used in the VSK robot cell. These included entering the cell safely, verifying component states and unlocking the robot system. In total, eight subtasks were defined (ten in the MR condition, where two were split for technical reasons). Four tasks intentionally omitted critical information to prompt participants to submit problem reports, thereby simulating realistic uncertainties in industrial workflows. In the expert role, participants had to provide asynchronous responses to these types of issues.

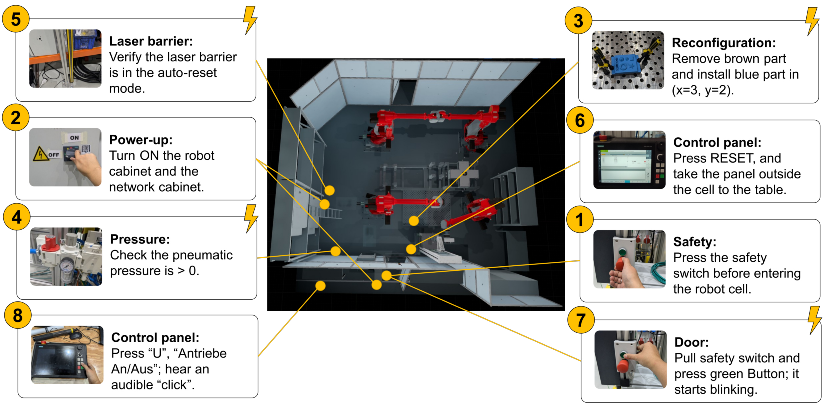


Figure 6.11: Experiment task: participants conducted steps 1-8 to start up the production system, whereby tasks 3, 4, 5, 7, and 8 had intended errors to trigger asynchronous assistance using the Q&A function.

Each condition was presented with the same core instructions, differing only in their delivery modality. The MR system included immersive features such as spatial anchoring and avatar embodiment, while the *2D Video* system was limited to standard playback without interactivity or depth cues.

6.2.5 Ethical Considerations

All participants provided informed consent prior to the study. They were made aware of their rights, including the ability to withdraw at any time without penalty. All data were anonymized in accordance with GDPR guidelines.

6.2.6 Measurements

In order to thoroughly assess the effects of the ACIE framework, a multi-faceted measurement strategy was utilized. The research investigated outcomes related to tasks, including efficiency and accuracy, as well as subjective experiences such as user satisfaction, cognitive load, and perceived presence. This section outlines the metrics and tools employed to evaluate these aspects, setting the stage for the following analysis and discussion of the findings.

Efficiency

To analyze the efficiency of asynchronous guidance, the total time taken by participants to complete the guided tasks was measured. For each task, the start time and end time were recorded, and the total duration was calculated as:

$$\Delta t_{\text{task}} = t_{\text{end}} - t_{\text{start}} \tag{6.1}$$

As participants were given the freedom to decide their approach to the guidance, measuring time for individual subtasks was deemed neither possible nor necessary. The instructional video could either be watched in full before starting the task or viewed simultaneously while performing the task. Because of variations in interaction styles, time per subtask would not have yielded a meaningful metric.

Additionally, precise logging of interaction events across different applications was prevented by technical constraints. In the *2D Video* condition, a switch was made by participants between a messaging app (for guidance) and a camera app (for execution), whereas in the *MR Avatar* condition, interactions occurred within a single application. Given these differences, the total time per task was measured and related to the number of correctly completed tasks, as described in the accuracy analysis.

As participants were free to rewatch instructions and execute tasks at their own pace, a strict correlation between time and accuracy is not expected. However, an analysis is conducted to determine whether more accurate task completions were associated with greater efficiency and whether different conditions influenced the time-accuracy trade-off.

Effectiveness

The effectiveness of the conditions regarding asynchronous guidance is determined by measuring the accuracy of the participants' solutions, or in other words, the completeness of their solution and the error rate. The scoring system used to evaluate the accuracy of the tasks is summarized in Table 6.2, and the method for determining the overall accuracy score per participant and condition is described in equation 6.2.

Table 6.2: Scoring system used to assess task accuracy

Outcome	Points	Description
✔ Completed as expected	2	Followed instructions correctly or correctly reported errors when required.
⚠ Partially completed	1	The task was completed partially or in an unintended, yet still correct, way (e.g., an error was fixed instead of reported, or an issue was reported where none were expected).
✖ Not completed or incorrect	0	Completely incorrect execution or failure to report an issue despite the need for assistance.

Each task is assigned a score based on how well it was completed. Participants receive 2 points if they completed the task exactly as instructed, 1 point if they completed it in an unexpected but still correct way or only partially, and 0 points if they did not complete the task or completed it incorrectly.

In some cases, a task could not be attempted due to a missing prerequisite action in an earlier phase, leading to a *"Not Applicable"* (NA) outcome. To ensure fairness, NA tasks are excluded from the total task count rather than being assigned a score of 0. This means that a participant's accuracy score is only calculated based on tasks they had the opportunity to complete, making comparisons between participants more meaningful.

The final accuracy score is computed as the sum of earned points divided by the sum of possible points for all attempted tasks (excluding NA), expressed as a percentage:

$$\text{Accuracy Score} = \left(\frac{\sum \text{Earned Points}}{\sum \text{Total Possible Points (excluding NA)}} \right) \times 100 \tag{6.2}$$

The study employed a diverse set of standardized and custom questionnaires to evaluate different dimensions of system interaction, including usability, cognitive load, user experience, and presence, which are summarized in Table 6.3.

ASQ: The After-Scenario Questionnaire (ASQ) is a widely used measure for assessing perceived ease of task completion in human-computer interaction studies [98]. It evaluates task difficulty, efficiency and support requirements, providing insights into the usability of different interaction methods. Higher ASQ scores indicate a more positive user experience. Here the ASQ was used additionally to the SUS and UEQ to specifically assess user satisfaction.

SUS: The System Usability Scale (SUS) is a standardized questionnaire designed to measure the perceived usability of a system [22]. It consists of ten

Table 6.3: Overview of questionnaires used to evaluate system usability, user experience, presence, collaboration, and participant demographics

Questionnaire	Purpose	Key Dimensions
ASQ [98]	Task ease and satisfaction	Task difficulty, efficiency, support requirements
SUS [22]	System usability	Ease of use, complexity, learnability
NASA-TLX [70]	Workload assessment	Mental, physical, temporal demand; effort, performance, frustration
UEQ [93]	User experience evaluation	Attractiveness, efficiency, perspicuity, dependability, stimulation, novelty
IPQ [138]	Sense of presence in XR	Spatial presence, involvement, realism
NMSPI [15]	Social presence in collaboration	Co-presence, attentional engagement, emotional contagion, comprehension, behavioral interdependence
Collaboration A.7	Effectiveness of collaboration	Communication, cooperation, coordination
Demographic A.6	Background information	Age, gender, education, XR/robotics experience

items evaluating aspects such as ease of use, complexity, consistency and learnability. SUS provides a single usability score ranging from 0 to 100, with higher scores indicating better usability. In this study, the SUS was used to assess how intuitive and accessible the asynchronous collaboration methods were for participants.

NASA-TLX: The NASA Task Load Index (NASA-TLX) is a widely used tool for measuring subjective workload [70]. It evaluates six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. Participants rate each dimension on a scale from 0 to 100, with higher values indicating higher workload. In this study, NASA-TLX was employed to compare cognitive and physical task load across the asynchronous collaboration conditions.

UEQ: The User Experience Questionnaire (UEQ) is a comprehensive tool for measuring user experience across six dimensions: Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation and Novelty [93]. Each dimension captures different aspects of usability and user satisfaction, providing a holistic evaluation of user experience. Higher scores indicate a more positive perception. In this study, the UEQ was used to assess the subjective experience of participants using different asynchronous collaboration methods.

IPQ: The Igroup Presence Questionnaire (IPQ) is designed to measure the sense of presence in virtual environments [138]. It consists of three primary dimensions: Spatial Presence (SP), Involvement (INV) and Realism (REAL). Higher scores indicate a stronger sense of being "inside" the virtual environment and engaged in the experience.

NMSPI: The Networked Minds Social Presence Inventory (NMSPI) assesses the perceived social presence in mediated interactions [15]. It measures dimensions such as co-presence, perceived attentional engagement, perceived emotional contagion, perceived comprehension and perceived behavioral interdependence. This questionnaire helps to determine how closely connected users feel to those around them in collaborative settings.

Collaboration Questionnaire: A custom Collaboration Questionnaire was designed to assess communication, cooperation and coordination aspects of the collaborative experience (❯ **Appendix A.7**). The items evaluate clarity of instructions, effectiveness of information exchange, responsiveness of the system and ease of coordinating tasks in an asynchronous environment.

Demographic Questionnaire: The Demographic Questionnaire collected background information, including age, gender, education level and technology experience (❯ **Appendix A.6**). Additionally, it included items related to familiarity with digital devices, experience with virtual and augmented reality, prior exposure to robotics and potential discomfort experienced during the experiment.

6.3 Results

In this section, the results of the experiments are presented. The findings are systematically analyzed to provide insights into the observed phenomena.

The collected data, including questionnaire responses, observational notes, and time measurements, was preprocessed and prepared for statistical analysis. This involved exporting the data into Excel, organizing it by questionnaire and condition group, converting responses into numerical format, and inverting scores for negatively worded items where necessary.

Statistical tests were used to determine if the differences between the groups were significant or just occurred by chance, allowing for a reliable assessment of the study's hypotheses and ensuring any observed results are due to the experiment itself. More information on the statistical tests and the procedure can be found in the ❯ **Appendix A.4**, Table A.5 and Figure A.4.

6.3.1 Usability and Workload Results

The evaluation results for the perceived usability and workload are presented in this section. By analyzing these aspects, the aim is to understand whether the perceived ease of use or mental effort is affected by the immersive *MR Avatar* condition compared to the conventional *2D Video* condition.

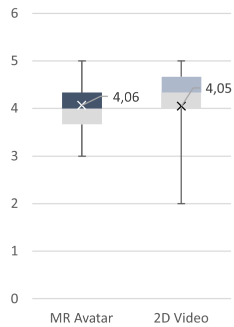


Figure 6.12: Average ASQ scores for MR Avatar and 2D Video conditions

User Satisfaction

The analysis of ASQ scores revealed no significant difference between the *MR Avatar* condition ($M = 4.061$, $SD = 0.521$) and the *2D Video* condition ($M = 4.048$, $SD = 0.825$), $t(41) = 0.52$, $p = 0.604$, $r = 0.079$.

The very small effect size ($r = 0.079$) suggests that the difference in perceived ease of task completion was negligible. This indicates that participants found both methods equally effective in providing task guidance, and that the introduction of immersive MR elements did not introduce any additional usability burden compared to traditional *2D Video* instructions. As depicted in Figure 6.12, the high ASQ scores in both conditions suggest that the overall task structure was well-designed and that participants were able to complete the tasks efficiently, regardless of the method of instruction.

Discussion: These findings suggest that the *MR Avatar* method, despite its immersive nature, did not significantly alter the perceived ease of task completion compared to the *2D Video* method. This aligns with previous research indicating that while immersive elements can enhance engagement and presence, they do not necessarily impact usability unless they introduce complexity or hinder interaction [154, 156]. Related work suggests a higher user satisfaction with immersive interaction over 2D [73, 154, 156]. The lack of a difference in ASQ scores suggests that both methods provided clear, structured and effective task guidance.

Future research could explore whether variations in task complexity or additional interactive features in the *MR Avatar* condition influence usability perceptions [36, 154]. Additionally, incorporating user-adaptive elements, such as gaze-based interaction or personalized instructional pacing, might provide further insights into how immersive systems impact task efficiency in asynchronous collaboration settings [48, 77].

System Usability

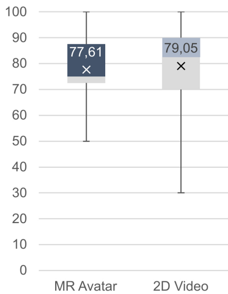


Figure 6.13: System Usability Scale (SUS) scores comparing MR Avatar and 2D Video conditions

The analysis of SUS scores revealed no significant difference between the *MR Avatar* condition ($M = 77.614$, $SD = 12.308$) and the *2D Video* condition ($M = 79.048$, $SD = 16.63$), $t(41) = 0.81$, $p = 0.422$, $r = 0.123$.

The small effect size ($r = 0.123$) indicates that any difference in perceived system usability was minimal. Figure 6.13 shows the achieved SUS scores in both conditions. Both collaboration methods achieved high SUS scores, suggesting that participants found both approaches highly usable. The lack of a usability advantage for the *MR Avatar* condition implies that while immersive elements did not hinder usability, they also did not provide a noticeable improvement in system usability compared to conventional *2D Video* instructions.

Discussion: The high usability ratings across both conditions indicate that neither method introduced usability barriers, reinforcing the idea that asynchronous collaboration methods can be intuitive and user-friendly, regardless of whether they are immersive or traditional. This is particularly important, as one concern with MR-based approaches revolves around the possibility that added complexity could negatively affect usability. Some related studies showed usability issues with viewpoint and playback control [36, 154], other studies suggest improved usability in AR instructions [34, 35]. The results of this study suggest no negative usability effect with the *MR Avatar*. However, it also did not prove to be substantially superior to *2D Video* instructions.

One possible explanation for this finding is that the tasks used in the study were well-structured and did not require extensive interaction with the system, beyond following instructions. In scenarios where more complex decision-making or adaptive user input is required, usability differences may become more apparent [73]. Future studies could investigate whether more dynamic and interactive MR-based systems, such as real-time avatar feedback or adaptive task guidance, impact usability perceptions differently [116].

Overall, the outcome shows that *MR Avatar* and *2D Video* are both effective for asynchronous collaboration in terms of usability, with no clear advantage for either method. This results support prior findings that immersive approaches, although they can enhance engagement and social presence, do not inherently improve usability unless they incorporate targeted interaction enhancements [73].

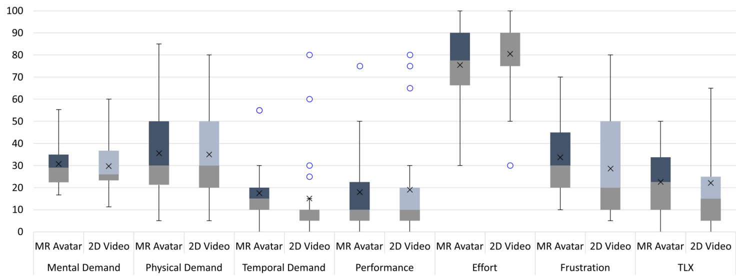


Figure 6.14: NASA-TLX results across all six workload subscales: Mental Demand, Physical Demand, Temporal Demand, Effort, Performance and Frustration.

Cognitive Workload

The NASA-TLX results, as depicted in Figure 6.14 and listed in Table 6.4, showed no significant differences in perceived workload between the *MR Avatar* condition ($M = 30.697$, $SD = 10.573$) and the *2D Video* condition ($M = 29.714$, $SD = 12.171$), $t(41) = 0.87$, $p = 0.389$, $r = 0.044$.

Across all six workload subscales – Mental Demand, Physical Demand, Temporal Demand, Effort, Performance and Frustration – participants reported comparable levels of cognitive and physical workload, suggesting that both asynchronous collaboration methods imposed a similar burden on users.

Table 6.4: Summary of NASA-TLX workload ratings comparing MR Avatar and 2D Video conditions

Measure	**MR Avatar**		**2D Video**		**p**	**r**
	M	SD	M	SD		
Weighted TLX Score	30.697	10.573	29.714	12.171	0.389	0.044
Mental Demand	35.455	18.703	35.000	22.638	0.713	0.056
Physical Demand	17.500	14.618	15.000	19.875	0.127	0.233
Temporal Demand	17.955	20.567	19.048	24.475	0.971	0.006
Performance	75.455	18.186	80.476	16.651	0.306	0.156
Effort	33.636	16.986	28.571	22.257	0.221	0.187
Frustration	22.500	14.454	22.143	20.711	0.591	0.082

Discussion: The absence of significant differences in NASA-TLX scores suggests that neither the *MR Avatar* nor the *2D Video* method increased workload or led to noticeably greater cognitive strain. While immersive systems can sometimes introduce additional cognitive load due to increased sensory input and interaction complexity [139], this was not the case here. This aligns with related studies suggesting that adding assistive MR cues may help reduce cognitive load [34, 35]. The structured nature of both guidance methods, combined with clear task instructions, likely contributed to keeping workload at a manageable level.

Although minor numerical differences were observed, such as slightly higher Mental Demand in the *MR Avatar* condition and marginally higher Performance scores in the *2D Video* condition, effect sizes were consistently small, indicating that these differences were not significant in practice. If designed well, additional MR content seem rather to reduce task load, as related work suggests [34, 35]. The Effort and Frustration ratings also remained similar across conditions, suggesting that participants did not perceive either method as particularly difficult or discouraging.

User Experience

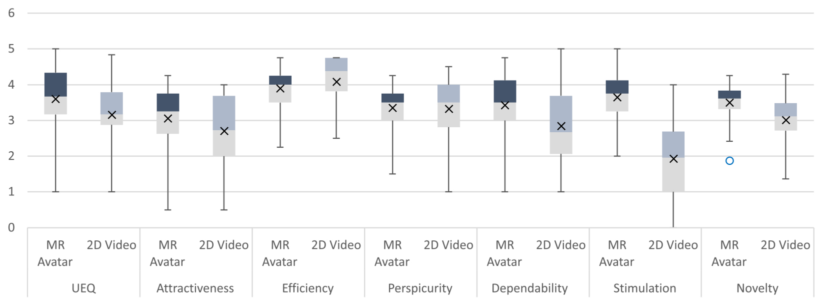


Figure 6.15: User Experience ratings for MR Avatar and 2D Video across dimensions

User experience was assessed using the User Experience Questionnaire (UEQ), measurering perceived usability and engagement across multiple dimensions. Table 6.5 presents the results comparing the *MR Avatar* and *2D Video* conditions. A significant difference was observed in the overall UEQ score, with the *MR Avatar* condition ($M = 1.496$, $SD = 0.585$) receiving higher ratings than the *2D Video* condition ($M = 1.007$, $SD = 1.007$), $t(43) = 2.297, p = 0.027$, $r = 0.331$. Figure 6.15 depicts the UEQ results including the six dimensions Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation and Novelty. Additional insights into user experience ratings are illustrated in Figure 6.16.

Table 6.5: Statistical comparison of UEQ dimension ratings between MR Avatar and 2D Video conditions

Measure	**MR Avatar**		**2D Video**		**p**	**r**
	M	SD	M	SD		
UEQ*	1.496	0.585	1.007	1.007	0.027	0.331
Attractiveness	1.598	0.998	1.159	1.058	0.116	0.234
Efficiency	1.057	0.829	0.702	0.984	0.238	0.176
Perspicuity	1.898	0.651	2.083	0.733	0.199	0.192
Dependability	1.352	0.597	1.321	0.958	0.891	0.020
Stimulation*	1.420	0.934	0.845	1.214	0.041†	0.263
Novelty*	1.648	0.789	-0.071	1.163	$<.001$	0.664

*Statistically significant results.
† Statistically significant difference could only be found in the one-tailed test.

Discussion: The results suggest that the *MR Avatar* condition led to a more engaging and positively perceived user experience compared to the *2D Video* condition. The significant difference in the overall UEQ score ($p = 0.027$) indicates that participants rated the avatar-guided experience as more usable and satisfying. Previous research indicates that shared virtual content [66], the presence of an avatar [154, 171], and the interactive nature of collaboration [63, 73, 159] can enhance user engagement. However, in our results, most usability-related dimensions – Attractiveness, Efficiency, Perspicuity and Dependability

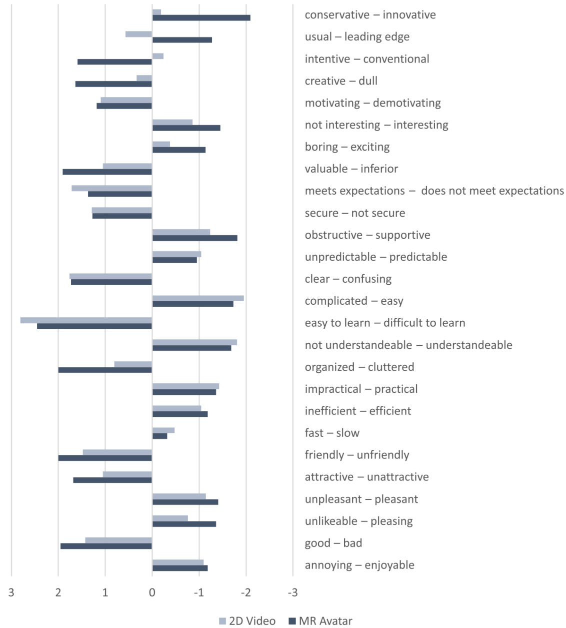


Figure 6.16: Detailed UEQ subscale comparisons between MR Avatar and 2D Video conditions

– showed no significant differences between conditions, suggesting that both methods provided a similar level of perceived usability.

Among the subscales, Stimulation was rated significantly higher in the *MR Avatar* condition ($t(43) = 1.787$, $p = 0.041$, $r = 0.263$), implying that participants found the avatar-based interaction more engaging compared to the video-based approach. A more pronounced difference was observed in Novelty ($t(43) = 5.826$, $p < 0.001$, $r = 0.664$), with the *MR Avatar* condition being perceived as a significantly more innovative experience. Related work reported similar findings. This suggests that the immersive and interactive nature of the avatar replay method provided a unique and engaging form of guidance compared to the more conventional *2D Video* format. However, given that most participants were unfamiliar with MR, the observed effect on Novelty may

– at least partially – reflect a *"novelty-effect"*[2] rather than solely the design characteristics of the avatar condition [30, 112].

The strong effect size for Novelty indicates that the avatar-based approach was not only effective in capturing participants' attention but was also perceived as a more modern and distinctive way of delivering instructional content. This is particularly relevant for XR-based collaboration systems, since a strong sense of engagement is crucial for maintaining user motivation and ensuring high-quality interaction.

Despite these advantages, the absence of significant differences in core usability dimensions suggests that while avatars enhance engagement and novelty, they do not necessarily improve the perceived effectiveness of task execution. This aligns with previous research indicating that immersive systems can enhance user experience without necessarily increasing task efficiency (➜ Efficiency Results 6.3.4). Future work could explore how to integrate adaptive avatar behaviors to further enhance usability while maintaining the advantages of engagement and novelty.

6.3.2 Collaboration Results

Collaboration was assessed across three subcomponents: Communication, Cooperation and Coordination. Table 6.6 presents the results comparing the *MR Avatar* and *2D Video* conditions, with the corresponding subscale scores visualized in Figure 6.17. No statistically significant differences were observed in any of the subscales ($p > 0.05$ for all comparisons), suggesting that both asynchronous collaboration methods were perceived as similarly effective.

[2] The novelty-effect describes the boost in attention, motivation, or perceived effectiveness that often occurs during initial exposure to a new technology, independent of actual performance.

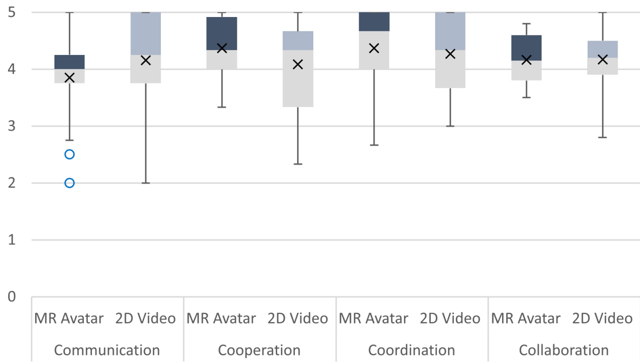


Figure 6.17: Collaboration ratings across three subcomponents: Communication, Cooperation, and Coordination

Table 6.6: Collaboration results summary comparing MR Avatar and 2D Video conditions.

Measure	MR Avatar		2D Video		p	r
	M	SD	M	SD		
Collaboration	4.159	0.120	4.167	0.135	0.779	0.043
Communication	3.852	0.722	4.155	0.893	0.120	0.237
Cooperation	4.364	0.534	4.079	0.809	0.348	0.143
Coordination	4.364	0.650	4.270	0.727	0.774	0.044

Although no significant differences were found, the mean values suggest some minor trends. The *MR Avatar* condition showed slightly higher ratings in Communication and Cooperation, indicating a potential tendency for participants to feel more engaged or supported by the avatar than by the video instructions. However, the effect sizes were small, suggesting that these differences were not substantial enough to be meaningful. Coordination, on the other hand, was rated nearly identically in both conditions, implying that the structured, step-by-step approach provided a consistent level of task organization across modalities.

Discussion: No statistically reliable differences were observed between the *MR Avatar* and *2D Video* conditions on the collaboration measures (communication, cooperation, coordination). Accordingly, the present data do not support the conclusion that either modality improves or worsens perceived collaboration relative to the other. Descriptively, scores were high in both conditions (scale $[0, 5]$), which is consistent with the interpretation that clear, stepwise guidance enabled participants to communicate, cooperate and coordinate effectively regardless of format. Because no equivalence or non-inferiority analysis with a pre-specified margin was conducted, equivalence cannot be claimed; rather, only the absence of detected differences within the current sample and measures can be reported. Prior work has argued that interactive or 3D shared content can support collaboration [63, 73], although these studies did not directly operationalize the three C's, limiting direct comparability. Implementing stepwise playback with pauses aligns with recommendations in related work [16, 154]; however, the present results do not show a statistically reliable advantage of the *MR Avatar* condition on the collaboration scales. Future work should employ instruments that target the three C's more granularly, plan a priori power to detect small effects, and consider equivalence or Bayesian analyses to quantify evidence for similarity; qualitative feedback may also reveal nuanced differences not captured by the current questionnaire.

6.3.3 Spatial and Social Presence Results

Spatial and social presence were evaluated using two established instruments: the Igroup Presence Questionnaire (IPQ) and the Networked Minds Social Presence Inventory (NMSPI). These tools measured the immersive and interpersonal aspects of the participants' experiences across conditions.

Spacial Presence

Since the IPQ is specifically intended for evaluating immersive experiences, it was initially not planned to be used for the *2D Video* condition, as traditional video lacks the depth and interactivity of XR-based environments. However, it was later decided to include the IPQ for a small subset of participants in the *2D Video* group (five participants) to explore whether any differences in presence perception could still be observed. The results are summarized in Figure 6.18.

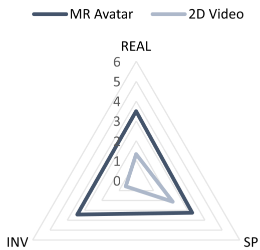


Figure 6.18: Comparison of IPQ subscales (Involvement, Realism, Spatial Presence) between MR Avatar and 2D Video conditions

Despite the smaller sample size in the *2D Video* group, the statistical analysis revealed significant differences with large effect sizes, reinforcing the robustness of the findings. Table 6.7 presents the results, showing that spatial presence scores were significantly higher in the *MR Avatar* condition ($M = 3.258$, $SD = 1.191$) compared to the *2D Video* condition ($M = 2.125$, $SD = 2.097$). An independent samples t-test confirmed this difference as statistically significant, $t(24) = 3.401, p = 0.002$, with a large effect size ($r = 0.570$). This result strongly suggests that *MR Avatar* interactions significantly enhance the participants' experience of spatial presence compared to traditional video-based methods.

Table 6.7: IPQ results summary comparing MR Avatar and 2D Video conditions.

Measure	MR Avatar		Smartphone Video		p	r
	M	SD	M	SD		
REAL*	3.489	0.657	1.333	1.155	0.024	0.470
SP**	3.258	1.191	2.125	2.097	0.002	0.570^{+}
INV***	3.409	0.599	0.600	0.894	$<$.001	0.658^{+}

* Statistically significant results.
** Highly significant result ($p < .01$).
*** Highly significant result ($p < .001$).
$^{+}$ Large effect size ($r > .5$).

Similarly, involvement scores were significantly higher in the *MR Avatar* condition ($M = 3.409$, $SD = 0.599$) than in the *2D Video* condition ($M = 0.6$, $SD = 0.894$), with a large effect size ($U = 57$, $p < .001$, $r = 0.658$). Additionally, realism was rated significantly higher in the *MR Avatar* condition ($M = 3.489$, $SD = 0.657$) compared to *2D Video* ($M = 1.333$, $SD = 1.155$), showing another statistically significant difference ($U = 110$, $p = 0.024$, $r = 0.470$). These findings suggest that the *MR Avatar* method fosters a stronger sense of immersion, spatial presence and realism than the *2D Video* condition. Figure 6.19 illustrates the distribution of individual IPQ scores, further emphasizing the variation in perceived presence between conditions.

Discussion: The large effect sizes across all subscales indicate that immersive MR technology has a strong impact on presence perception. The involvement and realism scores suggest that participants in the *MR Avatar* condition felt more engaged and found the virtual environment more believable than those using the *2D Video* method. Prior studies indicate that greater engagement, realism and richer interaction within a shared virtual space are associated with increased presence [73, 120, 139]. However, while the *MR Avatar* approach significantly enhanced presence, its mean values remained below the theoretical maximum of the scale[3], indicating that there is still room for improvement by increasing the immersive qualities of avatar-based asynchronous collaboration.

The variation in presence perception, as indicated by the relatively large standard deviations, suggests that individual differences played a role in how participants experienced immersion. Factors such as prior experience with XR, cognitive styles and personal preferences likely contributed to the extent to which participants felt present in the environment. Related work showed similar dependence of co-presence results between 3D and 2D conditions [156]. Notably, the *2D Video* group exhibited the highest variability ($SD = 2.097$), suggesting that while some participants found the video-based instructions engaging, others felt significantly less immersed.

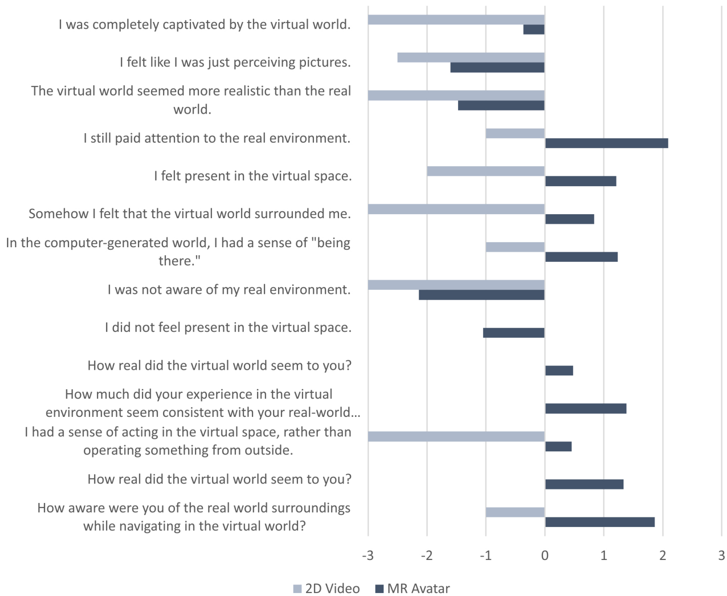


Figure 6.19: Distribution of individual IPQ scores across subscales and conditions

3 within the value range $[-3, +3]$

It is important to consider that the IPQ is designed for immersive experiences, such as XR. Although it is not ideal for evaluating non-immersive conditions, a small set ($N = 5$) of participants were asked to complete IPQ for the *2D Video* condition. The statistical testing revealed significant differences with large effect sizes; however, the sample size is too small for robust conclusions. Since IPQ is specifically tailored to immersive environments, the observed differences align well with the expectation that the *MR Avatar* condition would naturally induce a greater sense of spatial presence. Related work suggests that presence is positively correlated with user performance [139], yet our data do not show improvements in perceived performance (➜ Cognitive Workload 6.3.1) or efficiency (➜ Efficiency Results 6.3.4). Instead, the immersive condition's higher presence aligns with accuracy gains on spatially demanding steps (➜ Accuracy Results 6.3.5).

Social Presence

The Networked Minds Social Presence Inventory (NMSPI) was used to measure the degree of perceived social presence (co-presence). The corresponding distribution of NMSPI ratings is visualized in Figure 6.20.

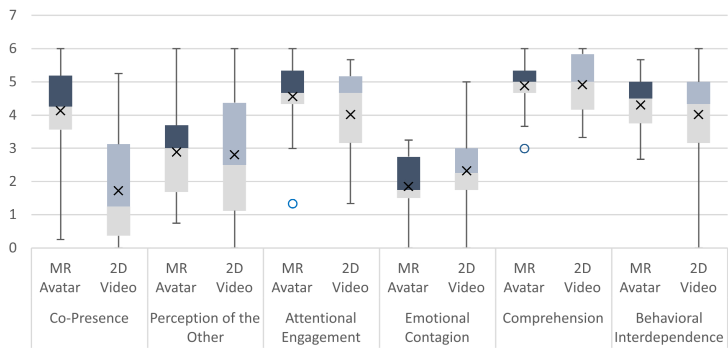


Figure 6.20: NMSPI results by subscale scores showing social presence perceptions in both conditions

Table 6.8 summarizes the results of the NMSPI groups, presenting a significant difference in co-presence between the two conditions. Participants in the *MR Avatar* condition reported a higher sense of shared space and connection with the instructor ($M = 1.136$, $SD = 1.467$) compared to the *2D Video* condition ($M = -1.031, SD = 1.717$). The Mann-Whitney U test confirmed that this difference was statistically significant ($U = 54, p < .001$), with a large effect size ($r = 0.624$). This indicates that the avatar-based replay method was considerably more effective at fostering a sense of belonging compared to traditional video instructions.

Table 6.8: Statistical comparison of NMSPI subscale scores between MR Avatar and 2D Video conditions

Measure	**MR Avatar**		**Smartphone Video**		**p**	**r**
	M	SD	M	SD		
Co-Presence* **	1.136	1.467	-1.031	1.717	$<$.001***	0.624$^+$
Perception of the Other	-0.114	1.420	-0.104	1.104	0.893	0.022
Attentional Engagement	1.561	1.091	1.047	1.286	0.252	0.179
Emotional Contagion	-1.148	0.981	-0.615	1.458	0.227	0.193
Comprehension	1.879	0.787	1.857	0.938	0.895	0.021
Behavioral Interdependence	1.303	0.942	1.028	1.449	0.822	0.035

*** Statistically highly significant result ($p < .001$).
$^+$ Large effect size ($r > .5$).

The mean value in the *MR Avatar* condition suggests a moderate level[4] of co-presence, though not at the highest possible level. This implies that while avatar-based asynchronous collaboration significantly enhances social presence, it does not fully replicate the feeling of real-time interaction. The large standard deviation ($SD = 1.467$) also highlights substantial variability among participants, indicating that some individuals experienced a strong sense of

[4] within the value range $[-3, +3]$

co-presence while others did not. In contrast, the *2D Video* condition not only resulted in lower mean co-presence scores but also exhibited an even larger standard deviation ($SD = 1.717$). This suggests that while some participants may have found the video-based instructions engaging, the overall perception of co-presence was lower and highly inconsistent.

Further examining the subcomponents of co-presence, the results for perception of the other, attentional engagement, emotional contagion, comprehension and behavioral interdependence showed no statistically significant differences between conditions. This suggests that while the *MR Avatar* method enhances the overall sense of co-presence, it does not necessarily influence all aspects of social presence equally. Notably, the comprehension scores were similar across both conditions, indicating that the effectiveness of information transfer was not significantly impacted by the method of delivery.

Discussion: The findings suggest that the increased co-presence in the *MR Avatar* condition likely stems from the spatial alignment and embodied interaction cues provided by the avatar, which are absent in *2D Video* instructions [139, 154]. Similar patterns have been observed in prior work, where 3D interfaces were found to enhance the sense of co-presence and interaction compared to 2D alternatives, although the latter were sometimes perceived as more intuitive to use [73]. However, the variability in responses indicates that personal factors such as prior XR experience, cognitive preferences or familiarity with avatar-based interactions may influence how participants perceive social presence. Earlier studies have presented comparable outcomes, indicating divergent user preferences based on personal learning styles and comfort [156]. Given that neither condition reached the extreme ends of the scale, it is obvious that while avatar-based collaboration is more effective than video, it is not a perfect substitute for real-time interaction like in synchronous collaboration.

The results emphasize the importance of designing asynchronous collaboration tools that optimize co-presence while acknowledging individual differences. Enhancing avatar realism, integrating adaptive behavioral cues and offering personalized interaction settings may further improve the experience [154]. Additionally, while MR Avatars are more effective in fostering presence, *2D Video* instructions remain a viable alternative for scenarios where high co-presence is not a priority.

Furthermore, the avatar lacked facial expressions and lighting was not implemented perfectly, which resulted in uncanny lighting/shadows of the avatar. Also, according to observations and participants' feedback, the co-presence feeling mitigated when the avatar was not reacting, intruded the personal space of the participants, collided with physical objects like walls or walked through physical objects like a ghost [36, 154, 162, 172].

6.3.4 Efficiency Results

Efficiency was operationalized by measuring the time participants needed to complete various task across conditions. This allows insight into how different instructional modalities affect the pace of task execution and problem-solving.

Recorded Time Data

The analysis of task completion times in the *MR Avatar* and *2D Video* conditions revealed significant differences in efficiency across various phases of the experiment, as depicted in Figure 6.21. Overall, participants in the *2D Video* condition completed tasks faster than those in the *MR Avatar* condition, particularly in the solution application phase (Part III).

Table 6.9 provides a detailed breakdown of time differences across the task phases. During novice tasks (Part I+II), which involved both task execution and error reporting, participants using the *MR Avatar* took slightly longer ($M = 26.00$ min, $SD = 4.44$) compared to those in the *2D Video* condition ($M = 23.57$ min, $SD = 10.95$), with a statistically significant difference, $p = .023$, $r = .346$. A closer examination revealed that while task execution times were

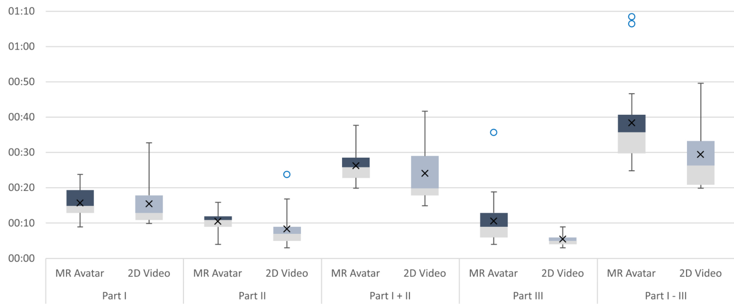


Figure 6.21: Boxplot visualization of completion times across task phases for MR Avatar and 2D Video conditions

similar between conditions in Part I, in Part II error reporting was significantly slower in the *MR Avatar* condition ($M = 10.18$ min, $SD = 2.16$) compared to *2D Video* ($M = 8.00$ min, $SD = 5.18$), $p = .013$, $r = .383$. This suggests that detecting and reporting errors required more time when interacting with an avatar, possibly due to increased cognitive load or differences in how issues were perceived in the MR interface.

Table 6.9: Mean task completion times (in minutes) and standard deviations across phases for MR Avatar and 2D Video conditions

Measure	**MR Avatar**		**2D Video**		**p**	**r**
	M	SD	M	SD		
Part I	15.82	4.10	15.14	6.17	0.429	0.121
Part II*	10.18	2.16	8.00	5.18	0.013	0.383
Part I + II*	26.00	4.44	23.57	10.95	0.023	0.346
Part III* **	10.00	6.86	5.43	1.50	$<.001$	0.567
$\sum$ Part I - III	38.71	11.30	28.14	11.98	0.448	0.004

* Statistically significant result ($p < .05$).
*** Highly significant result ($p < .001$).

In Part III, where participants applied solutions from prior steps, the *2D Video* condition showed a substantial efficiency advantage. Participants using the *MR Avatar* took nearly twice as long ($M = 10.00$ min, $SD = 6.86$) as those in the *2D Video* condition ($M = 5.43$ min, $SD = 1.50$), with a highly significant effect, $p < .001$, $r = .567$. This large effect size suggests that video-based instructions provided clearer references, enabling faster retrieval and application of knowledge. In contrast, avatar-based interactions might have introduced additional complexity in recalling and executing learned solutions.

When considering the total task duration (Part I – III), participants in the *MR Avatar* condition showed a numerically higher completion time ($M = 38.71$ min, $SD = 11.30$) compared to those in the *2D Video* condition ($M = 28.14$ min, $SD = 11.98$). However, this difference was not statistically significant, $p = .448$, indicating high inter-individual variability.

These findings suggest that the *2D Video* condition facilitated more efficient task completion, particularly in solution application, possibly due to its static and structured nature, which can aid in information retrieval. In contrast, while MR Avatars may provide a more engaging and immersive experience, they appear to introduce a higher cognitive load, which could slow down performance in asynchronous collaboration settings. Future research should explore how to optimize avatar-based interactions for efficiency without compromising engagement.

Discussion: The findings underscore a clear trade-off between immersive engagement and task efficiency. While the *MR Avatar* condition offers a more interactive and embodied experience, this benefit may come at the cost of increased time investment during task execution – especially in complex phases including response review and solution application. Longer instruction time, but a statistically similar assembly time, was observed in related work with AR instructions [43].

One possible explanation lies not in the MR modality itself, but in the discrepancy between the technological maturity and user familiarity of the two systems. Smartphone-based video interactions – such as timeline interfaces – are highly optimized through decades of iterative design and are used daily by all participants [34, 154]. In contrast, the *MR Avatar* interface represents a relatively novel and still maturing form of interaction. Its unfamiliarity, combined with the early-stage nature of the prototype, may have introduced friction in navigation, comprehension, and control, leading to slower performance despite the immersive potential.

Interestingly, the greatest difference was observed during error reporting and solution application (Parts II and III), where clarity and quick access to prior information are critical. This suggests that while MR can enhance presence, it may be less suited for tasks requiring speed – unless specifically optimized.

Another contributing factor is the fundamental trade-off between naturalistic and abstracted interaction paradigms. Immersive interfaces allow embodied, spatial interactions that mimic real-world behavior – such as pointing or gesturing – whereas smartphone interfaces use highly abstracted but efficient interactions like tapping, scrolling, or swiping [16]. For example, selecting or highlighting an object on a touchscreen is typically faster than gesturing toward it in a roomscale space. However, while slower, these embodied interactions more closely resemble natural human behavior and do not require the same degree of learned interface conventions [16, 38]. This trade-off between intuitiveness and efficiency must be carefully considered when designing collaborative XR systems, especially in asynchronous contexts where not just timing but also precision are crucial (❯ Accuracy Results 6.3.5).

These results highlight the need for better interface design in MR applications. Rather than fundamental limitations of MR, the inefficiencies likely reflect a gap in usability design and user adaptation. Incorporating more intuitive controls, improving pacing, and aligning design with users' mental models can help reduce the temporal overhead and support spatial memory [35, 154]. Future research should also explore whether combining MR with traditional instructional modalities or input interfaces can offer the best of both worlds – enhanced engagement without sacrificing efficiency.

6.3.5 Accuracy Results

To assess the effectiveness of each instructional condition, task accuracy was measured as the percentage of correctly completed steps. Completion rates were compared across experimental phases for both the *MR Avatar* and *2D Video* conditions, as summarized in Table 6.10.

Recorded Completion Rate

The accuracy scores, measured as the percentage of correctly completed tasks, were analyzed to compare the effectiveness of asynchronous guidance between the *MR Avatar* and *2D Video* conditions (see Table 6.10 and Figure 6.22).

Table 6.10: Task accuracy comparison across phases between MR Avatar and 2D Video conditions

Measure	**MR Avatar**		**Smartphone Video**		**p**	**r**
	M (%)	SD (%)	M (%)	SD (%)		
Part I**	92.41	11.35	75.43	19.19	0.001**	0.513
Part II	90.34	16.55	94.25	12.03	0.530	0.098
Part I + II*	91.38	9.20	80.35	19.65	0.012*	0.381
Part III	94.76	6.80	90.50	8.87	0.110	0.250
$\sum$ **Part I - III***	91.07	9.48	82.70	15.03	0.024*	0.343

* Statistically significant results.
** Highly significant result ($p < .01$).

A significant difference was found in **Part I**, where participants using the *MR Avatar* condition achieved higher accuracy ($M = 92.98\%$, $SD = 10.38$) compared to the *2D Video* condition ($M = 77.65\%$, $SD = 19.88$), $p = .003$, with a moderate effect size ($r = .450$). In contrast, in **Part II**, the *2D Video* group achieved slightly higher accuracy ($M = 95.57\%$, $SD = 7.40$) than the *MR Avatar* group ($M = 92.16\%$, $SD = 16.41$), but this difference was not statistically significant ($p = .987$, $r = .0025$).

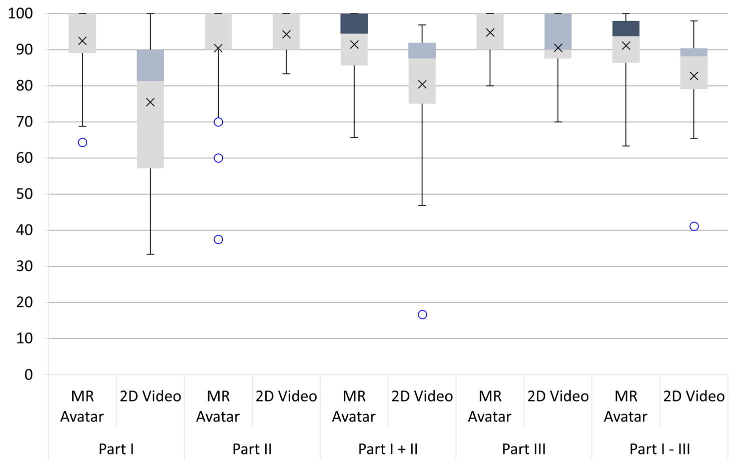


Figure 6.22: Boxplot visualization of accuracy across task phases for MR Avatar and 2D Video conditions

When combining **Part I and II**, a significant difference was found, with the *MR Avatar* condition showing higher accuracy ($M = 92.57\%$, $SD = 8.73$) compared to the *2D Video* condition ($M = 82.06\%$, $SD = 19.53$), $p = .019$, with a moderate effect size ($r = .357$). In **Part III**, no significant difference was observed ($p = .110$, $r = .250$). Finally, the overall accuracy score across all phases was higher in the *MR Avatar* condition ($M = 91.86\%$, $SD = 9.32$) compared to the *2D Video* condition ($M = 83.44\%$, $SD = 15.55$), with a statistically significant difference, $p = .014$, and a moderate effect size ($r = .373$).

These results suggest that while the *MR Avatar* condition led to higher accuracy overall, particularly in **Part I**, the *2D Video* condition did not show a significant advantage in **Part II**. The significant difference in the combined **Part I + II** score and overall accuracy suggests that the *MR Avatar* guidance method was more effective in ensuring accurate task execution across the experiment.

Discussion: The accuracy results indicate that MR-based asynchronous guidance can foster more precise task execution, particularly during guided phases. The significantly higher accuracy in Part I for the *MR Avatar* condition suggests that immersive, embodied instructions may support participants' comprehension and recognition during task execution. Related work suggests that additional interactivity and traceability of movements, e.g. hand gestures, in the local user's environment can improve auditory and spatial recall and understanding of key information [36, 38, 63]. Additionally, our findings align with related studies that suggest accuracy positively correlates with spatial presence (➜ Spatial Presence 6.3.3). Although not measured in this study, related work further suggests a better long-term recall after training with MR systems [43].

A possible explanation for this advantage lies in the visual clarity and spatial embedding of the instructions provided by the avatar, as previous studies suggest [153]. Additionally, participants in the *2D Video* condition often experienced a split-attention problem: their focus constantly shifted between the handheld screen and the physical workspace. Most participants attempted to follow the video while simultaneously performing tasks in the real environment. As a result, when moving through the workspace, they frequently missed critical visual cues in the video because their attention was directed toward their surroundings. In contrast, the *MR Avatar* instructions were embedded directly into the environment, eliminating the need to shift attention between two separate contexts. This could explain the higher error rate and lower accuracy observed in the video group. Similar findings related to split attention have been reported in previous work, where camera framing limitations made it difficult to capture both the workspace and the instructional actions simultaneously, often resulting in omissions or misunderstandings [154].

Interestingly, participants rarely paused or rewind the video, likely to avoid missing important on-screen visuals, and instead continued navigating the physical space while listening to the audio instructions. Still, some participants replayed the video several times to clarify specific steps. A similar observation was noted in related work, where the authors pointed out that students rarely re-wound recordings to check for errors; our findings provide empirical support for this claim [154]. Future studies could focus on whether more concise step-by-step videos – comprising more steps yet shorter duration – might contribute to a reduction in errors, a notion stated in previous research [16, 154].

It is important to note that some technical limitations may have also affected MR performance, similar to previous work [156]. Participants occasionally encountered playback problems, including audio delays, avatar desynchronization, or incomplete body tracking. These were known limitations of the early prototype. In rare cases (two participants), tracking instability led to positional misalignment, although these were quickly corrected during the session.

Despite these issues, the overall higher accuracy in the *MR Avatar* condition suggests that immersive guidance can support better spatial and procedural understanding. The avatar's embodied actions and spatial anchoring likely contributed to more robust cognitive encoding, as suggested in related work [35, 43].

These findings reinforce the potential of MR-based instruction to enhance accuracy in XR collaboration scenarios. Future developments should focus on improving technical reliability, introducing more flexible playback controls, and offering feedback mechanisms to help participants monitor and verify their progress during complex or error-prone tasks.

6.3.6 Qualitative Results

The questionnaire included additional comment sections where participants could provide further feedback in text form. These feedback results are categorized into three groups: usability, avatar interaction and learning aspects that are further presented and discussed in this section.

Usability: Participants considered the systems to be intuitive and user-friendly in general, but some suggested specific improvements – especially among HMD users. Smartphone users valued the familiar interface, describing it as straightforward and aligned with everyday apps like messengers. Related work suggest hands-free conditions to be more preferable and effective [16, 38, 139]. In our study, HMD users appreciated the clarity of the menu and hands-free interaction, but noted issues such as difficulty reaching avatar-related buttons, confusing button feedback and having to walk around or through the avatar to proceed. Hands-free interaction seem to be commonly favored as discussed in related work [16, 38, 73, 156]. One participant requested *"bigger record buttons"* to prevent misclicks, while another highlighted the need for improved positioning of controls for shorter users. That support reports in related work, where the positioning of virtual elements occluded the guidance if not placed ideally [35]. Smartphone participants focused more on interface ergonomics, such as accidentally selecting the wrong icon, or struggling to manage and identify multiple videos without labels. Overall, users across both platforms called for improved layout, clearer visual cues and reduced cognitive demand when multitasking between content and task execution.

Avatar Interaction: The avatar received largely positive reactions, particularly in terms of its instructional value and presence. Users appreciated the spatial and gestural communication, which many said helped them understand procedures more easily than static instructions. That is supported by similar results in related work [154]. A few described their instinctive behavior of walking around the avatar as if it were a real person, showing a surprising level of social presence, similar to prior work [120]. This feature enabled participants to observe the playback from various perspectives, a capability that was valued by users in past studies [156, 159]. Some participants expressed discomfort when the avatar invaded their personal space or when they had to walk through it, finding that *"strange"* or *"unnatural"*. Similar observations were noted in related studies focusing on teleportation, where it was suggested to limit teleportation zones to prevent overlap with the avatar's playback areas

[36]. Some participants reported minor visual glitches, like mismatches in animation or expression. Interestingly, the avatar's human-like presence was often reported as beneficial, helping users stay engaged and follow instructions, despite occasional proximity uncanny moments. One user remarked that it *"felt like interacting with a person"*, and another mentioned they avoided walking through it, despite being aware it was digital, suggesting the avatar successfully triggered real-world social behavior. Nevertheless, a less realistic representation of the avatar may also be sufficient for a similar degree of co-presence, as indicated in related studies [154].

Learning Aspects: Comments showed a clear contrast between the prototypes in how participants experienced the learning process, even though no participant tested both. HMD users described learning as guided, immersive and embodied, frequently mentioning the advantage of seeing gestures in real time and following an expert's actions in the physical space. This real-time, avatar-based guidance was seen as intuitive and easy to remember, particularly in complex or spatial tasks. Similiar impressions about the intuitiveness, retention and usability were reported in related work [73, 154]. In contrast, smartphone users valued the ability to pause, rewatch and process instructions at their own pace. Our results support similar observations in prior studies [63, 154]. Several participants who tested the *2D Video* condition, noted that it would have been easier to *"reread or jump back"* with text and image based instructions compared to skipping through videos. While they appreciated the clarity of visual examples, some smartphone participants mentioned the effort of holding the phone while performing tasks – similar to prior studies [16, 38, 43] – and difficulties processing long sequences of video. Furthermore, it is suggested to auto-pause the playback when users perform tasks in their workspace, since it might help avoid split attention and prevent them from missing an instruction [154]. These patterns suggest that while the HMD approach offers a seamless task integration, the smartphone method supports greater flexibility and reflection.

6.4 Major Findings and Implications

This section presents the synthesized findings from the evaluation study that compared asynchronous collaboration in an immersive MR environment – ACIE approach – with more conventional asynchronous collaboration using flat media – video messaging approach. The objective is to illustrate the DSR rigor cycle and to derive design-related implications that extend beyond specific implementations. Additionally, the empirical data will be interpreted through a scientifically substantiated perspective.

Finding 1

> ℹ **Equivalent collaboration quality:** ACIE had no significant differences in communication, cooperation or coordination compared to the video-based messaging for asynchronous collaboration.

This finding reinforces the assumption that the implemented ACIE components for communication, cooperation and coordination are sufficiently supporting asynchronous collaboration and are not perceived as less effective as conventional methods, such as video messaging. When implemented well and thoroughly to support all 3C's, immersive media are well suited for collaboration on complex tasks and do not suffer from a lack of familiarity.

Design Implications: To support asynchronous collaboration with XR, immersive approaches for communication, cooperation and coordination have to be supported.

Finding 2

> ℹ **Usability trade-offs in immersive collaboration:** Although ACIE resulted in significantly higher accuracy, especially for error-sensitive tasks, it was less efficient.

Task durations under ACIE were longer, likely due to added complexity by unfamiliar menu paradigms in MR, while participants could navigate through the video messages with ease. However, ACIE enabled more accurate task execution. The spatial alignment of instructions and embodied cues from the proxy may have supported deeper understanding and better error prevention. Additionally, the seamless integration of proxy-avatar instructions within the workspace may help maintain focus, while switching between the display and the work space splits the attention [154]. This presents a trade-off: immersive media provide higher accuracy, but may require optimization to match the speed of traditional media.

Design Implications: Complex task procedures, especially those relying on body motion, spacial navigation or spatial referencing, benefit from immersion and embodiment. However, the interaction needs optimization to be able to compete with traditional familiar media and interaction paradigms. Further investigation is needed to determine whether more training/familiarity also reduces this efficiency gap. For other tasks or job types, immersive media adds an extra layer of complexity and unfamiliarity, which in turn leads to inefficiency.

Finding 3

> **Enhanced user experience without added cognitive load:** ACIE significantly increased perceived user experience, particularly in stimulation and novelty, while maintaining statistically equivalent usability and workload levels.

Despite their lack of familiarity with immersive technologies, participants rated ACIE's usability (ASQ, SUS) and perceived workload (NASA-TLX) as equivalent across conditions. This indicates that immersive asynchronous systems can enrich user experience without introducing additional cognitive burden. Higher stimulation and novelty in the ACIE condition suggest elevated engagement potential, which may translate to stronger motivation and learning retention over time.

Design Implications: Since usability was comparable, future research should focus on optimizing task efficiency and user engagement, rather than usability per se. Refining interaction modalities, such as gesture-based interactions or adaptive guidance, could enhance efficiency without compromising usability. Since neither method led to excessive cognitive load, system designers can introduce additional interactive elements (e.g. user-driven exploration, embedded quizzes) to increase engagement without overwhelming users.

Summarizing the evaluation, it can be concluded that immersive asynchronous collaboration (ACIE) provides benefits in terms of experience and learning – including increased spatial presence, co-presence, accuracy and user interaction – without increasing cognitive load or reducing perceived usability. However, these benefits come at the cost of reduced efficiency, largely due to unfamiliar interaction paradigms and increased complexity. The findings suggest that

ACIE is a promising method for asynchronous collaboration on tasks requiring physicality or spatial thinking. However, further design optimizations are needed to achieve the efficiency of traditional methods.

Final Takeaway

In summary, the evaluation shows that immersive asynchronous collaboration – like in ACIE – offers meaningful experiential and instructional benefits, including enhanced spatial presence, co-presence, task accuracy, and user engagement, without increasing cognitive workload or compromising usability compared to traditional flat-media collaboration. Table 6.11 provides a concise overview of the most important findings at a glance.

Table 6.11: Summary of core findings across presence, collaboration, performance, and experience dimensions

Finding	Summary
F1: Collaboration Quality	No significant difference in perceived communication, cooperation, or coordination.
F2: Accuracy vs. Efficiency	ACIE improved task accuracy but increased task duration.
F3: User Experience	ACIE enhanced user experience (novelty, stimulation) without adding cognitive load.

However, ACIE's immersive advantages come at the cost of reduced task efficiency, likely due to unfamiliar interaction paradigms and additional system complexity. Table 6.12 summarizes the empirical validation of the hypotheses, which were defined in ❯ Section 6.2. These findings suggest that immersive asynchronous tools are particularly valuable in tasks involving embodiment, spatial reasoning or collaborative instruction, but require further optimization to match the speed and operational familiarity of conventional methods.

Table 6.12: Empirical validation of XR-related hypotheses in the Rigor Cycle, based on observed efficiency, accuracy, presence, and satisfaction outcomes

Hypothesis	Result	Evidence
H2.1	✖	XR does not improve efficiency in ACCT compared to non-immersive representations. In fact, the *MR Avatar* condition showed slower performance; the *2D Video* was significantly more efficient.
	✔	XR improves effectiveness in ACCT compared to non-immersive representations. The *MR Avatar* condition led to higher task accuracy and better performance in complex phases.
	✔	XR improves satisfaction in ACCT compared to non-immersive representations. Significantly higher stimulation and novelty were reported in the *MR Avatar* condition (UEQ).
H2.2	✔	XR improves spatial presence in ACCT compared to non-immersive representations. This was reflected in significantly higher IPQ scores (REAL, SP, INV).
	✔	XR improves social presence in ACCT compared to non-immersive representations. The *MR Avatar* condition significantly enhanced co-presence (NMSPI) with a large effect size.

✔ Supported
✖ Rejected

Answer to RQ2.1: In the ACIE condition, task effectiveness was significantly higher – particularly in complex and error-sensitive phases – reflected by increased task accuracy, while efficiency was significantly lower, with longer completion times than in the non-immersive baseline (*2D Video*). User satisfaction was also significantly higher (UEQ: stimulation, novelty). Together, these results support **H2.1** and indicate a trade-off between accuracy/satisfaction and speed in immersive asynchronous guidance.

Answer to RQ2.2: Compared with flat video messaging, ACIE yielded significantly higher spatial presence (IPQ: REAL, SP, INV) and co-/social presence (NMSPI), with a large effect size on co-presence. These findings support **H2.2** and demonstrate that immersive media enhance both the feeling of being in the environment and the sense of shared experience, even in asynchronous contexts.

6.5 Summary and Outlook

This chapter addressed **RQ2**, which focuses on the evaluation of asynchronous collaboration using immersive technologies in applied engineering scenarios (see ❯ **Chapter 1**, Figure 1.1). Building on the conceptual and technical foundations established in the previous chapters, an MR-based avatar guidance system was deployed in a real-world robotic cell setting to assess its practical relevance and effectiveness.

The comparative study contrasted the immersive MR-based avatar system with a traditional smartphone-based *2D Video* baseline. To thoroughly assess the two approaches, both quantitative and qualitative data were collected across multiple metrics, including task performance, usability, user experience, presence and collaboration quality.

- **Task Completion and Usability ❯ RQ2.1**: Both conditions enabled participants to successfully complete the procedural task. Usability scores (SUS, ASQ) showed no significant differences, confirming the practical applicability of both methods.
- **User Experience & Workload ❯ RQ2.1**: The participants reported similar workload levels (NASA-TLX), but significantly higher hedonic and experiential quality in the immersive condition. In particular, the UEQ subscales for stimulation and novelty showed large effects.
- **Perceived Presence ❯ RQ2.2**: The immersive condition led to significantly higher spatial presence (SP), involvement (INV) and co-presence, supporting the hypothesis that embodied avatar guidance enhances the subjective sense of presence and belonging.
- **Collaboration Quality ❯ RQ2.2**: Ratings of collaboration, cooperation, and coordination were high in both conditions. However, participants in the MR-based condition reported stronger spatial comprehension and engagement based on qualitative feedback.

- **Qualitative Insights**: The participants appreciated the intuitive nature of embodied guidance in XR, the contextualization of actions in space, and the enhanced sense of shared presence. Reported limitations included reduced control over playback and occasional occlusion or clumsiness in avatar gestures within the physical environment.

Answer to RQ2 (synthesis): Relative to the non-immersive baseline, the immersive system increased effectiveness (higher procedural accuracy) and satisfaction (higher stimulation and novelty), maintained usability and workload, and increased spatial presence and co-presence; efficiency was lower (longer completion times). Collaboration quality (communication, cooperation, coordination) was rated as equivalent across conditions.

A detailed summary of all quantitative results is provided in **Appendix A.8**, including statistical significance and effect sizes. These findings validate the practical relevance of immersive asynchronous collaboration and highlight key strengths and limitations that inform its refinement.

In Simpler Terms

This chapter showed that engineering students using the *MR Avatar* system could follow complex instructions more intuitively and with a stronger sense of spatial context than those using traditional video guides. They felt more present, more engaged, and perceived the collaboration as more natural – even though it was asynchronous. While the immersive system did not necessarily make tasks faster, it made the experience more vivid and meaningful.

How to Proceed

The evaluation confirms that immersive asynchronous collaboration is both usable and beneficial in real-world engineering environments. However, it also reveals practical limitations and technical considerations that guide future design improvements.

➋ **Chapter 7** integrates a critical discussion of these findings in light of the research questions, design principles and broader implications for XR-supported collaboration. It concludes the thesis by summarizing the key contributions and outlining directions for future work on scalable, flexible and embodied asynchronous collaboration systems.

7 Conclusion

This thesis advances the field of asynchronous XR-supported collaboration by introducing a novel conceptual technical framework and validating it through implementation and empirical evaluation. The Design Science Research (DSR) methodology was used to investigate how Extended Reality (XR) can support Asynchronous Collaboration on Complex Tasks (ACCT) in engineering. The research followed the three canonical DSR cycles-Relevance, Rigor and Design – each contributing to the development, evaluation and contextual embedding of a novel XR-based collaboration framework. The following provides an integrated summary of the dissertation chapters, their contributions and their alignment with the DSR cycles.

- **Chapter 1** frames the motivation and research questions, establishing the conceptual need for asynchronous collaboration in spatially complex engineering tasks. By articulating the limitations of existing CSCW systems and introducing ACCT as a focal construct, this chapter contributes to the **Relevance Cycle**, ensuring the investigation is grounded in a real-world and practically significant problem space.
- **Chapter 2** provides the theoretical underpinnings of the research. Drawing from CSCW, HCI and XR literature, it introduces the 3C model as an analytical structure for asynchronous collaboration in immersive environments. This conceptual foundation directly informs the design process, contributing to the Rigor Cycle by extending the academic knowledge base.

- **Chapter 3** presents a structured literature review highlighting the fragmented state of XR support for asynchronous collaboration. By identifying underexplored opportunities and design gaps, it supports both the Rigor and Relevance Cycles, providing a critical bridge between theoretical knowledge and design motivation.
- **Chapter 4** introduces the ACIE Framework, a modular system design that operationalizes the 3C model in XR environments. It specifies immersive message types, proxy representations and spatial coordination strategies. As the principal design artifact, this chapter is central to the Design Cycle, translating theoretical constructs into implementable system components.
- **Chapter 5** details the technical realization of the ACIE Framework across multiple XR platforms. It describes the development of asynchronous messaging protocols, spatial anchors and versioning tools. These iterative design efforts advance the Design Cycle by implementing and refining the artifact in alignment with real-world constraints.
- **Chapter 6** empirically evaluates the framework in a realistic engineering setting. A mixed-method study compares the XR system to a conventional video-based baseline, measuring spatial understanding, co-presence, usability and efficiency. The findings validate the artifact's impact, thereby completing the Relevance and Rigor Cycles through contextual evaluation and contribution to applied knowledge.
- **Chapter 7** integrates the key findings, reflects critically on limitations and outlines directions for future research. It synthesizes contributions across theoretical, methodological and practical dimensions. This chapter closes the DSR loop by feeding insights back into the knowledge base and practice, representing the culmination of all three DSR cycles.

Synthesis of Key Findings

The comparative evaluation against a smartphone-based video baseline indicated that the proposed framework supports asynchronous collaboration in immersive environments with several advantages:

Collaboration quality: Subjective ratings of communication, cooperation, and coordination in XR were equivalent to the non-immersive baseline despite lower familiarity.

Effectiveness & accuracy: While the XR system led to longer task durations, it resulted in higher procedural accuracy in spatially complex tasks, suggesting a trade-off between interaction complexity and task precision.

User experience: Despite increased complexity, the immersive system enhanced engagement and stimulation without raising perceived workload, indicating strong usability and cognitive sustainability.

Overall, the evaluation substantiates the viability of XR-based asynchronous collaboration, particularly for tasks involving spatial reasoning, embodied actions, or instructional complexity. Future system design should focus on interaction optimization and adaptive guidance to address current efficiency limitations while retaining the experiential benefits demonstrated.

7.1 Discussion

This section critically reflects on the research outcomes, contextualizing them within the broader frameworks of asynchronous collaboration, XR system design and CSCW theory. It interprets the empirical findings, evaluates the implemented system and explores conceptual, methodological and societal implications across multiple analytical levels.

Experimental Findings

The user study confirmed that immersive XR representations – particularly avatar-based asynchronous messaging – significantly enhanced spatial presence, clarity of communication and participant engagement, compared to a smartphone-based video baseline. Participants using the XR system exhibited fewer procedural errors, supporting prior research on embodied cognition and spatial memory in immersive environments [154]. These results validate the core hypothesis that XR can improve asynchronous collaboration on spatially complex tasks through embodied, persistent and context-rich interaction.

However, XR participants also required longer to complete the task, pointing to a trade-off between interaction richness and operational efficiency. This finding suggests that while immersive representations improve understanding, they can also increase task complexity due to navigation, control and cognitive processing demands.

System Implementation and Technical Lessons

The implementation of the ACIE framework across standalone XR headsets provided a functioning proof of concept for asynchronous interaction in immersive environments. Key features – such as avatar recording and playback, spatial annotations and version control – proved feasible and effective. Nonetheless, technical challenges emerged during deployment: synchronization issues between audio and motion, application crashes, avatar model glitches and UI occlusion reduced usability in some sessions.

These issues underscore the current technical limitations of XR platforms, particularly for persistent, multi-modal asynchronous systems. Improvements in spatial UI rendering, avatar proximity handling and system stability are required for production-ready deployments. Despite these limitations, the system remained usable and informative across diverse sessions, confirming its experimental robustness and operational capability.

Conceptual Insights and Theoretical Contributions

This research extends the understanding of asynchronous collaboration by integrating the 3C model – Communication, Cooperation, Coordination – into an immersive design framework. Each dimension was operationalized based on specific characteristics:

- **Communication:** Avatar-based messaging preserved gesture, gaze and timing.
- **Cooperation:** Proxy avatars and annotations enabled instruction and feedback across time.
- **Coordination:** Step-wise playback and version control facilitated alignment of actions.

Across design, implementation, and evaluation, immersive XR was found to improve procedural accuracy and perceived presence while maintaining usability and workload, at the cost of longer completion times relative to a smartphone-based 2D video baseline (❯ **Chapters 5** and **6**). Taken together, these results indicate that XR supports asynchronous collaboration on complex tasks by preserving spatial context and embodied cues, with an efficiency trade-off that warrants interaction optimization. Within the stated scope (tasks with strong spatial reasoning and embodied action), the research objective formulated in ❯ **Chapter 1** was advanced.

Overall, these mechanisms show that asynchronous collaboration does not have to compromise presence or effectiveness. Instead, well-designed XR systems can reconstruct the social and spatial context in deferred interactions. Consequently, this work challenges the traditional CSCW assumption that high-quality collaboration requires co-temporality.

Answer to main RQ (synthesis): Across design, implementation and evaluation, immersive XR was found to improve procedural accuracy and perceived presence while maintaining usability and workload, at the cost of longer completion times relative to a smartphone-based 2D video baseline (**Chapters 4**, **5**, **6**). Taken together, these results indicate that XR supports asynchronous collaboration on complex tasks by preserving spatial context and embodied cues, with an efficiency trade-off that warrants interaction optimization. Within the stated scope (tasks with strong spatial reasoning and embodied action), the research objective formulated in **Chapter 1** was advanced.

Positioning within the State of the Art

Prior work in XR-supported asynchronous collaboration has largely focused on isolated aspects, such as embodiment, annotation or replay, without fully addressing coordination or cooperation. As shown in Table 7.1, most systems lack full coverage of the collaborative functions required for complex task support.

In contrast, ACIE was designed as an integrated framework explicitly supporting all three 3C dimensions. It is the only system in the comparative analysis that scores across all five criteria-embodiment, complexity, communication, coordination and cooperation. This reflects a deliberate effort to close design gaps in the literature and move toward comprehensive asynchronous support in XR environments.

7.2 Contributions

Framed within Design Science Research (DSR) (**Chapter 1**, Section 1.4), this dissertation proceeded from the knowledge base to iterative design and instantiation, through empirical evaluation, and back to the knowledge base, with several cycles. Along the way, different contributions were made: **C1**

Table 7.1: Comparison of core XR systems for asynchronous collaboration using a five-dimension scoring framework, including the proposed ACIE system

Related Work	Embodiment	Task Complexity	Communication	Coordination	Cooperation	Score
Guided Tour [63]	●○○	●●○	●○○	●○○	○○○	5
Time Travellers [34]	●○○	●●●	●●○	●○○	●●○	9
MAVRC [36]	●○○	●●●	●●●	●○○	●○○	9
ReliveInVR [159]	●●●	●●○	●○○	●●○	●●○	10
XR-LIVE [154]	●●○	●●●	●○○	●●○	●●○	10
VRGit [172]	●●○	●●●	●○○	●●●	●●○	11
Virtual Triplets [171]	●●○	●●●	●●○	●●○	●●○	11
Loki [156]	●●●	●●●	●●○	●○○	●●○	11
Async Reality [52]	●●●	●●●	●●●	●●○	●●○	13
ACIE (this work)	●●●	●●●	●●●	●●●	●●●	15

Note: Each criterion is scored on a scale from 0 to 3 and visually encoded using circle symbols. An empty circle (○) indicates no or negligible support, while up to three filled circles (●) reflect increasing levels of functionality or completeness.

laid the conceptual groundwork and defined the design space, **C2** informed the realization and application of the approach, and **C3** evaluated the approach and fed the results back to the knowledge base.

C1 – Mapping and conceptualization: This work investigated how the 3Cs could be supported in XR and mapped established and ongoing XR approaches for asynchronous collaboration to requirements of complex engineering tasks. Based on this, individual components were brought together into a framework concept for ACCT – the ACIE framework (**Chapter 2**, **Chapter 3**, **Chapter 4**).

C2 – Implementation and application: The implementation and validation of the framework were reported, and a reference revision was provided as open source Mayer [1]. The framework and its components were applied to engineering use cases to show how it could be used for asynchronous collaboration in situ (**Chapter 5**, **Chapter 6**).

C3 – Evaluation and feedback to the knowledge base: The proposed approach was evaluated in a realistic experiment, comparing it with a baseline. The results were reported with regard to usability and performance, and the main findings were summarized together with design implications and further considerations for improvement (**Chapter 6**).

> **Research Objective:** Propose an XR-based approach to facilitate ACCT through improved representation and communication of spatial context.

Together, **C1–C3** contributed to the thesis objective. **C1** clarified what needed to be represented and communicated, and organized the design options around the 3Cs. **C2** showed how these approaches were implemented and combined for application in realistic engineering settings. Finally, **C3** evaluated the approach against a baseline and reported findings and implications that were fed back into the knowledge base.

7.3 Limitations

While this study provides a novel framework and promising empirical results, certain limitations impacting the technical robustness as well as the methodological and interpretive scope of the findings must be considered.

Technical Limitations: Despite functional implementation across multiple XR platforms, the prototype showed some stability and synchrony issues during operation. Audio and avatar playback were occasionally out of sync;

the application experienced crashes, tracking loss or avatar rendering failures, particularly under prolonged usage or occlusion conditions. Additionally, the usability and spatial behavior of certain interface elements need improvement. For example, the avatar interaction menu was sometimes rendered inside walls or objects, making it difficult to access. Similarly, the proxy avatars occasionally entered uncomfortable proximity to the user, leading to potential collisions or breaks in immersion. These issues underscore the importance of further refining spatial UI design, as well as robustness and runtime stability.

Methodological Limitations: The evaluation study was conducted with a moderately sized participant pool and used an artificially constructed engineering task. While this enabled control and replicability, it limits the ecological validity of the results. Although the task was representative of common spatial workflows, further testing in diverse, domain-specific scenarios – such as medical, architectural, or field-service contexts – would strengthen the generalizability of the findings. Participants were also predominantly non-expert users in XR, which may have influenced their performance or perceptions.

Measurement Limitations: Although the study employed validated instruments like the System Usability Scale (SUS), NASA-TLX and IPQ to assess usability and presence, some of the collaboration-related measures were collected using a custom-designed questionnaire. Despite its foundation in earlier studies, this instrument has not been validated in the academic literature. Therefore, it is essential to approach its findings with a degree of caution. Future studies should include standardized instruments or further validate custom scales through factor analysis or expert review.

These limitations are characteristic of early-stage research involving novel interactive systems and do not diminish the core contributions of this study. Instead, they emphasize practical challenges and design alternatives that could be considered in future iterations to enhance the system's robustness, user-friendliness and domain applicability.

Impact

The relevance and applicability of this research goes beyond academic experimentation. The ACIE framework was not only developed in collaboration with practising engineers in a research factory – a hybrid space that combines laboratory research and industrial practice – but is also actively used in operational environments.

The system is being integrated into the KIT-nova[1], a unique environment where education, collaboration and manufacturing meet. The ACIE system is used to create and deliver spatial safety training content and task guidance, enabling users to engage with complex procedures in a more embodied, persistent and context-rich manner.

In parallel, the ACIE framework is being explored for educational integration at the ATS Institute[2] at Mannheim University of Applied Sciences. As depicted in Figure 7.1, immersive avatar-based instructions are employed to digitalize hands-on pump-station procedures, allowing students to repeatedly review and practice critical operations directly at the physical pump system during their practical coursework.

This direct transfer into applied use demonstrates the framework's practical viability and supports its long-term relevance for XR-supported knowledge sharing, safety and education beyond the industrial domain.

This thesis contributes to the European Union's Sustainable Development Goals (SDGs)[3] by advancing digital, educational and sustainable practices in industrial collaboration and training. Figure 7.2 presents the variety of Sustainable

[1] KIT-nova is the *Learning and Application Center for Mechatronics* at KIT, where student co-working spaces meet real production systems. ACIE is used there for immersive safety instructions and training in VR/MR. See: https://www.laz.kit.edu/english/index.php

[2] ATS – Institut für Anlagentechnik und Anlagensicherheit, Technische Hochschule Mannheim: https://www.ats.hs-mannheim.de/

[3] International Sustainable Development Goals by the UN: international-partnerships.ec.europa.eu

Figure 7.1: Application of immersive avatar-based instruction for pump-station procedures at the ATS Institute at TH Mannheim

Development Goals (SDGs) that frame global efforts toward inclusive, responsible, and future-ready innovation. The proposed XR-based system supports several core SDGs through its conceptual design and practical deployment:

- **SDG 4 – Quality Education:** By enabling immersive, persistent and replayable training scenarios, the ACIE framework enhances procedural understanding and supports accessible, repeatable learning methods in engineering contexts.
- **SDG 9 – Industry, Innovation and Infrastructure:** The system promotes digital innovation in industrial workflows by introducing asynchronous collaboration tools that improve coordination, reduce errors and extend the usability of XR into production-related environments.
- **SDG 12 – Responsible Consumption and Production:** ACIE reduces the need for physical co-presence and paper-based documentation through digital guidance, thus supporting resource-efficient training and minimizing wasteful redundancy in industrial learning.

Figure 7.2: Overview of the Sustainable Development Goals (SDGs), including SDG 4, SDG 9, and SDG 12 to which this work contributes

Together, these impacts show the potential of immersive, asynchronous collaboration systems to shape future industrial workflows and educational practices in alignment with technological innovation and social responsibility.

7.4 Outlook

Building on the findings and technical foundation of this thesis, several promising research directions emerge that can shape future XR-based collaboration and learning technologies.

Societal Impact and Educational Integration: The broader potential of this research lies in its contribution to sustainable, accessible education and industrial transformation. XR-based asynchronous systems could democratize training across geographical locations and different knowledge and skill levels. Applying the ACIE framework in learning centers like KIT-nova or therapeutic institutions could foster inclusive, embodied learning environments, aligning with the digital transformation goals of the European education and innovation agenda. These efforts also support upskilling in technical professions and immersive safety training, helping to close knowledge gaps in increasingly complex workflows.

Beyond these initial domains, the approach is broadly applicable across sectors such as manufacturing, robotics, healthcare, cultural heritage, and logistics. This cross-domain flexibility makes immersive asynchronous systems a powerful tool for lifelong learning and scalable knowledge transfer.

WebXR, Robotics and Data-Centric Infrastructures: Achieving such broad accessibility and scalability requires robust and flexible technical foundations. One promising direction is the deployment of WebXR-based platforms, which allow immersive content to run directly in browsers across devices – without the need for specialized applications or hardware. This significantly lowers the barrier to access and fosters more inclusive participation in XR-based collaboration and training.

Additionally, 5G and edge computing enable low-latency capture, processing, and streaming of immersive content directly in the field. This technological backbone not only enhances the scalability and technical feasibility of asynchronous and remote collaboration, but also supports distributed simulations and real-time synchronization of digital twins across geographically dispersed sites.

In parallel, robotic digital twins could simulate human procedures based on ACIE-style recordings, generating high-quality synthetic datasets to train spatially- and socially-aware physical agents. As AI and robotics become increasingly embedded in everyday life – across domains such as healthcare, logistics, industrial maintenance, and personal assistance – the demand for context-rich, embodied training data will continue to grow. Systems like ACIE provide a scalable means of acquiring such spatially grounded data from real-world interactions.

These data-centric infrastructures are particularly relevant for procedural automation, remote diagnostics, and the advancement of embodied AI systems –

laying the groundwork for more intelligent, adaptive, and context-aware collaborative technologies.

Seamless XR Guidance in Industrial Spaces: For widespread adoption in industrial settings, XR guidance systems must integrate seamlessly into real-world workflows – whether through Spatial Augmented Reality (SAR)[4] or through immersive asynchronous systems like the ACIE approach presented in this thesis. Both aim to provide intuitive, in-situ assistance without disrupting the user's physical environment, while capturing not only human actions but also the evolving state of the environment – including spatial layout, equipment identifiers, process parameters, and other task-relevant metadata that are not directly visible.

Volumetric capture further exemplifies this direction by enabling accurate, holistic representations of full-body movement and appearance – without requiring wearable trackers (Figure 7.3).[5] Unlike conventional XR avatars, which rely on limited input data and often simplify body geometry and motion, volumetric representations preserve the detailed physicality and visual fidelity of human actions. This makes them particularly well-suited for high-precision tasks and scenarios where authenticity, realism, and trust are critical. When combined with environment capture and semantic overlays (e.g. CAD/DT links, tolerances, and sensor readouts), non-visible context can be exposed directly in the workspace to support reliable, auditable execution.

From Replay to Responsive Agents – Adaptive Human-AI Interaction: Looking further ahead, immersive asynchronous systems are likely to

[4] SAR enables projection-based augmentation directly onto real-world surfaces, offering spatially aligned guidance without requiring head-mounted displays.

[5] VOHLO is a project developed at Vicomtec that provides an end-to-end platform for capturing, analyzing, enhancing, and rendering realistic 3D human recordings in collaborative settings. See: `https://www.vicomtech.org/en/rdi-tangible/projects/project/voluai-volumetric-video-on-the-web`.

Figure 7.3: Volumetric human motion capture: compared to conventional avatars, volumetric recordings provide a holistic, trackerless, and highly accurate representation of both movement and appearance.

evolve from static playback tools into interactive, knowledge-driven agents. The ability to recontextualize previously recorded behavior using AI and machine learning not only supports knowledge preservation, but also enables scalable transfer across organizational boundaries. Such systems could assist learners through reflection, provide on-demand support for novices in the field, and foster human–AI teaming based on real-world, domain-specific expertise. This shift represents a promising convergence of XR embodiment and AI-driven personalization – paving the way for more intelligent, responsive, and sustainable knowledge transfer ecosystems.

By leveraging recorded data – whether avatar-based or volumetrically captured – future systems could browse past interactions to provide contextual, dialog-based guidance. LLM-powered agents would allow users to ask clarifying questions and receive animated responses, especially when combined with trained motion models. Generative models trained on recordings could further provide estimated responses, helping to bridge the gap of asynchrony. However, such synthesized responses may be inaccurate – or in some cases, even harmful – if applied blindly. Therefore, their use must be approached with caution. Future research should explore in which contexts AI-generated agent responses are appropriate, when actual human responses are necessary, and how both types can be meaningfully combined in different collaborative or instructional scenarios.

7.5 Closing Remarks

The findings of this thesis mark a step toward more human-centered and spatially embedded collaboration technologies. As the demand for flexible work and asynchronous workflows continues to grow, the ACIE framework offers a validated conceptual and technical foundation for future XR systems – enabling users not just to communicate, but to engage, interact, and co-create across time and space. By bridging immersive interaction with real-world application, this work contributes to the evolution of collaborative systems that are not only technically innovative, but also accessible, adaptable, and aligned with the broader goals of sustainable and inclusive digital transformation.

A Appendix

A.1 Search Keywords and Boolean Strings

The following keyword sets and Boolean expressions were used during the structured search (❯ **Chapter 3**). For transparency and reproducibility, exact strings are listed here.

Core: "asynchronous collaboration"

XR: "extended reality", "virtual reality", "augmented reality", "mixed reality", "XR", "VR", "AR", "MR", "embodiment", "presence"

3C: "coordination", "cooperation", "communication"

Context: "manufacturing", "production", "robotics", "maintenance", "assembly", "procedural tasks"

Combined search term: "asynchronous collaboration" AND ("extended reality" OR "virtual reality" OR "augmented reality" OR "mixed reality" OR "XR" OR "VR" OR "AR" OR "MR") AND ("coordination" OR "cooperation" OR "communication") AND ("manufacturing" OR "production" OR "robotics" OR "maintenance" OR "assembly" OR "procedural tasks")

A.2 Scoring of XR AC Systems

Table A.1: This scoring was used for the abstract comparison of XR asynchronous collaboration systems in ❯ **Chapter 3**

Dimension	0–3 Scale (summary criteria)
Embodiment	0: none; 1: minimal (e.g. head/hand only); 2: moderate (expressive gestures, ghost replays); 3: high-fidelity (full-body or volumetric, situated).
Task Complexity	0: trivial or static content only; 1: simple, step-wise tasks; 2: moderate spatial/procedural complexity; 3: complex spatial, physical, cognitive, and technical tasks.
Communication	0: none; 1: basic, one-way; 2: richer cues, limited reactivity; 3: expressive, multi-modal, contextually anchored.
Coordination	0: none; 1: basic sequencing; 2: temporal navigation or checkpoints; 3: versioning/branching, state restoration, explicit workflow support.
Cooperation	0: individual consumption only; 1: minimal (e.g., annotations); 2: task continuation; 3: joint editing, branching, feedback loops.
Overall score interpretation: 0–5 limited; 6–10 moderate; 11–15 good; 16–18 excellent	

A.3 Avatar Bone and Expression Mappings

Skeletal Mesh Structures

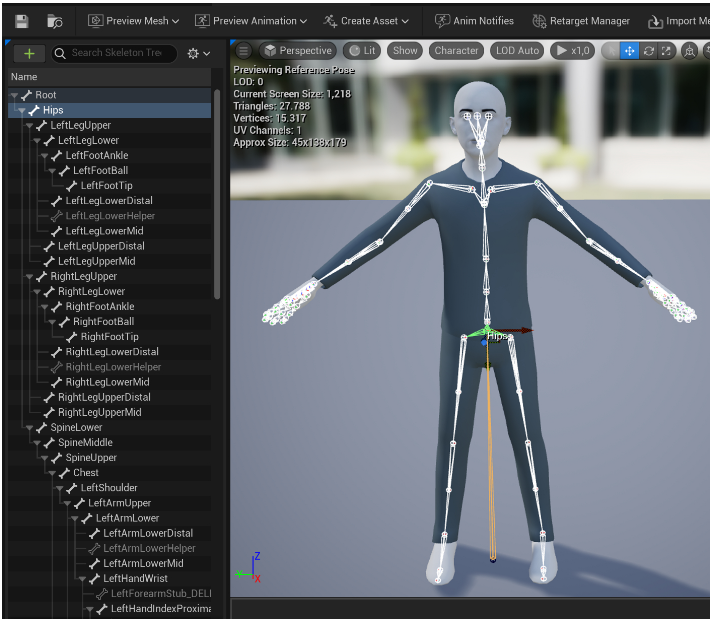


Figure A.1: MetaXR skeletal mesh structure in the Unreal Engine Editor view

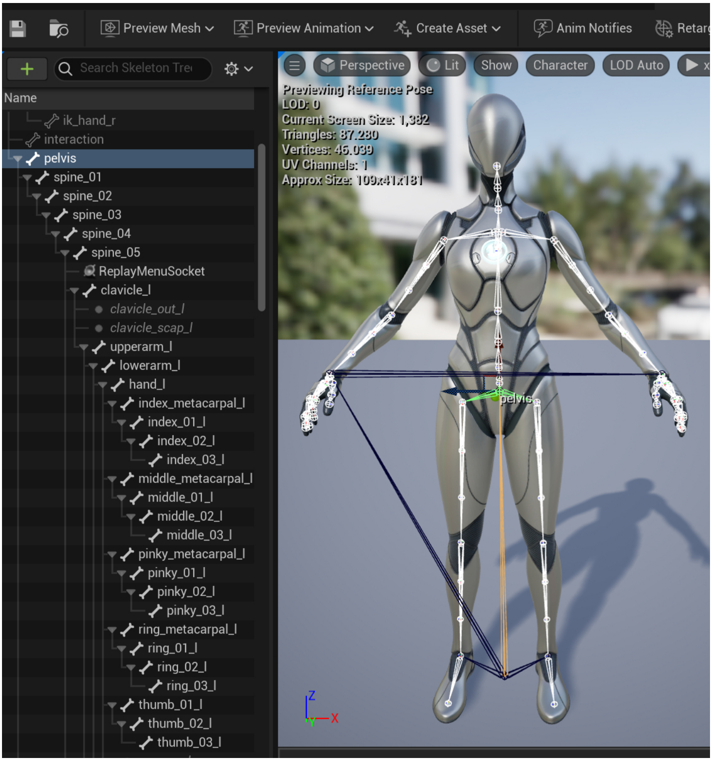


Figure A.2: UE Quinn skeletal mesh structure in the Unreal Engine Editor view

Avatar Skeletal Mapping

Table A.2: Skeletal joint mapping between MetaXR Avatar, UE Metahuman, and UE Mannequin. Body Joint Visual Reference for Movement SDK for Unreal: `https://developers.meta.com/horizon/documentation/unreal/move-ref-body-joints`

Movement SDK	Metahuman	Skeletal Mesh
XR_BODY_JOINT_ROOT_FB	root	root
XR_BODY_JOINT_HIPS_FB	pelvis	pelvis
XR_BODY_JOINT_SPINE_LOWER_FB	spine_01	spine_01
XR_BODY_JOINT_SPINE_MIDDLE_FB	spine_02	spine_02
XR_BODY_JOINT_SPINE_UPPER_FB	spine_03	spine_03
XR_BODY_JOINT_HEAD_FB	head	head
...	...	...
thigh_l	thigh_l	thigh_l
calf_l	calf_l	calf_l
foot_l	foot_l	foot_l

Hand Tracking Mapping

Table A.3: Skeletal mapping between UE5 hands and Meta Quest Pro hand tracking. Body Joint Visual Reference for Movement SDK for Unreal: `https://developers.meta.com/horizon/documentation/unreal/move-ref-body-joints`

UE5 Skeletal Mesh Hierarchy	Meta Quest Pro Hand Tracking Data
hand_l	Wrist Root
thumb_01_l	Thumb0
thumb_02_l	Thumb1
thumb_03_l	Thumb2
metacarpal_index_l	Index1
index_01_l	Index1
index_02_l	Index2
index_03_l	Index3
metacarpal_middle_l	Middle1
middle_01_l	Middle1
middle_02_l	Middle2
middle_03_l	Middle3
metacarpal_ring_l	Ring1
ring_01_l	Ring1
ring_02_l	Ring2
ring_03_l	Ring3
metacarpal_pinky_l	Pinky0
pinky_01_l	Pinky1
pinky_02_l	Pinky2
pinky_03_l	Pinky3

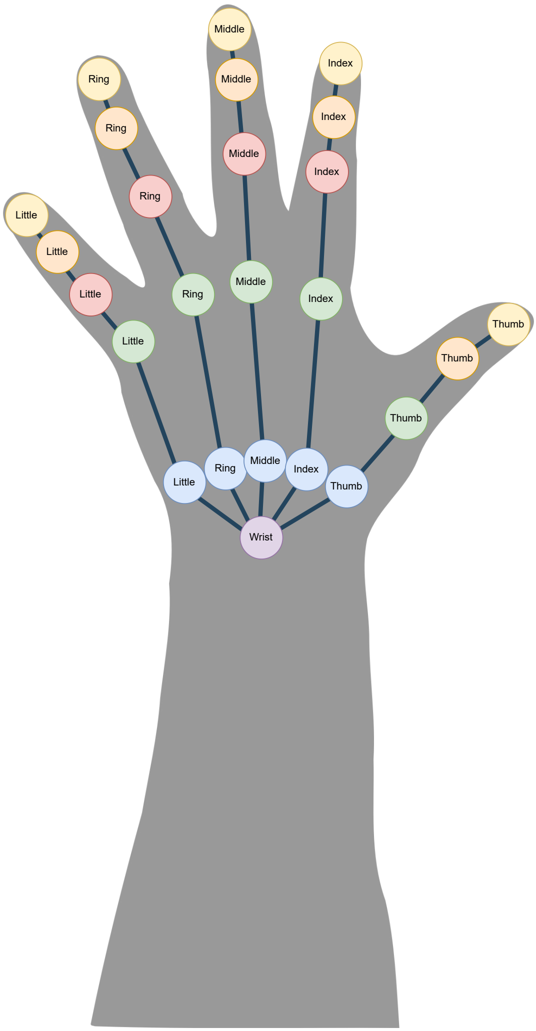


Figure A.3: Hand bones and names reference illustration

Facial Expression Mapping

Table A.4: Facial expression mapping between Metahuman expressions and MetaXR blendshapes. MetaXR Blendshape Visual Reference—Movement SDK for Unreal: `https://developers.meta.com/horizon/documentation/unreal/move-ref-blendshapes`

Metahuman Expression Pose	**Movement SDK Blendshape**
BrowDownLeft	BROW_LOWERER_L
CheekPuff (both)	CHEEK_PUFF_L
CheekSquintLeft	CHEEK_RAISER_L
–	CHEEK_SUCK_L
MouthShrugLower	CHIN_RAISER_B
...	...
MouthUpperUpLeft	UPPER_LIP_RAISER_L

A.4 Evaluation Appendix

Statistical Testing

Table A.5: Overview of statistical tests used for data analysis, including checks for normality, variance, and appropriate group comparison methods

Step	Test	Purpose
Normality Check	Shapiro-Wilk Test	Assesses whether the data follows a normal distribution. If $p > 0.05$, normality is assumed.
	Mann-Whitney U Test	Used if normality is violated in one or both groups. This non-parametric test compares distributions without assuming normality.
Variance Check	Levene's Test	Assesses homogeneity of variances. If $p > 0.05$, variances are considered equal.
	Welch's t-Test	Used if normality is satisfied but variances are unequal.
Group Comparison	Independent Samples t-Test	Used when both normality and equal variances are satisfied to compare group means.
	Mann-Whitney U Test	Used for non-normal distributions to compare group medians.

Normality Check: Normality was assessed separately for each group using the Shapiro-Wilk test. If the p-value for Shapiro-Wilk was greater than 0.05 ($p > 0.05$), the variable was assumed to follow a normal distribution. If the p-value was less than or equal to 0.05 ($p \leq 0.05$), normality was considered violated.

Homogeneity of Variance: For variables that passed the normality check, Levene's test was conducted to assess the equality of variances. If the p-value for Levene's test was greater than 0.05 ($p > 0.05$), variances were assumed to be equal. If the p-value was less than or equal to 0.05 ($p \leq 0.05$), unequal variances were assumed, and Welch's t-test was used instead of the standard independent samples t-test.

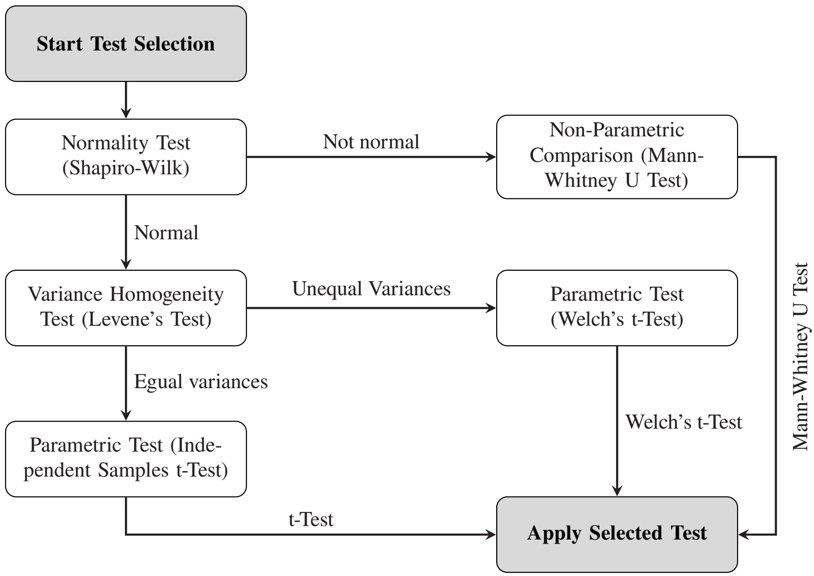


Figure A.4: Decision logic for selecting appropriate statistical tests based on data distribution and variance assumptions

Subjective Questionnaires

Table A.6: Demographics and Preferences Questionnaire

Category	ID	Item
Demographic	1	What is your age (in years)?
	2	What is your gender identity? (Male / Female / Non-binary / Prefer to self-describe: ___ / Prefer not to say)
	3	What is the highest level of education you are currently pursuing or have completed? (Bachelor's / Master's / Doctoral / Other: ___)
	4	What is your field of study or profession?
	5	Are you left-handed, right-handed, or ambidextrous?
	6	Do you usually wear glasses? (Yes/No)
Tech BG & Exp.	7	Your familiarity with computer technology (including smartphones).
	8	The amount of time you typically spend using digital devices.
	9	Your experience with video gaming.
	10	How frequently you play video games.
	11	Your experience with Augmented Reality (AR), Virtual Reality (VR), or Mixed Reality (MR).
	12	How frequently you use AR, VR, or MR technologies.
	13	Your experience with robotics.
	14	How frequently you interact with robots.
	15	Your comfort level when using new technology.
	16	Your level of discomfort (e.g., cybersickness, nausea, headache) during the experiment.
Setup & Menu	17	Which prototype did you test?
	18	During the experiment, did you wear glasses with the Headset? (Contact lenses do not count)
	19	How comfortable did you feel using the Menu to view, create, and navigate the recordings?
	20	Why? (open-ended)
Prefs & Qual.	21	Which medium would you prefer for asynchronous collaboration with the Expert (when you are the Novice)?
	22	Why? (open-ended)
	23	Which medium would you prefer for asynchronous collaboration with the Novice (when you are the Expert)?
	24	Why? (open-ended)
	25	Which medium would you prefer for learning the process?
	26	Why? (open-ended)
	27	How did you feel about the avatar?
	28	Why? (open-ended)
	29	Do you have any additional comments?

Table A.7: Custom Collaboration Questionnaire Items

Category	ID	Item
Communication	1	The [video/avatar] instructions were clear and easy to understand.
	2	The [video/avatar] provided a detailed explanation of the tasks.
	3	I was satisfied with the quality of the information provided by the [video/avatar].
	4	The [video/avatar] effectively communicated the necessary steps for completing the task.
Cooperation	5	The [video/avatar] Question and Response feature effectively facilitated collaboration with the instructor.
	6	The responses I received from the [video/avatar] were helpful and informative.
	7	I was satisfied with the level of cooperation enabled by the [video/avatar] during the task.
Coordination	8	The step-by-step paradigm helped me effectively organize my tasks.
	9	I was able to sequence my tasks efficiently using the step-by-step paradigm.
	10	It was easy to navigate through the tasks using the step-by-step paradigm.
Scoring (Likert 1–5): 1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree		

Table A.8: Comprehensive results summary of MR Avatar and 2D Video conditions

Measure	**MR Avatar**		**2D Video**		**p**	**r**
	M	SD	M	SD		
ASQ	4.061	0.521	4.048	0.825	0.604	0.079
SUS	77.614	12.308	79.048	16.630	0.422	0.123
Weighted TLX Score	30.697	10.573	29.714	12.171	0.389	0.044
Mental Demand	35.455	18.703	35.000	22.638	0.713	0.056
Physical Demand	17.500	14.618	15.000	19.875	0.127	0.233
Temporal Demand	17.955	20.567	19.048	24.475	0.971	0.006
Performance	75.455	18.186	80.476	16.651	0.306	0.156
Effort	33.636	16.986	28.571	22.257	0.221	0.187
Frustration	22.500	14.454	22.143	20.711	0.591	0.082
UEQ*	1.496	0.585	1.007	1.007	0.027*	0.331
Attractiveness	1.598	0.998	1.159	1.058	0.116	0.234
Efficiency	1.057	0.829	0.702	0.984	0.238	0.176
Perspicuity	1.898	0.651	2.083	0.733	0.199	0.192
Dependability	1.352	0.597	1.321	0.958	0.891	0.020
Stimulation*	1.420	0.934	0.845	1.214	0.041†	0.263
Novelty* **	1.648	0.789	-0.071	1.163	$<$.001***	0.664
Collaboration	4.159	0.120	4.167	0.135	0.779	0.043
Communication	3.852	0.722	4.155	0.893	0.120	0.237
Cooperation	4.364	0.534	4.079	0.809	0.348	0.143
Coordination	4.364	0.650	4.270	0.727	0.774	0.044
REAL*	3.489	0.657	1.333	1.155	0.024*	0.470
SP**	3.258	1.191	2.125	2.097	0.002**	0.570
INV***	3.409	0.599	0.600	0.894	$<$.001***	0.658
Co-Presence***	1.136	1.467	-1.031	1.717	$<$.001***	0.624
Perception of the Other	-0.114	1.420	-0.104	1.104	0.893	0.022
Attentional Engagement	1.561	1.091	1.047	1.286	0.252	0.179
Emotional Contagion	-1.148	0.981	-0.615	1.458	0.227	0.193
Comprehension	1.879	0.787	1.857	0.938	0.895	0.021
Behavioral Interdependence	1.303	0.942	1.028	1.449	0.822	0.035

* Statistically significant results.
** Highly significant result ($p < .01$).
*** Highly significant result ($p < .001$).
† Statistically significant difference could only be found in the one-tailed test.
$^{+}$ Large effect size ($r > .5$).

Own Publications

Journal Articles

[1] Jakob Bönsch, Lucas Greif, Svenja Hauck, Simon Kreuzwieser, Anjela Mayer, Felix Longge Michels, and Jivka Ovtcharova. Virtual engineering: Hands-on integration of product lifecycle management, computer-aided design, extended reality, and artificial intelligence in engineering education. *Chemie Ingenieur Technik*, 96(11):1460–1474, 2024. ISSN 0009-286X, 1522-2640. doi:10.1002/cite.202300169.

[2] Anjela Mayer, Lucas Greif, Tim Markus Häußermann, Simon Otto, Kevin Kastner, Sleiman El Bobbou, Jean-Rémy Chardonnet, Julian Reichwald, Jürgen Fleischer, and Jivka Ovtcharova. Digital twins, extended reality, and artificial intelligence in manufacturing reconfiguration: A systematic literature review. *Sustainability*, 17(5), 2025. ISSN 2071-1050. doi:10.3390/su17052318.

Book Chapters

[1] Simon Kreuzwieser, Anjela Mayer, Matthes Elstermann, and Jivka Ovtcharova. Collaboration, artificial intelligence and augmented reality enable purchasing x.0. In Florian Schupp and Heiko Wöhner, editors, *Digitalization in Procurement*, pages 265–282. Springer Fachmedien Wiesbaden, Wiesbaden, 2024. ISBN 978-3-658-45670-2. doi:10.1007/978-3-658-45670-2_16.

[2] Anjela Mayer, Jean-Rémy Chardonnet, Polina Häfner, and Jivka Ovtcharova. Collaborative work enabled by immersive environments. In Alexandra Shajek and Ernst Andreas Hartmann, editors, *New Digital Work: Digital Sovereignty at the Workplace*, pages 87–117. Springer International Publishing, Cham, 2023. ISBN 978-3-031-26490-0. doi:10.1007/978-3-031-26490-0_6.

[3] Anjela Mayer, Simon Otto, Jean-Rémy Chardonnet, Alexander Puchta, Jürgen Fleischer, and Jivka Ovtcharova. Extended intelligence for rapid cognitive reconfiguration. *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 120(s1):177–180, 2025. doi:10.1515/zwf-2024-0159.

Conference Papers

[1] Sebastian Amann, Nicolai Beisheim, Polina Haefner, Victor Haefner, Kevin Kastner, Anjela Mayer, Felix Longge Michels, and Tobias Ott. Using collaborative immersive environments and building information modeling technology for holistic planning of production lines. 41:443 – 452, 2023. ISSN 2352-751X. doi:10.3233/ATDE230638.

[2] Kevin Kastner, Ines Miguel-Alonso, Anjela Mayer, Jean-Rémy Chardonnet, Andres Bustillo, Jivka Ovtcharova, René Eiswirth, and Julian Reichwald. Evaluating the effectiveness of avatar-based collaboration in xr for pump station training scenarios. In *2024 IEEE 3rd German Education Conference (GECon)*, pages 1–7, 2024. doi:10.1109/GECon62014.2024.10734032.

[3] Tim Krodel, Vera Schott, Anjela Mayer, and Jivka Ovtcharova. Impact of xr-enabled collaboration in businesses—an economic, ecological, and social perspective. In Bahaaeddin Alareeni and Islam Elgedawy, editors, *AI and Business, and Innovation Research: Understanding the Potential and Risks of AI for Modern Enterprises*, pages 767–777. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-42085-6. doi:10.1007/978-3-031-42085-6_66.

[4] Anjela Mayer, Victor Häfner, and Jivka Ovtcharova. Extending opensg for real-time synchronization of immersive environments in distributed collaboration. In *ConVRgence (VRIC) Virtual Reality International Conference Proceedings*, pages 6–15, 2021. doi:10.20870/IJVR.2021.1.1.4797.

[5] Anjela Mayer, Théo Combe, Jean-Rémy Chardonnet, and Jivka Ovtcharova. Asynchronous manual work in mixed reality remote collaboration. In *Extended Reality : First International Conference, XR Salento 2022, Lecce, Italy, July 6–8, 2022, Proceedings, Part II. Ed.: L. De Paolis*, volume 13446 of *Lecture Notes in Computer Science*, page 17–33. Springer, 2022. ISBN 978-3-031-15552-9. doi:10.1007/978-3-031-15553-6_2.

[6] Anjela Mayer, Kevin Kastner, Julian Reichwald, Jean-Rémy Chardonnet, and Jivka Ovtcharova. Comparing avatar and face-to-face collaboration in vr education: Concept and preliminary insights. In *2023 IEEE 2nd German Education Conference (GECon)*, pages 1–5, 2023. doi:10.1109/GECon58119.2023.10295113.

[7] Anjela Mayer, Izel Kilinc, Kevin Sprügel, and Polina Häfner. Towards research gaps in collaborative virtual reality environments for education: A literature review. In Michael E. Auer, Reinhard Langmann, and Thrasyvoulos Tsiatsos, editors, *Open Science in Engineering*, pages 513–525. Springer Nature Switzerland, Cham, 2023. ISBN 978-3-031-42467-0. doi:10.1007/978-3-031-42467-0_47.

[8] Anjela Mayer, Augustin Rungeard, Jean-Rémy Chardonnet, Polina Häfner, and Jivka Ovtcharova. Immersive hand instructions in ar – insights for asynchronous remote collaboration on spatio-temporal manual tasks. In *2023 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pages 1–6, 2023. doi:10.1109/CIVEMSA57781.2023.10231018.

[9] Anjela Mayer, Kevin Kastner, Edgar Mühlbeier, Jean-Rémy Chardonnet, Julian Reichwald, Alexander Puchta, Jürgen Fleischer, and Jivka Ovtcharova. Towards reconfigurable cyber-physical-human systems: Leveraging mixed reality and digital twins to integrate human operations. *Procedia CIRP*, 130:524–531, 2024. ISSN 2212-8271.

doi:https://doi.org/10.1016/j.procir.2024.10.124. 57th CIRP Conference on Manufacturing Systems 2024 (CMS 2024).

[10] Anjela Mayer, Jivka Ovtcharova, Jean-Rémy Chardonnet, Sebastian Amann, Nicolai Beisheim, and Xianbiao Jiang. Extended reality methods for transdisciplinary asynchronous engineering. In *Engineering For Social Change – Proceedings of the 31st ISTE International Conference on Transdisciplinary Engineering, London, United Kingdom, 9-11 July 2024. Ed.: A. Cooper*, volume 60 of *Advances in Transdisciplinary Engineering*, page 926 – 935. IOS Press, 2024. ISBN 978-1-64368-550-2. doi:10.3233/ATDE240947.

[11] Anjela Mayer, Ines Miguel-Alonso, Jean-Rémy Chardonnet, Andres Bustillo, and Jivka Ovtcharova. Enhancing asynchronous learning in immersive environments: Exploring baseline modalities for avatar-based ar guidance. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 1358–1359, 2025. doi:10.1109/VRW66409.2025.00325.

[12] Arelia McKern, Anjela Mayer, Lucas Greif, Jean-Rémy Chardonnet, and Jivka Ovtcharova. Ai-based interactive digital assistants for virtual reality in educational contexts. In *2024 IEEE 3rd German Education Conference (GECon)*, pages 1–5, 2024. doi:10.1109/GECon62014.2024.10734030.

[13] Ines Miguel-Alonso, Anjela Mayer, Jean-Rémy Chardonnet, Andres Bustillo, and Jivka Ovtcharova. Avatars in immersive virtual reality education: Preliminary insights and recommendations from a systematic review. In Michael E. Auer, Reinhard Langmann, Dominik May, and Kim Roos, editors, *Smart Technologies for a Sustainable Future*, pages 267–274. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-61891-8. doi:10.1007/978-3-031-61891-8_26.

Source Code

[1] Anjela Mayer. Aciie – asynchronous collaboration in immersive environments. `https://codeberg.org/ImmersiveCreators/aciie`, 2025. Open-source research framework. DOI: 10.35097/gs7ge5w59we6ed26.

Bibliography

[1] José Abdelnour Nocera, Marta Kristín Lárusdóttir, Helen Petrie, Antonio Piccinno, and Marco Winckler, editors. *A Survey of Computer-Supported Remote Collaboration on Physical Objects*, volume 14144 of *Lecture Notes in Computer Science*. Springer Nature Switzerland, Cham, 2023. ISBN 978-3-031-42285-0 978-3-031-42286-7. doi:10.1007/978-3-031-42286-7.

[2] Leonor Adriana Cárdenas-Robledo, Óscar Hernández-Uribe, Carolina Reta, and Jose Antonio Cantoral-Ceballos. Extended reality applications in industry 4.0. – A systematic literature review. *Telematics and Informatics*, 73:101863, September 2022. ISSN 07365853. doi:10.1016/j.tele.2022.101863.

[3] Pierre America. Object-oriented software construction: By bertrand meyer. prentice-hall, hemel hempstead, united kingdom, 1988, price £19.95, isbn 0-13-629031-0. *Science of Computer Programming*, 12(1):88–90, 1989. ISSN 0167-6423. doi:https://doi.org/10.1016/0167-6423(89)90034-8.

[4] Christopher Andrews, Michael Southworth, Jennifer Silva, and Jonathan Silva. Extended reality in medical practice. *Current Treatment Options in Cardiovascular Medicine*, 21, 03 2019. doi:10.1007/s11936-019-0722-7.

[5] Bryony Archer, Scott D. Young, Gina M. Grinberg, and Juno Kim. Odour enhances the sense of presence in virtual reality. *PLOS ONE*, 17(3): e0265039, 2022. doi:10.1371/journal.pone.0265039.

[6] Bita Astaneh Asl and Carrie Sturts Dossick. Immersive vr versus bim for aec team collaboration in remote 3d coordination processes. 2022. doi:10.3390/buildings12101548.

[7] Ronald T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, 1997. doi:10.1162/pres.1997.6.4.355.

[8] David Baccarini. The concept of project complexity—a review. *International Journal of Project Management*, 14(4):201–204, 1996. ISSN 0263-7863. doi:10.1016/0263-7863(95)00093-3.

[9] Kaushal Kumar Bhagat, Wei-Kai Liou, and Chun-Yen Chang. A cost-effective interactive 3D virtual reality system applied to military live firing training. *Virtual Reality*, 20(2):127–140, June 2016. ISSN 1359-4338, 1434-9957. doi:10.1007/s10055-016-0284-x.

[10] Sharmistha Bhattacharya. Towards asynchronous vr collaboration with virtual agents as placeholder users. 2024. doi:10.1145/3652988.3673945.

[11] M. Billinghurst, S. Bee, J. Bowskill, and H. Kato. Asymmetries in collaborative wearable interfaces. In *Digest of Papers. Third International Symposium on Wearable Computers*, pages 133–140, 1999. doi:10.1109/ISWC.1999.806693.

[12] Mark Billinghurst and Hirokazu Kato. Collaborative mixed reality. In *Proceedings of the first international symposium on mixed reality*, volume 1, 1999. `https://www.academia.edu/download/3272154/Collaborative_mixed_reality.pdf`.

[13] Mark Billinghurst, Adrian Clark, and Gun Lee. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3): 73–272, 2015. doi:10.1561/1100000049.

[14] Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality*. A K Peters/CRC Press, 0 edition, August 2005. ISBN 978-1-4398-6494-4. doi:10.1201/b10624.

[15] Frank Biocca, Chad Harms, and Judee K. Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, 12(5):456–480, 10 2003. doi:10.1162/105474603322761270.

[16] Răzvan Gabriel Boboc, Florin Gîrbacia, and Eugen Valentin Butilă. The Application of Augmented Reality in the Automotive Industry: A Systematic Literature Review. *Applied Sciences*, 10(12), January 2020. ISSN 2076-3417. doi:10.3390/app10124259.

[17] Zahra Borhani. Annotation in asynchronous collaborative immersive analytic environments using augmented reality. 2022. doi:10.1109/VRW55335.2022.00326.

[18] Lyle Bourne, James Kole, and Alice Healy. Expertise: defined, described, explained. *Frontiers in Psychology*, 5:186, 03 2014. doi:10.3389/fpsyg.2014.00186.

[19] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. An introduction to 3-d user interface design. volume 10, pages 96–108. 2001. doi:10.1162/105474601750182342.

[20] Margot Brereton and Ben McGarry. An observational study of how objects support engineering design thinking and communication: implications for the design of tangible media. CHI '00, page 217–224, New York, NY, USA, 2000. Association for Computing Machinery. ISBN 1581132166. doi:10.1145/332040.332434.

[21] Stephen A. Brewster, Peter C. Wright, and Alistair D. N. Edwards. Experimentally derived guidelines for the creation of earcons. In *Proceedings of BCS HCI 1995 (People and Computers XI)*, pages 155–166, 1996. doi:10.1007/978-1-4471-3588-3_12.

[22] John Brooke. *SUS – a quick and dirty usability scale*, pages 189–194. 01 1996. `https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%2528sus%2529_comp%255B1%255D.pdf`.

[23] Alisa Burova, John Mäkelä, Hanna Heinonen, Paulina Becerril Palma, Jaakko Hakulinen, Viveka Opas, Sanni Siltanen, Roope Raisamo, and Markku Turunen. Asynchronous industrial collaboration: How virtual reality and virtual tools aid the process of maintenance method development and documentation creation. 2022. doi:10.1016/j.compind.2022.103663.

[24] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, Jose Neira, Ian Reid, and John J. Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016. doi:10.1109/TRO.2016.2624754.

[25] Filipe Calegario, João Tragtenberg, Johnty Wang, Ivan Franco, Eduardo Meneses, and Marcelo M. Wanderley. Open source dmis: Towards a replication certification for online shared projects of digital musical instruments. In Constantine Stephanidis, Aaron Marcus, Elizabeth Rosenzweig, Pei-Luen Patrick Rau, Abbas Moallem, and Matthias Rauterberg, editors, *HCI International 2020 - Late Breaking Papers: User Experience Design and Case Studies*, pages 84–97. Springer International Publishing, Cham, 2020. ISBN 978-3-030-60114-0. doi:10.1007/978-3-030-60114-0_5.

[26] Xaviéra Calixte and Pierre Leclercq. The interactive projection mapping as a spatial augmented reality to help collaborative design: Case study in architectural design. pages 143–152, 2017. doi:10.1007/978-3-319-66805-5_18.

[27] Luis M. Camarinha-Matos, Rosanna Fornasiero, Javaneh Ramezani, and Filipa Ferrada. Collaborative networks: A pillar of digital transformation. *Applied Sciences*, 9(24), 2019. ISSN 2076-3417. doi:10.3390/app9245431.

[28] Donald J. Campbell. Task complexity: A review and analysis. *The Academy of Management Review*, 13(1):40–52, 1988. ISSN 03637425. http://www.jstor.org/stable/258353.

[29] João Pedro Assunção Campos, Guilherme Afonso Madalozzo, Ana Luisa Sant'Anna Alves, and Rafael Rieder. Arfood: an augmented-reality food diary app for asynchronous collaborative interaction. 2024. doi:10.5753/jis.2024.4346.

[30] David Checa, Ines Miguel-Alonso, and Andres Bustillo. Immersive virtual-reality computer-assembly serious game to enhance autonomous learning. *Virtual Real.*, 27(4):3301–3318, December 2021. ISSN 1359-4338. doi:10.1007/s10055-021-00607-1.

[31] Wenqian Chen, Yifei Shan, Yue Wu, Zihan Yan, and Xiangdong Li. Design and evaluation of a distance-driven user interface for asynchronous collaborative exhibit browsing in an augmented reality museum. 2021. doi:10.1109/ACCESS.2021.3080286.

[32] Hyunwoo Cho, Eunhee Chang, Zhuang Chang, Jiashuo Cao, Bowen Yuan, Jonathon Derek Hart, Gun A. Lee, Thammathip Piumsomboon, and Mark Billinghurst. Hybrid cross-reality collaborative system. 2023. doi:10.1109/ISMAR-Adjunct60411.2023.00170.

[33] Hyunwoo Cho, Bowen Yuan, Jonathon Derek Hart, Eunhee Chang, Zhuang Chang, Jiashuo Cao, Gun A. Lee, Thammathip Piumsomboon, and Mark Billinghurst. An asynchronous hybrid cross reality collaborative system. 2023. doi:10.1109/ISMAR-Adjunct60411.2023.00022.

[34] Hyunwoo Cho, Bowen Yuan, Jonathon Derek Hart, Zhuang Chang, Jiashuo Cao, Eunhee Chang, and Mark Billinghurst. Time travellers: An asynchronous cross reality collaborative system. 2023. doi:10.1109/ISMAR-Adjunct60411.2023.00186.

[35] Hyunwoo Cho, Eunhee Chang, Bowen Yuan, Gun A. Lee, Thammathip Piumsomboon, and Mark Billinghurst. Bridging time and space: A novel cross-reality interface for asynchronous task guidance. *International Journal of Human–Computer Interaction*, 0(0):1–22, 2025. doi:10.1080/10447318.2025.2499159.

[36] Kevin Chow, Caitlin Coyiuto, Cuong Nguyen, and Dongwook Yoon. Challenges and design considerations for multimodal asynchronous collaboration in VR. *Proceedings of the ACM on Human-Computer Interaction*, 3 (CSCW), 2019. ISSN 25730142. doi:10.1145/3359142.

[37] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. The cave: audio visual experience automatic virtual environment. *Commun. ACM*, 35(6):64–72, June 1992. ISSN 0001-0782. doi:10.1145/129888.129892.

[38] Ryan Anthony de Belen, Huyen Nguyen, Daniel Filonik, Dennis Favero, and Tomasz Bednarz. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. *AIMS Electronics and Electrical Engineering*, 3:181–223, 2019. doi:10.3934/ElectrEng.2019.2.181.

[39] Sara Deflorio, Paolo M. Rossini, Italo Porto, Simone Rossi, and Lorenzo Ricci. Mechanoreceptors and tactile perception: From receptors to neural coding. *Frontiers in Human Neuroscience*, 16:862344, 2022. doi:10.3389/fnhum.2022.862344.

[40] Iñigo Fernández del Amo, Elisa Galeotti, Riccardo Palmarini, Gino Dini, John Erkoyuncu, and Rajkumar Roy. An innovative user-centred support tool for augmented reality maintenance systems design: a preliminary study. *Procedia CIRP*, 70:362–367, 2018. ISSN 2212-8271. doi:10.1016/j.procir.2018.02.020.

[41] Marika Demers, Ophélie Martinie, Carolee Winstein, and Maxime T. Robert. Active video games and low-cost virtual reality: An ideal therapeutic modality for children with physical disabilities during a global pandemic. 2020. doi:10.3389/fneur.2020.601898.

[42] Jozef Doboš and Anthony Steed. 3d revision control framework. In *Proceedings of the 17th International Conference on 3D Web Technology*, Web3D '12, page 121–129, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450314329. doi:10.1145/2338714.2338736.

[43] Sanika Doolani, Callen Wessels, Varun Kanal, Christos Sevastopoulos, Ashish Jaiswal, Harish Nambiappan, and Fillia Makedon. A Review of Extended Reality (XR) Technologies for Manufacturing Training. *Technologies*, 8(4):77, 2020. ISSN 2227-7080. doi:10.3390/technologies8040077.

[44] Paul Dourish. *Where the Action Is: The Foundations of Embodied Interaction.* The MIT Press, 09 2001. ISBN 9780262256056. doi:10.7551/mitpress/7221.001.0001.

[45] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, pages 107–114, 1992. doi:10.1145/143457.143468.

[46] Clarence A. Ellis, Simon J. Gibbs, and Gail Rein. Groupware: some issues and experiences. *Commun. ACM*, 34(1):39–58, January 1991. ISSN 0001-0782. doi:10.1145/99977.99987.

[47] Mica R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64, 1995. doi:10.1518/001872095779049543.

[48] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131, 2019. doi:10.1016/j.ijhcs.2019.05.011.

[49] Selim Erol, Andreas Jäger, Philipp Hold, Karl Ott, and Wilfried Sihn. Tangible industry 4.0: A scenario-based approach to learning for the future of production. *Procedia CIRP*, 54:13–18, 2016. ISSN 2212-8271. doi:https://doi.org/10.1016/j.procir.2016.03.162.

[50] S. Esser, S. Vilgertshofer, and A. Borrmann. Graph-based version control for asynchronous bim collaboration. *Adv. Eng. Informatics*, 53:101664, 2022. doi:10.1016/j.aei.2022.101664.

[51] Nico Feld and Benjamin Weyers. Overview of Collaborative Virtual Environments using Augmented Reality. page 10.18420/muc2019. Gesellschaft für Informatik e.V., 2019. doi:10.18420/muc2019-ws-617.

[52] Andreas Rene Fender and Christian Holz. Causality-preserving asynchronous reality. 2022. doi:10.1145/3491102.3501836.

[53] Shuo Feng, Weiping He, Xiaotian Zhang, Mark Billinghurst, and Shuxia Wang. A comprehensive survey on AR-enabled local collaboration. *Virtual Reality*, 27(4):2941–2966, December 2023. ISSN 1359-4338, 1434-9957. doi:10.1007/s10055-023-00848-2.

[54] Cristina Fenu and Fabio Pittarello. Svevo tour: The design and the experimentation of an augmented reality application for engaging visitors of a literary museum. *International Journal of Human-Computer Studies*, 114: 20–35, June 2018. ISSN 10715819. doi:10.1016/j.ijhcs.2018.01.009.

[55] Guo Freeman, Dane Acena, Nathan J. Mcneese, and Kelsea Schulenberg. Working together apart through embodiment. *Proceedings of the ACM on Human-Computer Interaction*, 6:1 – 25, 2022. doi:10.1145/3492836.

[56] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human–Computer Interaction*, 19:273 – 309, 2004. doi:10.1207/s15327051hci1903_3.

[57] Catherine Garbay, Fabien Badeig, and Jean Caelen. Normative multi-agent approach to support collaborative work in distributed tangible environments. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, pages 83–86, Seattle Washington USA, February 2012. ACM. ISBN 978-1-4503-1051-2. doi:10.1145/2141512.2141548.

[58] Inma García-Pereira, Cristina Portalés, Jesús Gimeno, and Sergio Casas. A collaborative augmented reality annotation tool for the inspection of prefabricated buildings. *Multimedia Tools and Applications*, 79(9-10):6483–6501, March 2020. ISSN 1380-7501, 1573-7721. doi:10.1007/s11042-019-08419-x.

[59] Inma García-Pereira, Pablo Casanova-Salas, Jesús Gimeno, Pedro Morillo, and Dirk Reiners. Cross-device augmented reality annotations method for asynchronous collaboration in unprepared environments. 2021. doi:10.3390/info12120519.

[60] S. Garner and Phebe Mann. Interdisciplinarity: perceptions of the value of computer-supported collaborative work in design for the built environment. *Automation in Construction*, 12:495–499, 2003. doi:10.1016/S0926-5805(03)00035-9.

[61] Danilo Gasques, Janet G Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin, and Xu Enrique Zavala. Artemis: A collaborative mixed-reality system for immersive surgical telementoring. 2021. doi:10.1145/3411764.3445576.

[62] William W. Gaver. Auditory icons: Using sound in computer interfaces. *Human–Computer Interaction*, 2(2):167–177, 1986. doi:10.1207/s15327051hci0202_3.

[63] Alexander Giovannelli, Leonardo Pavanatto, Shakiba Davari, Haichao Miao, Vuthea Chheang, Brian Giera, Timo Bremer, and Doug A. Bowman. Investigating the influence of playback interactivity during guided tours for asynchronous collaboration in virtual reality - 2025 ieee conference virtual reality and 3d user interfaces (vr). pages 23–33, 3 2025. ISSN 2642-5254. doi:10.1109/VR59515.2025.00027.

[64] Mar Gonzalez-Franco and Tabitha C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018. doi:10.3389/frobt.2018.00074.

[65] C. Greenhalgh, M. Flintham, J. Purbrick, and S. Benford. Applications of temporal links: recording and replaying virtual environments. In *Proceedings IEEE Virtual Reality 2002*, pages 101–108, 2002. doi:10.1109/VR.2002.996512.

[66] Anhong Guo, Ilter Canberk, Hannah Murphy, Andrés Monroy-Hernández, and Rajan Vaish. Blocks: Collaborative and Persistent Augmented Reality Experiences. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–24, September 2019. ISSN 2474-9567. doi:10.1145/3351241.

[67] Carl Gutwin and Saul Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):411–446, 2002. doi:10.1023/A:1021271517844.

[68] Ramy Hammady, Minhua Ma, and Nicholas Temple. Augmented Reality and Gamification in Heritage Museums. In Tim Marsh, Minhua Ma,

Manuel Fradinho Oliveira, Jannicke Baalsrud Hauge, and Stefan Göbel, editors, *Serious Games*, volume 9894, pages 181–187. Springer International Publishing, Cham, 2016. ISBN 978-3-319-45840-3 978-3-319-45841-0. doi:10.1007/978-3-319-45841-0_17.

[69] Janeen Hammond, Richard J. Koubek, and Craig M. Harvey. Distributed collaboration for engineering design: A review and reappraisal. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 11(1):35–52, 2001. doi:10.1002/1520-6564(200124)11:1<35::AID-HFM3>3.0.CO;2-E.

[70] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988. doi:https://doi.org/10.1016/S0166-4115(08)62386-9.

[71] Rachel S. Herz. The role of odor-evoked memory in psychological and physiological health. *Brain Sciences*, 6(3):22, 2016. doi:10.3390/brainsci6030022.

[72] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in information systems research. *MIS Q.*, 28(1):75–105, March 2004. ISSN 0276-7783. `https://www.uni-kassel.de/fb07/index.php?eID=dumpFile&t=f&f=4899&token=fda52302d42e5e5c05d7f89c2d578f80fedf7b28`.

[73] Weidong Huang, Leila Alem, Franco Tecchia, and Henry Been Lirn Duh. Augmented 3d hands: a gesture-based mixed reality system for distributed collaboration. *Journal on Multimodal User Interfaces*, 12:77–89, 6 2018. ISSN 17838738. doi:10.1007/s12193-017-0250-2.

[74] Y. Huang, W. Han, and D. Macbeth. The complexity of collaboration in supply chain networks. *Supply Chain Management: An International Journal*, 2019. doi:10.1108/SCM-11-2018-0382.

[75] Tomoko Imai, Andrew Johnson, Jason Leigh, Dave Pape, and T.A. DeFanti. Supporting transoceanic collaborations in virtual environment. pages 1059 – 1062 vol.2, 1999. ISBN 7-5635-0402-8. doi:10.1109/APCC.1999.820446.

[76] Andrew Irlitti, Stewart Von Itzstein, Leila Alem, and Bruce Thomas. Tangible interaction techniques to support asynchronous collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–6, Adelaide, Australia, October 2013. IEEE. ISBN 978-1-4799-2869-9. doi:10.1109/ISMAR.2013.6671840.

[77] Andrew Irlitti, Ross T. Smith, Stewart Von Itzstein, Mark Billinghurst, and Bruce H. Thomas. Challenges for Asynchronous Collaboration in Augmented Reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 31–35, Merida, September 2016. IEEE. ISBN 978-1-5090-3740-7. doi:10.1109/ISMAR-Adjunct.2016.0032.

[78] ISO. Usability: Definitions and concepts. Standard, International Organization for Standardization, 2018. `https://cdn.standards.iteh.ai/samples/16883/44acafdfd9a24edd9c66ed2f0e2a50e2/ISO-9241-11-1998.pdf`.

[79] S. Jarvenpaa and L. Välikangas. Toward temporally complex collaboration in an interorganizational research network. *Strategic Organization*, 20:110 – 134, 2021. doi:10.1177/14761270211042934.

[80] Jérŏme Jetter, Jörgen Eimecke, and Alexandra Rese. Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits? *Computers in Human Behavior*, 87:18–33, 2018. ISSN 0747-5632. doi:https://doi.org/10.1016/j.chb.2018.04.054.

[81] Allison Jing, Theophilus Teo, Jeremy Mcdade, Chenkai Zhang, Yi Wang, Andrei Mitrofan, Rushil Thareja, Heesook Shin, Yongho Lee, Youn-Hee Gil, Mark Billinghurst, and Gun A. Lee. Superpowering emotion through multimodal cues in collaborative vr. 2024. doi:10.1109/ISMAR62088.2024.00030.

[82] Robert Johansen. *GroupWare: Computer Support for Business Teams*. The Free Press, USA, 1988. ISBN 0029164915. `https://dl.acm.org/doi/abs/10.5555/542298`.

[83] Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 957–966, New York, NY, USA, April 2015. Association for Computing Machinery. ISBN 978-1-4503-3145-6. doi:10.1145/2702123.2702345.

[84] Shunichi Kasahara, Valentin Heun, Austin S. Lee, and Hiroshi Ishii. Second surface: multi-user spatial collaboration system based on augmented reality. In *SIGGRAPH Asia 2012 Emerging Technologies*, pages 1–4, Singapore Singapore, November 2012. ACM. ISBN 978-1-4503-1912-6. doi:10.1145/2407707.2407727.

[85] Konstantina Kilteni, Raphaela Groten, and Mel Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012. doi:10.1162/PRES_a_00124.

[86] Seungwon Kim, Mark Billinghurst, and Gun Lee. The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration. *Computer Supported Cooperative Work (CSCW)*, 27(3):569–607, December 2018. ISSN 1573-7551. doi:10.1007/s10606-018-9324-2.

[87] Paul A Kirschner, John Sweller, and Richard E Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2):75–86, 2006. doi:10.1207/s15326985ep4102_1.

[88] Roberta L. Klatzky and Susan J. Lederman. Haptic object perception: Foundations and progress. *Philosophical Transactions of the Royal Society B*, 366(1581):3097–3105, 2011. doi:10.1098/rstb.2011.0153.

[89] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Proceedings of the 2007 6th IEEE and ACM International*

Symposium on Mixed and Augmented Reality (ISMAR), pages 225–234, 2007. doi:10.1109/ISMAR.2007.4538852.

[90] Gerd Kortuem, Martin Bauer, and Zary Segall. [No title found]. *Mobile Networks and Applications*, 4(1):49–58, 1999. ISSN 1383469X. doi:10.1023/A:1019122125996.

[91] Falko Kübler. Immersive versionsverwaltung von kollaborativer arbeit in virtual reality, 2022. `https://publikationen.bibliothek.kit.edu/1000182579`.

[92] Thomas Kvan. Collaborative design: What is it? *Automation in Construction*, 9:409–415, 07 2000. doi:10.1016/S0926-5805(99)00025-4.

[93] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In Andreas Holzinger, editor, *HCI and Usability for Education and Work*, pages 63–76, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. doi:10.1007/978-3-540-89350-9_6.

[94] Steven M. LaValle. *Virtual Reality*. Cambridge University Press, Cambridge, 2017. ISBN 9781107198937. `https://msl.cs.uiuc.edu/vr/vrbookbig.pdf`.

[95] Jan Marco Leimeister. *Collaboration Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. ISBN 978-3-642-20890-4. doi:10.1007/978-3-642-20891-1. Publication Title: Collaboration Engineering.

[96] Maximilian Letter and Katrin Wolf. Tangible version control: Exploring a physical object's alternative versions. 2022. doi:10.1145/3491101.3519686.

[97] Maximilian Letter, Ceenu George, and Katrin Wolf. A survey of computer-supported remote collaboration on physical objects. In *Human-Computer Interaction – INTERACT 2023: 19th IFIP TC13 International Conference, York, UK, August 28 – September 1, 2023, Proceedings, Part III*, page 346–368, Berlin, Heidelberg, 2023. Springer-Verlag. ISBN 978-3-031-42285-0. doi:10.1007/978-3-031-42286-7_19.

[98] James R. Lewis. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq. *SIGCHI Bull.*, 23(1):78–81, January 1991. ISSN 0736-6906. doi:10.1145/122672.122692.

[99] David Lindlbauer and Andy Wilson. *Remixed Reality: Manipulating Space and Time in Augmented Reality*, pages 1–13. 04 2018. doi:10.1145/3173574.3173703.

[100] Peng Liu and Zhizhong Li. Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, 42: 553–568, 2012. doi:10.1016/J.ERGON.2012.09.001.

[101] Thomas Ludwig, Oliver Stickel, Peter Tolmie, and Malte Sellmer. shARe-IT: Ad hoc Remote Troubleshooting through Augmented Reality. *Computer Supported Cooperative Work (CSCW)*, 30(1):119–167, February 2021. ISSN 1573-7551. doi:10.1007/s10606-021-09393-5.

[102] Andreea Gabriela Lupascu, Aurelia Ciupe, Serban Meza, and Bogdan Orza. ARThings – enhancing the visitors' experience in museums through collaborative AR. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 669–670, Lisbon, Portugal, March 2021. IEEE. ISBN 978-1-6654-4057-8. doi:10.1109/VRW52623.2021.00217.

[103] Abdullah Maghazil and Peter Yellowlees. Novel Approaches to Clinical Care in Mental Health: From Asynchronous Telepsychiatry to Virtual Reality. In Margaret Lech, Insu Song, Peter Yellowlees, and Joachim Diederich, editors, *Mental Health Informatics*, volume 491, pages 57–78. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. ISBN 978-3-642-38549-0 978-3-642-38550-6. doi:10.1007/978-3-642-38550-6_4.

[104] Bernardo Marques, Samuel Silva, Antonio Rocha, Paulo Dias, and Beatriz Sousa Santos. Remote asynchronous collaboration in maintenance scenarios using augmented reality and annotations. 2021. doi:10.1109/VRW52623.2021.00166.

[105] Bernardo Marques, Samuel Silva, António Rocha, Paulo Dias, and Beatriz Sousa Santos. Remote asynchronous collaboration in maintenance scenarios using augmented reality and annotations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 567–568, 2021. doi:10.1109/VRW52623.2021.00166.

[106] Bernardo Marques, Samuel Silva, Paulo Dias, and Beatriz Sousa-Santos. Which notification is better- comparing visual, audio and tactile cues for asynchronous mixed reality (mr) remote collaboration: A user study. 2022. doi:10.1145/3568444.3570587.

[107] Bernardo Marques, Carlos Ferreira, Samuel Silva, Paulo Dias, and Beatriz Sousa Santos. How to notify team members during asynchronous remote collaboration supported by mixed reality: Comparing visual, audio and tactile notifications. 2023. doi:10.1145/3565066.3608687.

[108] Tariq Masood and Johannes Egger. Augmented reality in support of industry 4.0—implementation challenges and success factors. *Robotics and Computer-Integrated Manufacturing*, 58:181–195, 2019. ISSN 0736-5845. doi:https://doi.org/10.1016/j.rcim.2019.02.003.

[109] A. Massey, Mitzi M. Montoya, Binny M. Samuel, and J. Windeler. Presence and team performance in synchronous collaborative virtual environments. *Small Group Research*, 55:290 – 329, 2023. doi:10.1177/10464964231185748.

[110] Gregory McGowin, Nathan A. Sonnenfeld, Stephanie G. Fussell, and Sarah Ahmed. Which reality are we in? deconstructing xr terms and taxonomies—survey of who cares and why it matters. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 0(0): 10711813251363189, 0. doi:10.1177/10711813251363189.

[111] Eduardo Medina, Anatoly Compeau, Mark K. Renault, David Clavel, and Jess Hart. Virtusphere: Walking in virtual environments and the perception of motion. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 52, pages 2102–2106, 2008. doi:10.1177/154193120805202704.

[112] Ines Miguel-Alonso, David Checa, Henar Guillen-Sanz, and Andres Bustillo. Evaluation of the novelty effect in immersive virtual reality learning experiences. *Virtual Real.*, 28(1), January 2024. ISSN 1359-4338. doi:10.1007/s10055-023-00926-5.

[113] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: a class of displays on the reality-virtuality continuum. In Hari Das, editor, *Telemanipulator and Telepresence Technologies*, volume 2351, pages 282 – 292. International Society for Optics and Photonics, SPIE, 1995. doi:10.1117/12.197321.

[114] Simone Mora, Alessandro Boron, and Monica Divitini. CroMAR: Mobile Augmented Reality for Supporting Reflection on Crowd Management. *International Journal of Mobile Human Computer Interaction*, 4(2):88–101, April 2012. ISSN 1942-390X, 1942-3918. doi:10.4018/jmhci.2012040107.

[115] Mikhail Morozov, Alexey Gerasimov, Mikhail Fominykh, and Andrey Smorkalov. Asynchronous immersive classes in a 3d virtual world: Extended description of vacademia. In Marina L. Gavrilova, C. J. Kenneth Tan, and Arjan Kuijper, editors, *Transactions on Computational Science XVIII*, pages 81–100. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-38803-3. doi:10.1007/978-3-642-38803-3_5.

[116] Huyen Nguyen, Thomas Bruhn, Christian Sandor, and Patrick Bourdot. Absence agents: Mitigating interruptions in extended reality remote collaboration. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 554–555, 2022. doi:10.1109/VRW55335.2022.00128.

[117] Nelusha Nugegoda, Marium-E Jannat, Khalad Hasan, and Patricia Lasserre. Exploring the effect of viewing attributes of mobile ar interfaces on remote collaborative and competitive tasks. 2024. doi:10.1109/TVCG.2024.3456177.

[118] Juha Ojala, Jukka Selin, Timo Partala, and Markku Rossi. Virtual construction: Interactive tools for collaboration in virtual reality. 2020. doi:10.1007/978-3-030-39442-4_26.

[119] Gary M. Olson and Judith S. Olson. Distance matters. *Hum.-Comput. Interact.*, 15(2):139–178, September 2000. ISSN 0737-0024. doi:10.1207/S15327051HCI1523_4.

[120] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, page 741–754, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341899. doi:10.1145/2984511.2984517.

[121] Jivka Ovtcharova, Polina Häfner, Victor Häfner, Jurica Katicic, and Christina Vinke. *Innovation braucht Resourceful Humans Aufbruch in eine neue Arbeitskultur durch Virtual Engineering*, pages 111–124. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. doi:10.1007/978-3-662-45915-7_12.

[122] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 2021. doi:10.1136/bmj.n71.

[123] Shiva Pedram, Stephen Palmisano, Richard Skarbez, Pascal Perez, and Matthew Farrelly. Investigating the process of mine rescuers' safety training with immersive virtual reality: A structural equation modelling approach. *Computers & Education*, 153:103891, 2020. ISSN 0360-1315. doi:10.1016/j.compedu.2020.103891.

[124] Jérôme Perret and Emmanuel Vander Poorten. *Touching Virtual Reality: a Review of Haptic Gloves*. June 2018. https://www.researchgate.net/publication/324562855_Touching_Virtual_Reality_a_Review_of_Haptic_Gloves.

[125] Florian Perteneder, Eva-Maria Grossauer, Yan Xu, and Michael Haller. Catch-Up 360: Digital Benefits for Physical Artifacts. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 105–108, Stanford California USA, January 2015. ACM. ISBN 978-1-4503-3305-4. doi:10.1145/2677199.2680564.

[126] Catlin Pidel and Philipp Ackermann. Collaboration in Virtual and Augmented Reality: A Systematic Overview. In Lucio Tommaso De Paolis and Patrick Bourdot, editors, *Augmented Reality, Virtual Reality, and Computer Graphics*, pages 141–156, Cham, 2020. Springer International Publishing. ISBN 978-3-030-58465-8. doi:10.1007/978-3-030-58465-8_10.

[127] Ulrich Piepenburg. Ein konzept von kooperation und die technische unterstützung kooperativer prozesse in bürobereichen. In Jürgen Friedrich and Karl-Heinz Rödiger, editors, *Computergestützte Gruppenarbeit (CSCW): Gemeinsame Fachtagung des German Chapter of the ACM mit der Gesellschaft für Informatik (GI) und der Universität Bremen, FB Mathematik/Informatik, vom 30.9. bis 2.10.1991 in Bremen*, volume 34 of *Berichte des German Chapter of the ACM*, pages 79–94. Teubner, 1991. http://dl.mensch-und-computer.de/handle/123456789/2553.

[128] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. doi:10.1145/3173574.3173620.

[129] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and

tangible interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–17, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi:10.1145/3290605.3300458.

[130] Ronald Poelman, Oytun Akman, Stephan Lukosch, and Pieter Jonker. As if being there: mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1267–1276, Seattle Washington USA, February 2012. ACM. ISBN 978-1-4503-1086-4. doi:10.1145/2145204.2145394.

[131] Jens Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3):257–266, 1983. doi:10.1109/TSMC.1983.6313160.

[132] Sharif Razzaque, Zachariah Kohn, and Mary C. Whitton. Redirected walking. In *EUROGRAPHICS 2001 – Short Presentations*, 2001. `https://diglib.eg.org/bitstreams/6bb0fac0-8555-471e-8728-209d57cfcde5/download`.

[133] Irene Reisner-Kollmann and Andrea Aschauer. Design and implementation of asynchronous remote support. *CEUR Workshop Proceedings*, 2618: 9–11, 2020. ISSN 1613-0073. `https://ceur-ws.org/Vol-2618/paper3.pdf`.

[134] Philippe Renevier and Laurence Nigay. Mobile Collaborative Augmented Reality: The Augmented Stroll. In Gerhard Goos, Juris Hartmanis, Jan Van Leeuwen, Murray Reed Little, and Laurence Nigay, editors, *Engineering for Human-Computer Interaction*, volume 2254, pages 299–316. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. ISBN 978-3-540-43044-5 978-3-540-45348-2. doi:10.1007/3-540-45348-2_25.

[135] Jen Rhymer. Location-independent organizations: Designing collaboration across space and time. *Administrative Science Quarterly*, 68:1 – 43, 2021. doi:10.1177/00018392221129175.

[136] Augustin Rungeard. A concept to provide immersive hand instructions for assembly tasks in asynchronous remote collaboration. Master's thesis, Karlsruher Institut für Technologie (KIT), 2022. `https://publikationen.bibliothek.kit.edu/1000180933`.

[137] Kjeld Schmidt and Liam Bannon. Taking cscw seriously: Supporting articulation work. *Computer Supported Cooperative Work (CSCW)*, 1(1): 7–40, 1992. doi:10.1007/BF00752449.

[138] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10(3):266–281, 06 2001. doi:10.1162/105474601300343603.

[139] Mickael Sereno, Xiyao Wang, Lonni Besancon, Michael J Mcguffin, and Tobias Isenberg. Collaborative Work in Augmented Reality: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2020. ISSN 1077-2626, 1941-0506, 2160-9306. doi:10.1109/TVCG.2020.3032761.

[140] Neal E. Seymour, Anthony G. Gallagher, Sanziana A. Roman, Michael K. O'Brien, Vipin K. Bansal, Dana K. Andersen, and Richard M. Satava. Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery*, 236(4):458–464, October 2002. ISSN 0003-4932. doi:10.1097/00000658-200210000-00008.

[141] Aashiq Shaikh, Linda Nguyen, Alireza Bahremand, Hannah Bartolomea, Frank Liu, Van Nguyen, Derrick Anderson, and Robert LiKamWa. Coordinate: A Spreadsheet-Programmable Augmented Reality Framework for Immersive Map-Based Visualizations. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 134–1343, San Diego, CA, USA, December 2019. IEEE. ISBN 978-1-7281-5604-0. doi:10.1109/AIVR46125.2019.00028.

[142] Vibhu Saujanya Sharma, Rohit Mehra, Vikrant Kaulgud, and Sanjay Podder. SImBa: An Augmented Reality approach for creating Smart Immersive

Bays in Software Delivery Environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 842–843, Atlanta, GA, USA, March 2020. IEEE. ISBN 978-1-7281-6532-5. doi:10.1109/VRW50115.2020.00272.

[143] A. Simões, Ana Sofia Brás Pinto, Joana Santos, Sofia Pinheiro, and David Romero. Designing human-robot collaboration (hrc) workspaces in industrial settings: A systemic literature review. *Journal of Manufacturing Systems*, 2022. doi:10.1016/j.jmsy.2021.11.007.

[144] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 1997. doi:10.1162/pres.1997.6.6.603.

[145] Branislav Sobota, Štefan Korečko, Marián Hudák, and Martin Sivý. Mixed reality: A known unknown. In Branislav Sobota and Dragan Cvetković, editors, *Mixed Reality and Three-Dimensional Computer Graphics*, chapter 10. IntechOpen, Rijeka, 2020. doi:10.5772/intechopen.92827.

[146] Hyungki Son, Inwook Hwang, Tae-Heon Yang, Seungmoon Choi, Sang-Youn Kim, and Jin Ryong Kim. Realwalk: Haptic shoes using actuated mr fluid for walking in vr. In *2019 IEEE World Haptics Conference (WHC)*, pages 241–246, 2019. doi:10.1109/WHC.2019.8816165.

[147] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi:10.1145/3290605.3300767.

[148] Mandayam A. Srinivasan and Cagatay Basdogan. Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics*, 21(4):393–404, 1997. ISSN 0097-8493. doi:https://doi.org/10.1016/S0097-8493(97)00030-7.

[149] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Marc Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE Transactions on Visualization and Computer Graphics*, 16 (1):17–27, 2010. doi:10.1109/TVCG.2009.62.

[150] Ibrahim A. Tahmid, Francielly Rodrigues, Alexander Giovannelli, Lee Lisle, Jerald Thomas, and Doug A. Bowman. Colt: Enhancing collaborative literature review tasks with synchronous and asynchronous awareness across the reality-virtuality continuum. pages 831–836, 2023. doi:10.1109/ISMAR-Adjunct60411.2023.00183.

[151] James Tam and Saul Greenberg. A framework for asynchronous change awareness in collaborative documents and workspaces. *International Journal of Human-Computer Studies*, 64(7):583–598, 2006. ISSN 1071-5819. doi:10.1016/j.ijhcs.2006.02.004.

[152] Yujie Tao, Jordan Egelman, and Jeremy N. Bailenson. I feel you: Impact of shared body sensations on social interactions in virtual reality. 2024. doi:10.1109/ISMAR62088.2024.00126.

[153] Franco Tecchia, Leila Alem, and Weidong Huang. 3d helping hands: A gesture based mr system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI ’12, page 323–328, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450318259. doi:10.1145/2407516.2407590.

[154] Santawat Thanyadit, Parinya Punpongsanon, Thammathip Piumsomboon, and Ting-Chuen Pong. Xr-live: Enhancing asynchronous shared-space demonstrations with spatial-temporal assistive toolsets for effective learning in immersive virtual laboratories. 2022. doi:10.1145/3512983.

[155] Ann Sofia Skou Thomsen, Daniella Bach-Holm, Hadi Kjærbo, Klavs Højgaard-Olsen, Yousif Subhi, George M. Saleh, Yoon Soo Park, Morten La Cour, and Lars Konge. Operating Room Performance Improves after Proficiency-Based Virtual Reality Cataract Surgery Training. *Ophthalmology*, 124(4):524–531, April 2017. ISSN 01616420. doi:10.1016/j.ophtha.2016.11.015.

[156] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 161–174, New Orleans LA USA, October 2019. ACM. ISBN 978-1-4503-6816-2. doi:10.1145/3332165.3347872.

[157] C. Traxler and K. Chmelina. Storytelling as collaboration support for 3d tunnel monitoring. 2023. doi:10.1201/9781003348030-307.

[158] David Uttal, Nathaniel Meadow, Elizabeth Tipton, Linda Liu Hand, Alison Alden, Christopher Warren, and Nora Newcombe. The malleability of spatial skills: A meta-analysis of training studies. *Psychological bulletin*, 139, 06 2012. doi:10.1037/a0028446.

[159] Cheng Yao Wang, Mose Sakashita, Upol Ehsan, Jingjin Li, and Andrea Stevenson Won. *Again, Together: Socially Reliving Virtual Reality Experiences When Separated*, page 1–12. CHI '20. Association for Computing Machinery, New York, NY, USA, 2020. ISBN 9781450367080. doi:10.1145/3313831.3376642.

[160] Peng Wang, Shusheng Zhang, Xiaoliang Bai, Mark Billinghurst, Weiping He, Mengmeng Sun, Yongxing Chen, Hao Lv, and Hongyu Ji. 2.5dhands: a gesture-based mr remote collaborative platform. *International Journal of Advanced Manufacturing Technology*, 102:1339–1353, 6 2019. ISSN 14333015. doi:10.1007/s00170-018-03237-1.

[161] Peng Wang, Xiaoliang Bai, Mark Billinghurst, Shusheng Zhang, Sili Wei, Guangyao Xu, Weiping He, Xiangyu Zhang, and Jie Zhang. 3dgam: using 3d gesture and cad models for training on mixed reality remote collaboration. *Multimedia Tools and Applications*, 80:31059–31084, 8 2021. ISSN 15737721. doi:10.1007/s11042-020-09731-7.

[162] Peng Wang, Xiaoliang Bai, Mark Billinghurst, Shusheng Zhang, Xiangyu Zhang, Shuxia Wang, Weiping He, Yuxiang Yan, and Hongyu Ji. Ar/mr remote collaboration on physical tasks: A review. *Robotics and Computer-Integrated Manufacturing*, 72:102071, 2021. ISSN 0736-5845. doi:https://doi.org/10.1016/j.rcim.2020.102071.

[163] Elizabeth M. Wenzel, Michael Arruda, David J. Kistler, and Frederic L. Wightman. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123, 1993. doi:10.1121/1.407089.

[164] Bob G. Witmer and Michael J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, 06 1998. doi:10.1162/105474698565686.

[165] Robert E Wood. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37(1):60–82, 1986. ISSN 0749-5978. doi:https://doi.org/10.1016/0749-5978(86)90044-0.

[166] Kieran Woodward, Eiman Kanjo, and Will Parker. DropAR: Enriching Exergaming Using Collaborative Augmented Reality Content. In Xiaowen Fang, editor, *HCI in Games*, volume 13334, pages 652–663. Springer International Publishing, Cham, 2022. ISBN 978-3-031-05636-9 978-3-031-05637-6. doi:10.1007/978-3-031-05637-6_42.

[167] Te-Yen Wu, Jun Gong, Teddy Seyed, and Xing-Dong Yang. Proxino: Enabling Prototyping of Virtual Circuits with Physical Proxies. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 121–132, New Orleans LA USA, October 2019. ACM. ISBN 978-1-4503-6816-2. doi:10.1145/3332165.3347938.

[168] You-Jun Wu, Linzhen Zhou, Pai Zheng, Yanqing Sun, and Kaikai Zhang. A digital twin-based multidisciplinary collaborative design approach for complex engineering product development. *Adv. Eng. Informatics*, 52: 101635, 2022. doi:10.1016/j.aei.2022.101635.

[169] Tae-Heon Yang, Hyungki Son, Sangkyu Byeon, Hyunjae Gil, Inwook Hwang, Gwanghyun Jo, Seungmoon Choi, Sang-Youn Kim, and Jin Ryong Kim. Magnetorheological fluid haptic shoes for walking in vr. *IEEE Transactions on Haptics*, 14(1):83–94, 2021. doi:10.1109/TOH.2020.3017099.

[170] Ungyeon Yang and Gerard Jounghyun Kim. Implementation and evaluation of "just follow me": An immersive, vr-based, motion-training system.

Presence: Teleoper. Virtual Environ., 11(3):304–323, 2002. ISSN 1054-7460. doi:10.1162/105474602317473240.

[171] Jingjing Zhang, Binyang Han, Ze Dong, Ruoyu Wen, Gun A. Lee, Simon Hoermann, Wendy Zhang, and Thammathip Piumsomboon. Virtual triplets: A mixed modal synchronous and asynchronous collaboration with human-agent interaction in virtual reality. 2024. doi:10.1145/3613905.3650766.

[172] Lei Zhang, Ashutosh Agrawal, Steve Oney, and Anhong Guo. Vrgit: A version control system for collaborative content creation in virtual reality. 2023. doi:10.1145/3544548.3581136.