DIGITAL, DYNAMIC AND RESPONSIBLE TWINS FOR XR

D5.1 – XR applications v1 and preliminary validation results

[WP5 – XR applications and validation]

Funded by the European Union under grant number 101092875. UK based consortium partner Trilateral Research Limited, UK is funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant number 10069394).

Funded by the European Union

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or UKRI. Neither the European Union nor the granting authority can be held responsible for them.

Lead Contributor	Àlex Ginés, Neàpolis
	alexgd@neapolis.cat
Other Contributors	Vera Boehner, FIWARE
	Tim Smyth, FIWARE
	Tanja Kojic, TUB
	Francesco Vona, HSHL
	Maximilian Warsinke, TUB
	Boulos El Asmar, Ideal Works
	Anthony Rizk, Ideal Works
	Etaf AL-Zarif, AUB
	Ibrahim El Hadka, AUB
	Marc Gálvez, DTT
	Adrián González, Neàpolis

Due Date	31.01.2025
Delivery Date	31.01.2025
Type	R (Report)
Dissemination level	PU = Public
Status	Final

Keywords	***Use case validation, Impact, User centric applications, Methodology***

Document History

Version	Date	Description	Reason for Change	Partner
V0.1	5.12.2024	Original draft		VNG
V0.2	23.01.2025	Draft	All contributions included	VNG, FW, IW, HSHL,TUB,
V0.9	28.01.2025	Version for internal Review		VNG
V1.0	31.01.2025	Final Version	Last corrections	JRS

About the project

The digital transformation and the availability of more diverse and cost-effective methods for 3D capture has led to the creation of digitised representations of parts of our public spaces, machinery and processes, so-called Digital Twins. These Digital Twins are an important basis for building advanced XR applications for cityscapes and the industrial environments. These XR applications need high-fidelity models for accurate localization, which are kept in-sync with the real world. The vision of DIDYMOS-XR is to enable advanced, more realistic and more dynamic extended reality (XR) applications, powered through artificial intelligence. The project thus focuses on advancing technologies for creating large-scale digital twins, synchronised with the real world. DIDYMOS-XR will research and develop methods for reconstruction and mapping from heterogeneous inputs, including static and mobile sensors, AI-based data fusion, scene understanding and rendering. Digital twin based XR applications also require means of accurately positioning in the environment, aware of the dynamics of the scene, also addressed by the project. The capture of scenes at scale, as well as using cameras and other sensor data for synchronising the digital representation, bears the risk of capturing personal and sensitive data. Hence, the technologies resulting from DIDYMOS-XR must be and will be ethical and privacy-aware by design. In order to demonstrate and validate the enabling technologies researched and developed in DIDYMOS-XR, the project will address use cases of XR applications in two domains that differ in their scale and characteristics. One is urban planning and smart mobility, the other is the collaboration with autonomous mobile robots at an industrial production facility.

Abstract

This deliverable aims to detail the first steps taken by the DIDYMOS-XR project regarding the use case applications and the subsequent validations. It describes all the process that has been done until the first application versions arrived to the first testers. It describes the validation techniques of the different applications as well as the methodology followed, qualitatively and quantitatively, in order to analyze the results of the same. The goal is to have an accurate overview of all the feedback gathered from the different users and stakeholders in relation to the different use cases activities. This aims to be a sum up of all the actions carried out in terms of being able to validate and analyze the results. These conclusions will serve to be able to trace the future lines of work to be able to bring these technologies closer to their end users.

Table of Contents

TABLE OF CONTENTS 4

1 INTRODUCTION 14

1 USE CASE GENERAL PRESENTATION 14

1.1 USER ENGAGEMENT 15

1.1.1 Stakeholder mapping 17

1.2 CITY USE CASES 18

1.3 USE CASE SPECIFIC FOR IDEAL WORKS 23

2. USE CASE VALIDATION 24

2.1 USE CASE VALIDATION IN ETTELN 24

2.1.1 Digital Twin Creation 26

Introduction 26

Subject Groups 27

Experiment Design 28

Procedure 29

Data Collection and Analysis 29

Results 31

Conclusion 32

2.1.2 City Planning 33

Introduction 33

Experimental Setup 33

Subject Groups 34

Experiment Design 34

Procedure 35

Data Collection and Analysis 35

Results 36
Conclusion 39
2.1.3 VR Virtual Tour 39
Introduction 39
Experimental Setup 39
Subject Groups 40
Experiment Design 40
Procedure 41
Task Execution 41
Data Collection and Analysis 42
Results 44
Conclusion 45
2.1.4 AR Virtual Tour 46
Introduction 46
Experimental Setup 46
Subject Groups 46
Experiment Design 47
Procedure 47
Data Collection and Analysis 48
Results 49
Conclusion 51
2.2 USE CASE VALIDATION IN VILANOVA 51
2.2.1 AR Virtual Tour 54
Introduction 54
Experimental Setup 54
Subject Groups: 54

Experiment Design 55
Procedure 56
Data Collection and Analysis 56
Results 57
Conclusion 58
2.3 USE CASE VALIDATION AT IDEAL WORKS 58
Experiment Goals 59
Experimental Setup 59
Experiment Design 60
Procedure 62
Training Phase: 62
TASK EXECUTION ORDER: 62
TASK EXECUTION: 62
Data Collection and Analysis 63
Data Collection from Participants' Input 63
Participant Details and Interface Selection 63
Task Execution and POI Input 63
Preparation for Transformation Between Planes Calculation 64
Transformation Calculations 64
Error Calculation 67
Euclidean Distance Error 67
Orientation Error (Theta) Calculation 67
Statistical Measures 68
Percentage Error Calculation 68
Time Tracking 68
NASA-TLX Survey 68

Six Dimensions of NASA-TLX 68
Statistics to be Calculated 68
3. RESULTS 69
3.1 DEMOGRAPHICS OF THE STUDY 69
3.2 ERROR CALCULATION 70
AnyFleet Distance Error 70
AnyFleet Orientation Error 71
HoloLens Distance Error 72
HoloLens Orientation Error 73
Percentage Error in AnyFleet and HoloLens 74
Time Analysis 74
NASA TLX 75
4. CONCLUSION 76
REFERENCES 77

List of figures

Figure 1: Stakeholder Map

Figure 2: Data from Bettair sensors placed at Neàpolis, Vilanova i la Geltrú.

Figure 3: Etteln: AR Virtual Tourism

Figure 4: Etteln: VR Virtual Tourism

Figure 5: Etteln: Digital Twin Creation

Figure 6: Etteln: City Planning

Figure 7: All three sample groups with ATI scores and age (N=27)

Figure 8: UEQ-S Benchmark for the Vilanova Remote Group (N=16) in Digital Twin Creation

Figure 9: Correlations between ATI and UEQ-S (Overall, Hedonic, Pragmatic) for aggregated sample group (N=27)

Figure 10: UEQ-S Benchmark for the Vilanova Remote Group (N=16) in City Planning

Figure 11: Correlations between ATI and significant UEQ-S values (Overall, Pragmatic)for aggregated sample group (N=27)

Figure 12: Correlations between ATI and the NASA-TLX score for the aggregated sample group (N=27)

Figure 13: ATI scores and age of the sample group (N=16)

Figure 14: UEQ-S Benchmark for the Etteln inhabitants Group (N=16) in VR Virtual Tour

Figure 15: Correlation between prior XR Experience and Cybersickness (N=16)

Figure 16: Correlation between IPQ dimensions (Realness, Involvement) and Cybersickness (N=16)

Figure 17: UEQ-S Benchmark for the Etteln inhabitants (N=16) in AR Virtual Tour

Figure 18: UEQ-S Graph AR Virtual tourism Etteln

Figure 19: Validation activity carried out at the fair

Figure 20: User tester doing a validation

Figure 21: 3D render of Vilanova i la Geltrú. Asset needed for constructing the UC_02C_DTC and UC03C_VRVT above all

Figure 22: Development of city maintenance use case

Figure 23: ATI scores and age for both subject groups (N=68)

Figure 24: UEQ-S Benchmark for the Vilanova Fair Group (N=50) in AR Virtual Tour

Figure 25: Correlation between UEQ-S dimensions (Overall, Hedonic) and ATI in the aggregated sample (N=68)

Figure 1 Vicon Layout

Figure 27: Warehouse with the Vicon System Setup

Figure 3 AnyFleet Interface

Figure 4 HoloLens

Figure 5: Vicon Tracking System

Figure 6 Dolly Near Distinct Objects for Easy Alignment

Figura 7: Dolly by Crates and Shelves for Moderate Alignment

Figure 33: Dolly In Open Space With No Clear References

Figura 34: Experimental Setup

Figura 35: Data Collection App

Figura 8 Data Collection System interface for entering task-level details, completion time, and POI placement coordinates

Figura 9 Coordinates for Transformation Calculation

Figura 38: AnyFleet Predefined POIs

Figura 10 Gender Distribution

Figura 11 Age Range Distribution

Figura 12 Previous Experience Percentage Comparison Between AnyFleet and HoloLens

Figura 42: AnyFleet Distance Error in Level 1(meters)

Figura 13 AnyFleet Distance Error in Level 2(meters)

Figura 14 AnyFleet Distance Error in Level 3(meters)

Figura 15 AnyFleet Orientation Error in Level

Figura 16 AnyFleet Orientation Error in Level 2

Figura 47 AnyFleet Orientation Error in Level 3

Figura 48: HoloLens Distance Error in Level 1 (meters)

Figura 17 HoloLens Distance Error in Level 2 (meters)

Figura 18 HoloLens Distance Error in Level 3 (meters)

Figura 19 HoloLens Orientation Error in Level 1

Figura 20 HoloLens Orientation Error in Level 2

Figura 53: HoloLens Orientation Error in Level 3

Figura 21 Percentage Error Comparison in Position and Orientation for AnyFleet and HoloLens

Figura 55 Average Time Comparison Between AnyFleet and HoloLens for Each Level (seconds)

Figura 56 Mean Scores of NASA TLX Dimensions Across Hololens and Anyfleet

Figura 22Standard Deviations of NASA TLX dimensions across HoloLens and Anyfleet.

List of tables

Table 1 List of acronyms/abbreviations.....11

Table 2 Glossary of terms .....**Fehler! Textmarke nicht definiert.**

Table 3 Sum up of all UC's responsibilities.....15

Table 4 Uses Cases.....16

List of acronyms/abbreviations

Abbreviation	Explanation
WP	Work Package
UC	Use Case
XR	Extended Reality
VR	Virtual Reality
AR	Augmented Reality
QHC	Quadruple Helix Collaboration
R&D	Research and Development
LoRaWAN	Long Range Wide Area Network
IoT	Internet of Things
e-car	Electric Car
LiDAR	Light Detection And Ranging
AI	Artificial Intelligence
VNG	Vilanova i la Geltrú
ICT	Information and Communications Technology
POI	Points of Interest
CERTH	Centre for Research and Technology Hellas
UAB	American University of Beirut
UPAT	University of Patras
IW	Ideal Works
JRS	Joanneum Research
DTT	Digital Twin Technology
HSHL	Hamm-Lippstadt University of Applied Sciences
TUB	Technische Universität Berlin
IEEE	Institute of Electrical and Electronics Engineers
HMD	Head-Mounted Display
API	Application Programming Interface
SUMO	Simulation of Urban Mobility

Abbreviation	Explanation
TCP	Transmission Control Protocol
SDK	Software Development Kit
BIM	Building Information Modelling
AMR	Autonomous Mobile Robots
MR	Mixed Reality
AMOLED	Active Matrix Organic Light Emitting Diode
UEQ	User Experience Questionnaire
CP	City Planning
DTC	Digital Twin Creation
CM	City Maintenance
VRVT	Virtual Reality Virtual Tourism
ARVT	Augmented Reality Virtual Tourism
UI	User Interface
ERP	Enterprise Resource Planning
WMS	Warehouse Management System

Glossary of terms

Term	Explanation
Digital Twin	A digital representation of an intended or actual real-world physical product, system, or process.
Stakeholder	A relevant actor (persons, groups, or organisations) who: (1) is affected by the project's results and findings; (2) have the potential to use or exploit the project's results and findings; or (3) have a stated interest or stake in the project results and findings.
End-user	A person, group or organisation who has the potential to use or exploit the project's results and findings. Each stakeholder is thus also an end-user, but not vice-versa.
Decision-makers	Individuals or groups who have the power and authority to make important choices that affect an organization, project, or policy.
Smart mobility	An innovative approach to transportation planning, management, and usage that leverages information and communication technologies (ICT) to improve the efficiency, sustainability, and quality of mobility for people and goods. It aims to create a more integrated, accessible, and user-centric transportation system.
Workshops	Interactive learning events, typically short-term, where a group of people gather to learn about a specific topic and develop practical skills through active participation.
Focus groups	A qualitative research method that involves gathering a small group of people to discuss a particular topic under the guidance of a moderator.
ettCAR	Is an initiative launched by the German municipality of Etteln to facilitate free, small-scale sustainable transport for local residents.
ettCARGO	The German municipality of Etteln has launched ettCARGO, an initiative promoting sustainable transportation through the use of electric cargo bikes for residents

Term	Explanation
Crossiety	Is a digital tool, specifically a village or community Etteln app, designed to enhance community engagement through communication and collaboration.
4.0 digital technologies	Refers to the current trend of automation and data exchange in manufacturing technologies. It encompasses a range of interconnected digital technologies that are transforming how products are designed, manufactured, operated, and serviced. It represents a shift towards interconnected, intelligent, and autonomous systems.
Blockchain	A blockchain is a distributed, decentralized, public digital ledger that is used to record transactions across many computers so that the record cannot be altered retroactively without the alteration of all subsequent blocks and the consensus of the network. It's a chain of blocks, where each block contains a set of transactions.
Forward-thinking	Favouring innovation and development; progressive.
Living lab	Is an open innovation ecosystem in a real-life environment.
AnyFleet	A technological solution that adapts to different types of fleets, regardless of their specific characteristics.
WebGL	Is a JavaScript API that allows you to create interactive 2D and 3D graphics within a web browser without the need for plugins. It leverages the power of your computer's graphics card (GPU) to render graphics smoothly and efficiently.
WebGPU	Is a new web API that exposes modern GPU capabilities for graphics and computation. It's designed to be more efficient and powerful than WebGL, offering better performance and access to advanced features.
PoTree	Is a data structure used to efficiently store and query point cloud data, often used in 3D graphics and spatial data analysis. It organizes points hierarchically, enabling fast neighbor searches and other spatial operations.
Microsoft Hololens	is a mixed reality (MR) headset developed by Microsoft. It allows users to interact with digital content overlaid onto the real world through augmented reality (AR) technology. Unlike virtual reality (VR) headsets, which fully immerse users in a digital environment, HoloLens blends holograms with the physical world, enabling interactive experiences without blocking the user's surroundings.

Table 1 Glossary of terms

Executive summary

This deliverable presents the research conducted and applications implemented during the first 24 months of the project.The DIDYMOS-XR project aims to enable advanced and dynamic extended reality (XR) applications through the development of high-fidelity digital twins synchronized with the real world. The project focuses on improving technologies for large-scale digital twin creation, high-accuracy localization, mapping, and real-time rendering. These technologies will be demonstrated and validated in urban planning, smart mobility, and industrial production facility use cases.

This document is focused on the development and validation of XR applications. The consortium's aim is to develop and evaluate six real-world XR applications within the smart city and industrial production domains to demonstrate the effectiveness of the developed technologies.

The validation process that has been followed is based on the following pillars:

- **Stakeholder Engagement**: Active stakeholder involvement in the design of XR technologies for citizen engagement in urban planning, tourism, smart mobility, and collaborative applications with autonomous mobile robots in industrial settings.
- **Technology Validation**: Rigorous validation of these XR technologies through real-world deployments.
- **Real-time Performance Assessment**: Development of methodologies for real-time evaluation of XR experiences, leveraging a continuously updated digital twin to provide real-time feedback to the application.
- **Project Evaluation:** Application of these developed methodologies to assess the progress and effectiveness of the project's digital twin technologies specifically within the context of XR experiences.

Key Tasks and Components:

TASK 5.1 - Urban Planning & Smart Mobility: XR applications v1 and preliminary validation results: Development of Initial XR Applications for Urban and Industrial environments, along with results of exploratory evaluations.

TASK 5.2 - Collaborative Mapping in Industrial environments: XR Applications v2 & final validation results: Development of Final XR Applications for Urban and Industrial environments, together with the results of final validation.

In conclusion, this deliverable serves as a valuable tool for the upcoming developments conducted on the project, as it offers valuable insights regarding how

users feel about the applications, it has to be a guide to still developing user-centric technologies.

Introduction

This deliverable outlines the core elements of the DIDYMOS-XR project, focusing on the validation of all use cases, objectives, and users to gather and analyze feedback on the development of all applications done across diverse contexts. It establishes a clear strategy and defines the technical criteria and guidelines governing the project's execution. A key aspect of this work is analyzing the potential applications of Digital Twins and their impact on stakeholders, including decision-makers, end-users, and professional teams. Crucially, this deliverable serves as a roadmap for future development and a valuable resource for partners seeking to understand the features and advantages of the DIDYMOS-XR framework.

At its core, the DIDYMOS-XR project aims to push the boundaries of extended reality (XR) by integrating high-fidelity digital twins with real-world data to create advanced XR applications. This integration enables real-time interaction and enhances user experiences in various domains, such as urban planning, smart mobility, and industrial environment. The project is structured into multiple work packages (WP), each dedicated to developing the technological components required for dynamic and interactive digital twins. A vital component of DIDYMOS-XR, emphasized throughout this deliverable, is the validation and testing of all developed technologies within real-world scenarios. This real-world testing is crucial for ensuring the effectiveness, robustness, and practical applicability of the project's outcomes, providing valuable feedback for iterative improvement and demonstrating the true potential of integrating digital twins with XR in real-world settings.

Above all, this deliverable is organised as it follows:

- **General presentation for use cases** - Description of the main six use cases
- **Specific use cases for Etteln** - Presentation for the use cases that involves the german city
- **Specific use cases for Vilanova** - Presentation for the use cases that involves the spanish city
- **Specific use case for Ideal Works** - Description of the single use case involved in an industrial place, Ideal Works in Munich
- **Use case validation** - Brief explanation about the methodology followed to analyze all the data obtained from the validation activities
- **Validations in Etteln** - sum up of the results from validation activities in Etteln
- **Validations in Vilanova** - sum up of the results from validation activities in Vilanova
- **Validations in Ideal Works** - sum up of the results from validation activities in Ideal Works

1 Use case general presentation

The DIDYMOS-XR project's use cases are divided into two primary areas: smart mobility and city management, and collaborative mapping in industrial environments.

The aim of this deliverable is to validate the technologies developed by Work Packages 3 and 4, which includes these partners: FICO, I2CAT, CERTH, AUB, UPAT, IW, JRS, DTT and Capgemini. These use cases are being tested in three distinct

locations: Etteln, a town in Germany; Vilanova i la Geltrú, a city in Spain; and Industrial Works, an R&D factory located in Munich, Germany.

Each of them has been chosen for their particular attributes, which allow the technologies under development within the project, to be validated in three different scenarios. Just for instance, Etteln is a town in a german rural area with less than 2,000 inhabitants, Vilanova is touristic city in catalan shore with more than 70,000 residents, meanwhile Ideal Works is an R+D german factory, the right place to test the industrial use case.

Across these locations, six specific use cases are being implemented to thoroughly test the developed technologies.

	Industry	City				
Name	Industrial DT Maintenance	Digital Twin Creation	AR Virtual Tourism	VR Virtual Tourism	City Maintenance	City Planning
Name (short)	UC01I_IDTM	UC02C_DTC	UC03C_ARVT	UC03C_VRVT	UC04C_CM	UC05C_CP
Owner	IW	VNG/FIWARE	VNG/FIWARE	VNG/FF	VNG	FIWARE
Tech Lead	AUB	DTT	UPAT	DTT	JRS	DTT

Table 3 overview on use cases and responsibilities

The use case descriptions were created by the WP2 leads, HSHL and TUB, who are also responsible for analysing and extracting data from all validation activities. All the information related to validation methodology will be described at point 2 of this document.

1.1 User engagement

In order to test all the applications made, we needed enough user testers that allowed validation methodology leaders to extract enough data to draw accurate conclusions.

Different user profiles were selected according to the end users of the applications to ensure user testing in an environment as realistic as possible. Consequently, four user typologies were identified: general citizens and society, specialized workers in the tourism sector, specialized workers in various fields, and professionals associated with the industrial sector.

Another key factor to take into account during validations is the nature of the developed application. There are some cases where validations must be tested in the same environment in which the final application will be used, while in other use cases the tests have been able to be done remotely. To sum up, these are the end users for each use case development:

Uses Cases	Industrial DT Maintenance	Digital Twin Creation	AR Virtual Tourism	VR Virtual Tourism	City Maintenance	City Planning
Code	UC01I_IDTM	UC02C_DTC	UC03C_ARVT	UC03C_VRVT	UC04C_CM	UC05C_CP
Validation location	Industrial Works, Munich	Remote	Etteln, Vilanova	Etteln	Vilanova	Etteln
End user	Ideal Works factory workers	No end users	VNG/FIWARE	VNG/FIWARE	VNG/FIWARE	VNG/FIWARE

Table 4 Uses Cases

Two different approaches to user tester engagement were implemented. Meanwhile, Ideal Works primarily conducted testing with its employees, for use cases related to city management and smart mobility, Etteln and Vilanova adopted a similar strategy to engage participants. Their approach focused on attracting individuals from city councils, local inhabitants - mostly young and elderly people, tourist agencies, tourist and cultural institutions, and local ecosystems.

Before carrying the validations there was an effort to attract as many different users as possible, following the user personas defined at WP2. As a result of this search, DIDYMOS obtained all types of profiles, young people and adults, people of different sexes, as well as socioeconomic profiles of all types, creating the richest universe possible.

The Quadruple Helix as a user identifier

The aim of the project is to get in touch with as many targets as possible. This is the main reason to apply the Quadruple Helix Collaboration (QHC) model to look for a holistic approach.

The QHC model emphasizes collaboration between four key stakeholders:

- **Public Authorities:** Provide regulatory frameworks, funding, and policy guidance.
- **Industry:** Offers expertise in production, market analysis, and commercialization.
- **Academia:** Contributes scientific knowledge, research methodologies, and educational resources.
- **Citizens (End-Users):** Represent the target audience and provide valuable insights into user needs and societal impact.

By incorporating these diverse perspectives into the validation process, we can achieve a more comprehensive and robust assessment of project value.

The QHC framework are integrated into project validation through a structured process involving the following stages:

- **Stakeholder Mapping and Engagement:** Identifying key stakeholders within each helix and establishing effective communication channels. This includes workshops, surveys, focus groups, and online platforms.
- **Needs Assessment:** Gathering input from citizens and other stakeholders to understand the real-world needs and challenges that the project aims to address.

This iterative process ensures that the project is continuously validated against the diverse perspectives of the QHC stakeholders and offers several significant advantages:

- **Enhanced Relevance and Impact:** By incorporating citizen input, the project is more likely to address real-world needs and have a positive societal impact.
- **Reduced Risk and Uncertainty:** Comprehensive validation minimizes the risk of project failure due to unforeseen technical, market, or societal challenges.
- **Strengthened Public Trust:** Demonstrating a commitment to public engagement and societal values enhances public trust and acceptance of projects.

1.1.1 Stakeholder mapping

This stakeholder map identifies four key categories representing different environments that could impact the project's progress and outcomes. All the information gathered through this map is crucial for maximizing the DIDYMOS-XR project's value and identifying target organizations for further actions. Specifically, they support coordinated partner efforts to deliver timely and meaningful results by understanding how human factors are involved regarding different use cases.

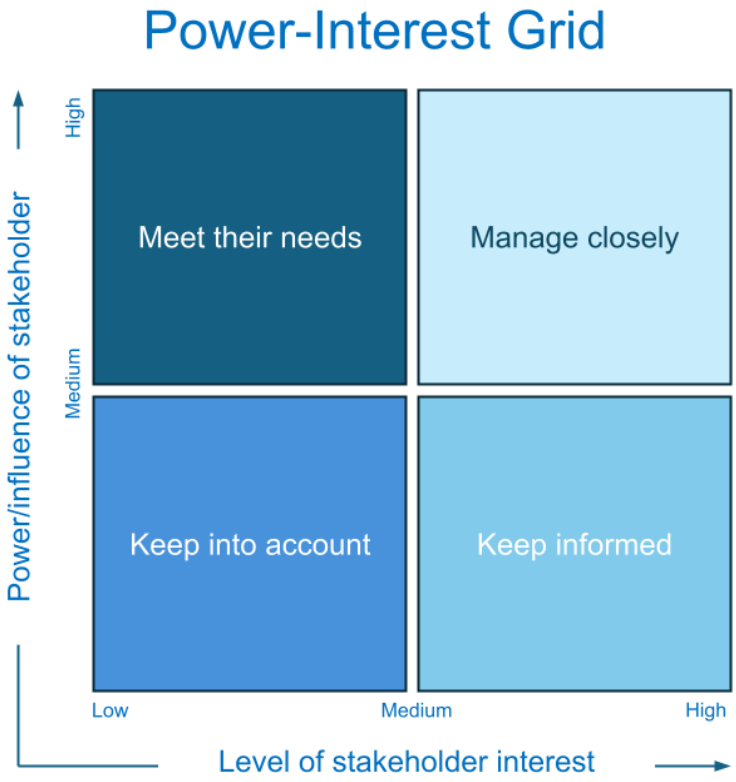

Figure 1 – Stakeholder Map

This kind of map if based on a power-interest grid, that is reflected through (Y) and (X) axis:

- Axis (Y): Power or influence of the stakeholder. This refers to their ability to influence the project's outcome. It can be derived from their position, authority, resources, or relationships.
- Axis (X): Level of interest: This refers to the stakeholder's level of concern or involvement in the project. It can be driven by their potential impact, personal stake, or vested interests.

Despite being in the early stages of use case validation, it is essential to identify stakeholders and determine their contributions to the project. The statements in each box provide recommendations for engaging with each stakeholder group.

1.2 City use cases

As urban and rural landscapes evolve in the digital age, communities worldwide are embracing innovative strategies to address the challenges of modernization, sustainability, and inclusivity. Two standout examples of this transformation are Etteln, a small German village pioneering rural digitalization, and Vilanova i la Geltrú (VNG), a Catalonian city leading advancements in smart city technology.

Etteln showcases the power of community-driven change, leveraging cutting-edge digital infrastructure, renewable energy, and mobility solutions to redefine rural living. Meanwhile, VNG exemplifies the integration of advanced technologies such as XR, digital twins, and AI to create a thriving ecosystem for innovation and entrepreneurship. Both communities, through the DIDYMOS-XR project, illustrate the potential of blending technology with human-centered design to shape sustainable, connected futures.

Explore how Etteln and VNG are paving the way for a smarter, more inclusive tomorrow. In this section there is a sum up of all use cases where both municipalities are involved, in the following order: City Planning, Digital Twin Creation, City Maintenance, Virtual Reality, Augmented Reality.

UC05C_CP, the City Planning use case is technically developed by DTT exclusively at Etteln. The project utilizes Virtual Reality (VR) technology accessible through both desktop applications and tethered head-mounted displays (HMDs) to provide users with tools for urban planning and simulation.

City Planning use case aims to empower users to interact with a digital twin of a city, enabling them to simulate various scenarios and assess their impact on urban development. Users can navigate the virtual environment, simulate traffic flow, and add or remove buildings and roads. This dual access method (desktop and HMD) allows for both detailed design work on desktop systems and immersive validation of changes within VR.

Users access the application via desktop computers or headsets. With HMD users having an adequate physical space for movement and interaction within the VR environment.Users can access from any location, making adjustments to the city model as needed, having the chance to create different simulation scenarios or modifying all the different buildings.. These changes can then be experienced and validated in VR using an HMD, providing a more immersive perspective.

The application offers several core functionalities:

- **Destination Earth API:** Integration for Weather Simulation: Weather data is fetched from the Destination Earth API, with a day and year selector for retrieving specific weather information.
- **Refined Rain Simulation:** Rain effects are enhanced for realism using Unity's Particle System, with rain intensity and occurrence linked to API data.
- **Refined Flood Simulation:** Realistic flooding is modelled using fluid simulation assets (e.g., Fluid Frenzy), incorporating soil data and Destination Earth API data for accurate visualization.
- **Refined Light Intensity Simulation:** Light intensity is dynamically adjusted based on the light source's inclination.

- **Traffic Simulation:** SUMO is used to create and connect traffic systems via TCP, visualizing traffic patterns within Unity.
- **Smart City Lights:** A configurable system allows for adjustments to smart city light positions, timing and synchronization.
- All this is possible thanks to the combination of different technologies like Unity, SUMO, Carla, Fluid Frenzi or Destination Earth API.

UC05C_CP provides a powerful platform for urban planning and simulation, empowering users to design and evaluate urban development strategies in a realistic and immersive environment. The planned future enhancements will further extend the application's capabilities and user experience with some new features such as:

- HoloMIT-SDK integration for real-time holoportation.
- AI agent integration for dynamic NPC interactions.
- Advanced analytics for user behaviour in VR.
- Cross-platform support for other XR platforms.

One of the very first use cases that is shared by FIWARE, at Etteln and Vilanova is UC_02C_DTC, the Digital Twin Creation. This use case is a collaborative effort between the two villages with technical leadership provided by DTT. The project utilizes Virtual Reality and web-based technologies accessible via desktop computers to create, manage, and maintain accurate and up-to-date digital twins of urban environments.

UC_02C_DTC serves as the foundational application for other related projects, providing the necessary digital twin environments. The system allows users to create new digital twins, update existing ones, add, delete and edit 3D assets, import sensor data, and check the overall quality of the digital twin models. This ensures that all downstream applications utilize the most current and accurate representation of the physical environment.

Digital twins created within this application are usable for other related applications. Therefore this application is the central starting point for all projects requiring digital twin data.

One of the main points is to get data from various sensors (equipped cars, drones, and city sensors) that enables the reconstruction and continuous updating of accurate digital twins. After logging in, users can either create a new digital twin or update an existing one. Creating a new digital twin involves selecting and confirming the data sources (sensors) to be used. Updating an existing digital twin allows for importing new sensor data or directly editing the 3D model through the integrated editor. This use case is being developed in two main locations: Vilanova i la Geltrú (Spain) and Etteln (Germany), with unique digital twin creation and management processes tailored to each city.

Regarding Digital Twin Creation, there are some interesting technological developments, such as:

- The implementation leverages Unity (2022.3.14f1), the XR Interaction Toolkit, NeRF/Gaussian Splatting for texture and mesh generation, WebGL for web-based access and Potree for point cloud rendering. The workflow encompasses several key processes:

- **Point-Cloud Loading:** A forked Unity plugin dynamically loads point clouds in Potree format using coroutines for WebGL compatibility. Future work will explore WebGPU for improved performance.
- **Terrain Improvement:** Unity terrain is replaced with 3D models featuring satellite imagery, enhanced with realistic vegetation, trees with wind shaders and advanced water shaders.
- **BIM Model Integration:** BIM models from other projects are integrated into the asset library, allowing users to instantiate them within the city environment.
- **Asset Manipulation:** Users can scale, rotate, and transform instantiated assets using integrated manipulation tools.
- **Point-Cloud Selection:** An interface allows users to select point clouds from various locations (Etteln, Vilanova) and dynamically load them into the scene.
- **Digital Twin Saving:** Changes to the digital twin (added assets, transformations) are persisted on the server using a JSON-based format.
- **User Management:** Keycloak is used for user authentication and role management, differentiating permissions for admin and regular users.
- **UI Refinement:** The UI is being redesigned for improved aesthetics and usability, using Unity's UI Toolkit for creating a panel for sensor selection during digital twin creation/updates.
- **Camera System:** Camera speeds are adapted to map size for intuitive navigation and rotation logic is updated to pivot around user-defined points.

UC_02C_DTC is a crucial component for enabling accurate and immersive XR experiences in other related projects. By providing a robust platform for digital twin creation and management, this use case facilitates data-driven urban planning, tourism, and other city-related applications. The ongoing development, including performance enhancements and UI refinements, ensures the application remains a valuable tool for representing and interacting with urban environments.

While the City Planning use case is mainly conducted by Etteln, the City Maintenance Use case is focused on Vilanova with technical leadership provided by JRS. The project focuses on utilizing sensor data to inform and optimize city maintenance operations, accessible through a web-based dashboard on various devices.

UC04_CM aims to provide the Vilanova i la Geltrú City Council's mobility department with real-time ground information derived from sensor networks. This data-driven approach facilitates prompt decision-making and efficient maintenance interventions, addressing urban infrastructure issues at the earliest possible stage. The core functionality involves visualizing sensor data related to road conditions (e.g., bumps, lines, potholes) and parking availability through an interactive web dashboard.

This use case assumes the availability of a network of sensors providing real-time data on urban infrastructure conditions. The primary objective is to enable timely maintenance decisions based on this sensor-derived information, directly contributing to improved city welfare. All sensors that are being used to enhance this dashboard are compatible with the FIWARE ecosystem and some of the most used communication protocols.

From a technological point of view, users access a web-based dashboard via various devices (smartphones, tablets or PCs) after logging in. The dashboard provides a top-down view of the Vilanova i la Geltrú area, visualizing data collected from sensors regarding road damage (bumps, lines, potholes) and parking areas. The visualization

incorporates various techniques, including filtering and heat-maps, to facilitate data analysis and interpretation. This allows maintenance personnel to quickly identify areas requiring attention and prioritize interventions.

UC04_CM offers a promising approach to proactive city maintenance by leveraging sensor data and visualization tools. By providing real-time information on urban infrastructure conditions, the project wants to empower the Vilanova i la Geltrú mobility department to make informed decisions and efficiently address maintenance needs. Ongoing research and development will further refine the system and maximize its impact on city maintenance operations.There are some interesting development areas such as:

- **Sensor Technology Specification:** There are different sensors sources as air pollution detector or noise detectors, among others.

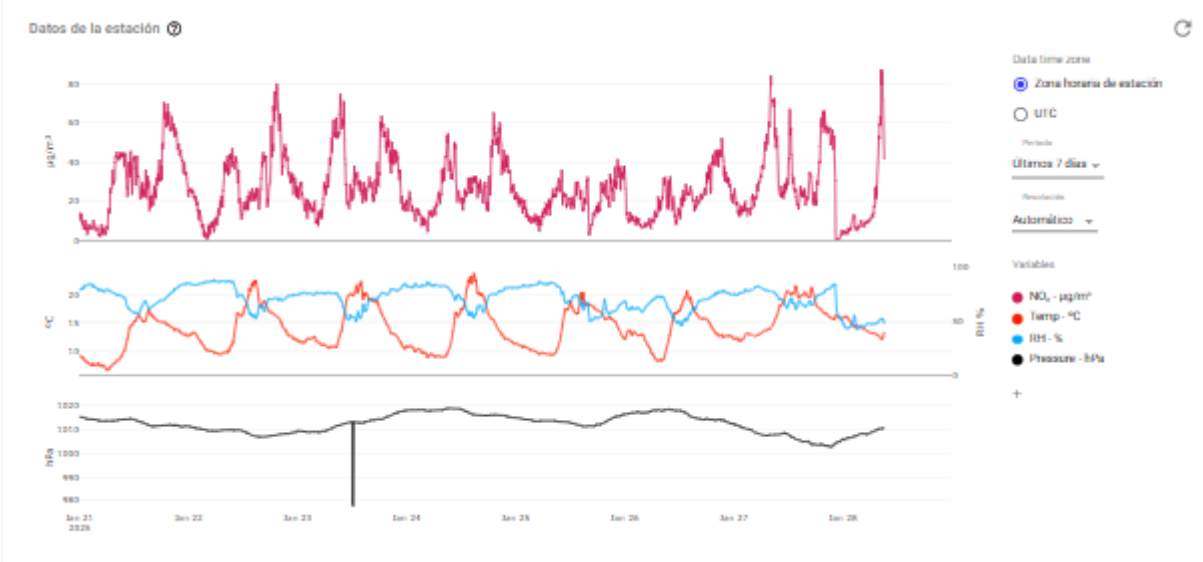

Figure 2: Data from Bettair sensors placed at Neàpolis, Vilanova i la Geltrú.

- **Data Integration and Processing:** Establishing robust data pipelines for integrating sensor data into the web dashboard and implementing appropriate data processing techniques.
- **Visualization Enhancements:** Exploring advanced visualization techniques to provide more insightful information to users.
- Integration with Maintenance Workflows: Defining how the data from the dashboard will be integrated into the existing maintenance workflows of the Vilanova i la Geltrú mobility department.
- **Performance Optimization:** Optimizing the performance of the web dashboard to ensure responsiveness and scalability.

One of the other use cases to be validated in Vilanova is UC03C_VRVT, a Virtual Reality (VR) Virtual Tourism initiative.This project is a collaborative use case between FIWARE and VNG, with technical leadership provided by DTT. The project leverages both Virtual and Augmented Reality technologies, deployed on headsets for remote experiences and mobile devices for potential on-site applications.

UC03C_VRVT provides users with immersive virtual tours of cities, allowing remote exploration of environments and key Points of Interest (POIs). Users navigate a virtual

environment, initially from a bird's-eye view, accessing the latest updated version of a digital twin of the city. These digital twin updates are managed within the "Digital Twin Creation" use case. The application allows users to visit highlighted indoor and outdoor locations within the virtual environment.

Users access the VR application through a wearable VR device, experiencing a virtual tour of the city. Navigation begins with a bird's-eye view, transitioning to ground-level exploration of the virtual environment. The application offers several core functionalities:

- **Free Movement, Navigation & Interaction:** Users can freely navigate the virtual environment using the XR Interaction Toolkit's Locomotion System (teleportation and continuous movement) and interact with objects using XR Grab Interactable.
- **Travel to POIs**: A UI map displays available POIs, allowing users to select locations to move between points.
- **Access Information**: Historical, cultural, or technical information is presented via pop-up or 3D panels triggered by proximity or gaze detection.
- **Take Snapshot:** Users can capture snapshots of the virtual environment using a virtual camera, saving images locally.
- **Change Weather/Time:** Dynamic weather and time changes are simulated using Unity's Shader Graph, with adjustable parameters controlled by sliders.
- **Tour Guide Role:** Configure POIs: An editor interface enables authorized users (Tour Guides) to define POI locations and add relevant information, stored in Scriptable Objects for runtime access.

Tools and Technologies: The technological suite used for this use case is composed by 5 elements: Unity Engine, XR Interaction Toolkit, Cinemachine, Shader Graph, PUN.

UC03C_VRVT provides a powerful tool for remote city exploration and tourism promotion. By leveraging VR technology and a regularly updated digital twin, the application offers immersive experiences and valuable information to users. The functionalities provided, combined with the Tour Guide role, allow for dynamic content management and enhanced user engagement. This use case will increase significantly to innovative approaches in urban tourism and digital city experiences.

UC03C_ARVT: Augmented Reality Virtual Tourism

This paragraph details the research, development, and implementation progress of UC03C_ARVT, an Augmented Reality (AR) Virtual Tourism. This project is a collaborative effort between FIWARE and VNG, with technical leadership provided by UPAT. The project leverages Augmented Reality technology deployed on both mobile devices for on-site experiences and headsets for remote access.

UC03C_ARVT aims to enhance tourist engagement and city promotion by providing users with immersive AR experiences at designated Points of Interest (POIs). The core functionality involves users scanning their environment using AR devices, triggering the display of virtual portals that reveal the interior of landmark buildings. Users can then virtually "enter" these buildings and explore their contents.

The project assumes users are geolocated at designated POIs, equipped with AR devices, and have sufficient physical space for movement and interaction within the AR environment.

The project has been deployed in Vilanova i la Geltrú and Etteln, focusing on key locations:

- **Vilanova i la Geltrú:** Victor Balaguer Museum (BMVB), Train Station, Neapolis Building.
- **Etteln:** Community Center.

The implementation follows a phased approach, allowing for continuous development, testing, and refinement based on user feedback and operational considerations gathered throughout 2024. This iterative process ensures a controlled rollout and optimizes the user experience.

At Vilanova, Victor Balaguer Museum is currently a key focus of the implementation phase. The museum is actively testing the AR application to assess its potential for broadening tourist engagement. Concurrently, the Vilanova i la Geltrú City Council's Tourism Department is evaluating the AR experience as a strategic tool for city promotion and enhanced visibility.

Future development will concentrate on incrementally expanding the AR experience within the Victor Balaguer Museum, adding more rooms and showcasing additional sculptures. This expansion reinforces the museum's role as a key cultural asset within the city.

By prioritizing these implementation strategies and potential applications, Vilanova i la Geltrú effectively leverages its "living lab" status to foster innovation and enhance the quality of life for both residents and visitors. The project's emphasis on the quadruple helix model—involving academia, industry, government, and citizens—is crucial for ensuring the responsible and beneficial development and implementation of these technologies for the entire community.

1.3 Use case specific for Ideal Works

Idealworks aims to enable logistics of the future by providing the most flexible & scalable robotics solutions. Idealworks offers its customers reliable hardware and software to support their logistics processes:

1. The “iw.hub” is Idealworks’ autonomous mobile transport robot, which operates within a fleet of robots. The iw.hub can transport up to 1000 kg and navigate autonomously in an intralogiscs area to transport material from a pick-up area to a drop off location.

2. “AnyFleet” is the central cloud-based fleet management system for autonomous mobile robots. AnyFleet steers the operations of heterogeneous AMR fleets through a unified communication interface. AnyFleet’s solution stack includes, among other modules, task assignment, routing, traffic management, parking and charging assignment. AnyFleet is also an integration platform, supporting industrial IoT solutions and ERP/WMS systems.

Deploying a new AMR fleet at a new customer site is a time-consuming process, requiring the following from a human operator:

To run one robot manually in mapping mode to generate a 2D SLAM map.

To upload the generated map to the cloud steering platform.

To place the relevant Interest Points (IPs) and draw the zones on the base map on the frontend of the AnyFleet platform.

The cloud steering platform (AnyFleet) broadcasts the map to all other robot agents of same type that are onboarded to operate in same location. The generated map thus impacts (1) the quality of robot localization and navigation, and (2) the quality of cloud-based traffic management.

Within the DIDYMOS-XR industry use case, novel approaches were proposed for facilitating the updating of the map and points of interest through human-in-the-loop experience.

2. Use case validation

2.1 Use case validation in Etteln

The use case validation in Etteln recruited people from three pools:

Youth Club: The “Morgenmacher Campus Party Etteln” is a youth club in Etteln that offers kids and young adults the opportunity to work with new technology under supervision. The goal is to interest the members in the technology to develop prototypes that contribute to the city goals. The club is equipped with 3D printers and XR Headsets. While the participants from this group are Etteln-native, tech-savvy and open for new technologies, their past experience results in an heightened standard/expectations towards the validated usecases resulting in feedback regarding the visual representation technics.

Senior citizens club: The “Senioren Computer Club Etteln” is a club to introduce senior citizens to the digital topics, starting with the general use of computers to the use of Etteln’s smart city apps. Most members spent their whole life in Etteln and are a valuable feedback regarding the use case’s content. Due to the limited experience with new technologies, this user group can provide feedback about usability.

Projects and municipality: Due to FIWARE’s participation in several Etteln based projects, a strong connection to both members of the local municipality and companies in the area exists. These people have a like-minded view on the digitalisation in Etteln and are open to explore new approaches and technics. While their technological background differs wildly, they have a deep knowledge of the location and understanding of use requirements/expectations in common.

The first validation round was run during Q4 2024 in the Etteln Community Hall using the bought Smart Phones and a Meta Quest 3 lent by the HSHL.

The use case UC03C_ARVT was being tested using Samsung S23 and S24 and focused on the augmented reality representation of the community hall. The

participants were able to explore the building through the portal mechanism and receive additional information about upcoming events.

Figure 3 – Etteln: AR Virtual Tourism

The use case UC03C_VRVT was tested with the supplied head-mounted-display and enabled the users to experience a big part of the town of Etteln. Via teleportation and free-exploration the participants were able to visit both the community hall and the church and get a feeling for the town.

Figure 4 – Etteln: VR Virtual Tourism

The use cases UC02C_DTC and UC05C_CP were tested with supplied notebooks in the community center and remotely. The participants were able to switch from the drone scan input to the digital twin model and update parts of the model with higher-resolution models. Utilizing a simple control mechanism the participants were able to

adjust the experience to different weather and time-of-day scenarios. Additionally to the local participants, a group of volunteers in Vilanova was used for this validation. This group of users was kindly recruited by our partners at VNG using vouchers as an incentive.

Figure 5: Etteln: Digital Twin Creation

Figure 6 – Etteln: City Planning

2.1.1 Digital Twin Creation

Introduction

Objective: The objective of the study was to evaluate the usability of the application. To achieve this, a supervised usability test was selected as the study's methodology. A supervised usability test is a structured evaluation method designed to assess the effectiveness, efficiency, and user satisfaction of a product, system, or interface. In this approach, participants complete a series of predefined tasks in a controlled testing environment under the observation of a facilitator or moderator. The facilitator provides guidance when necessary, records observations, and may ask follow-up questions to better understand the participants' thought processes, challenges, and overall experience.

Experiment Goals: The expected outcome of the study was to identify usability issues, uncover areas for improvement, and gather insights into how users interact

with the application. These findings would serve as a foundation for refining the application to better meet user needs and expectations.

Experimental Setup

Environment:

- The tests did not have any special requirements regarding the test environment, therefore the evaluation was done in multiple unrelated places (inside the Etteln community hall, remotely in VNG and in offices).

Hardware and Tools

- The use case utilizes the Unity Engine, specifically version 2022.3.14f1, as its development platform. For VR, MR, and AR interactions, it incorporates the XR Interaction Toolkit. Texture and mesh generation is achieved through Neural Radiance Fields (NeRF) and Gaussian Splatting. Web-based previews and accessibility are facilitated by WebGL, while Potree is employed for rendering point clouds.
- Notebooks used for the evaluation did not contain any specialized components but were recent (less than 3 years old) consumer models with a stable internet connection.

Subject Groups

- **Etteln inhabitants**: The Etteln inhabitants group consisted of 8 participants, comprising 1 elderly person, 2 youngsters, and 5 middle-aged persons. The mean age was 44.13 years, with ages ranging from 19 to 72 years. The group included 3 males and 5 females, and the average ATI score was 2.93.
- **Stakeholder board:** 3 persons from the Didymos Stakeholder board participated for this sample group. Among them 2 males and 1 female with an average age of 38. Their average ATI score was 4.7.
- **Vilanova remote:** The Vilanova remote group span 16 people among them 9 males and 7 females. Their average ATI score was 3.98.

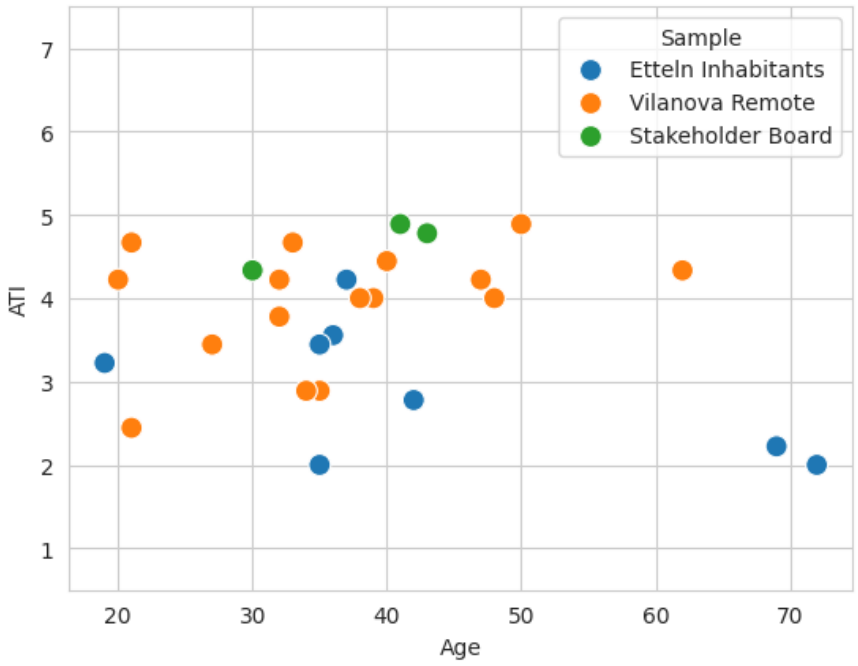

Figure 7: All three sample groups with ATI scores and age (N=27)

Experiment Design

Independent Variables:

- **Cognitive Task Load**: Measured using the NASA-TLX (Task Load Index), a widely used tool for assessing perceived workload. Participants rate their experiences on dimensions such as mental demand, physical demand, temporal demand, effort, performance, and frustration. This provides a comprehensive view of task difficulty.
- **User Experience:** Measured using the UEQ-S (User Experience Questionnaire - Short version). This tool assesses the user's perception of the application's usability and aesthetics, focusing on factors like attractiveness, efficiency, dependability, and stimulation.
- **Emotion**: Measured using the Self-Assessment Manikin (SAM), a non-verbal assessment technique. Participants rate their emotional state on dimensions of valence (positive/negative feelings), arousal (calmness/excitement), and dominance (sense of control).
- **Custom metrics:** Measured through a custom made questionnaire. It included a mix of quantitative and qualitative questions aimed at assessing enjoyment, ease of use, and preferences across different features and devices.

Dependent Variables:

- **Gender**: participants' gender is recorded as a categorical variable (e.g., male, female, non-binary, etc.). This helps in identifying potential gender-based differences in task load perception, user experience, and emotional responses.
- **Age**: the age of participants is recorded as a continuous variable or categorized into age groups (e.g., 18–25, 26–35, etc.). This allows for an analysis of how usability, workload, and emotions differ across age demographics.
- **Technological Affinity:** Measured using the Affinity for Technological Interaction (ATI) scale, understanding the participants' technological affinity helps determine

whether difficulties in testing applications stem from a general lack of interaction with technical systems or issues specific to the program.

Procedure

Task Execution

Participant Introduction: The validation session began with a warm welcome to all participants. They were provided with a short overview of the project, including its objectives and the core functionalities of the software. This introduction ensured that all participants had a clear understanding of the system they were about to test.

Explanation of Test Steps: Participants were given a detailed explanation of the test steps using a prepared manual. The manual contained structured instructions to guide them through the validation process, ensuring consistency in the execution of tasks.

Software Access: Participants were provided with notebooks to access the software's web front-end. They successfully logged into the system, demonstrating the accessibility and usability of the authentication process.

Inspection of Raw Point Cloud: Once logged in, participants examined the raw point cloud data within the software. This step allowed them to assess the initial visualization and representation of the digital environment before any transformations were applied.

Transformation to Digital Twin: Participants activated the transformation process by pressing the designated button. This action initiated the conversion of the raw point cloud into a structured digital twin, allowing an assessment of the software's processing capabilities and speed.

Asset Library Usage: Participants interacted with the asset library to replace a low-quality section of the digital twin with a high-quality version. Specifically, the town church was substituted during the test. This step demonstrated the functionality and ease of use of the asset library for enhancing digital twin accuracy.

Post-Test Questionnaires: Following the test, participants were introduced to a set of questionnaires designed to gather feedback on their experience. The questionnaires were completed by all participants, providing valuable insights into the usability, performance, and potential areas for improvement within the software.

Data Collection and Analysis

- **Demographics**
- **Affinity for Technological Interaction (ATI)**: The ATI provides insights into the interest and experience of subject groups in interacting with technology and technical systems. Using a 9-item questionnaire with 7-point Likert scales an overall ATI score can be calculated.
- **NASA Task Load Index (NASA-TLX):** The NASA-TLX is a widely used, subjective, multidimensional assessment tool designed to evaluate perceived workload during a task. It measures six dimensions:
    - **Mental Demand:** Assesses the mental and perceptual activity required.
    - **Physical Demand:** Evaluates the physical effort needed.
    - **Temporal Demand:** Considers the time pressure felt during the task.

  - **Performance:** Reflects the individual's self-assessment of task success.
  - **Effort:** Measures the amount of work exerted to achieve performance.
  - **Frustration Level:** Captures feelings of irritation, stress, and annoyance.

  Participants rate each dimension, and these ratings are combined to produce an overall workload score. The tool is versatile and has been applied in various fields, including aviation and healthcare. For more information and access to the questionnaire materials, visit the official NASA-TLX website.
- **User Experience Questionnaire - Short Version (UEQ-S):** The UEQ-S is a concise version of the original User Experience Questionnaire, tailored for scenarios where a quick assessment is needed. It evaluates user experience across two primary dimensions:
  - **Pragmatic Quality:** Focuses on usability aspects such as efficiency, clarity, and dependability.
  - **Hedonic Quality:** Addresses the product's ability to stimulate and engage users, including aspects like excitement and innovation.

  The UEQ-S consists of 8 items, each presented as a pair of contrasting adjectives (e.g., "complicated" vs. "easy"). Participants rate their experience on a 7-point scale between these adjectives.
- **Self-Assessment Manikin (SAM):** The Self-Assessment Manikin is a non-verbal pictorial assessment technique that measures emotional responses across three dimensions:
  - **Valence (Pleasure):** Ranges from happy to unhappy.
  - **Arousal:** Spans from excited to calm.
  - **Dominance:** Varies from controlled to in control.

  Participants select images that best represent their feelings in each dimension, providing insights into their emotional state during or after interacting with a product or system.
- **Custom made questionnaire:** The questionnaire consisted of the following sections:
  - **Overall Enjoyment:** Participants rated their overall enjoyment of the application on a **7-point Likert scale** (1 = very low enjoyment, 7 = very high enjoyment).
  - **Feature-Specific Usability Ratings:** Participants provided usability ratings for specific features of the application, such as:
    - **Placing a Frame**: Rated on a 7-point scale for ease of use.
    - **Finding Points of Interest (POIs)**: Rated on a 7-point scale for ease of discovery.
    - **Opening and Looking Through the Portal**: Rated on a 7-point scale to assess the intuitiveness and engagement of these features.
  - **Device Preferences:** Participants were asked to express their preferences for the device they would like to use with the application (e.g., smartphones or AR headsets such as the HoloLens).
  - **Open-Ended Feedback:** Participants were invited to share additional comments, providing qualitative insights into their experiences, challenges, and specific issues (e.g., difficulties in finding POI information).

Results

Task Load: The NASA-TLX results (on a scale from 1 to 7) reveal insights into perceived workload across three sample groups. The Etteln inhabitants group reported a workload score of 3.13, indicating a low-to-moderate level of perceived effort or demand during the task. The Stakeholder board group reported a lower workload score of 2.27, suggesting minimal perceived demands or effort. Similarly, Remote validations from Vilanova had the lowest score of 2.06, reflecting an easy and minimally taxing workload. Overall, the findings suggest that all groups experienced relatively low levels of task-related workload, with slight variations across sample groups.

User Experience: The UEQ-S results (on a scale from -3 to 3), provide an overview of user experience evaluations across three groups. Etteln inhabitants reported an Overall score of -0.34, indicating a negative experience, with Hedonic and Pragmatic scores of -0.09 and -0.59, respectively, suggesting a balanced but unfavorable perception of enjoyment and usability. The Stakeholder board showed an Overall score of 0.91, reflecting a positive experience, with a moderate focus on Hedonic aspects (0.41) and a stronger perception of Pragmatic qualities (1.41). Finally, Remote Vilanova recorded the most favorable evaluations, with an Overall score of 1.07. The scores for Hedonic (1.09) and Pragmatic (1.04) were similarly high, highlighting both enjoyment and usability (Figure 8). Following the UEQ benchmarks, Etteln inhabitants reported a “bad” experience, the Stakeholder board a “below average” and Remote Vilanova groups had a “above average” experience.

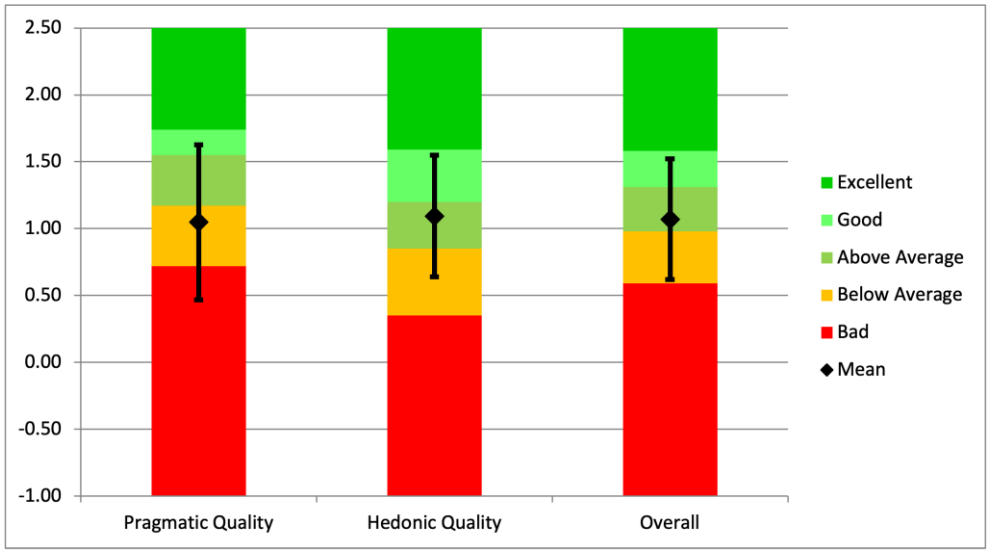

Figure 8: UEQ-S Benchmark for the Vilanova Remote Group (N=16) in Digital Twin Creation.

Emotion: The SAM results (on a scale of 1 to 5) for the group of Etteln inhabitants reveal a valence score of 2.837, indicating a neutral or indifferent emotional state, neither distinctly positive nor negative. The arousal score of 2.37 suggests a slight tendency towards lowered arousal, while the dominance score of 3.25 reflects a more balanced or average sense of control.

Custom Questionnaire: The custom questionnaire results (all answers on a scale of 1 to 7) for three groups—Etteln inhabitants, the stakeholder board, and the remote Vilanova group—show varying levels of satisfaction and ease. Regarding overall enjoyment of the experience, Etteln inhabitants rated it 4.13, the

stakeholder board gave it 3.67, and remote Vilanova had the highest rating with 5.19. When it comes to the ease of navigating the scene, Etteln inhabitants rated it 4.00, while the stakeholder board and remote Vilanova found it easier, scoring 5.00 and 5.69, respectively. As for creating the DT, Etteln inhabitants found it quite easy with a score of 5.50, compared to 4.67 for the stakeholder board, and remote Vilanova, once again, rated it the highest at 6.06. In terms of removing a low-quality building, all groups found it relatively easy, with Etteln inhabitants scoring 5.50, the stakeholder board at 5.34, and remote Vilanova at 6.00. Finally, when adding the high-quality building, Etteln inhabitants found it more challenging with a score of 3.38, while the stakeholder board and remote Vilanova rated it significantly easier, with scores of 5.67 and 6.19, respectively. The comments from the additional feedback section indicates, that the participants had problems with the navigation speed and the positioning of buildings, specifically the possibility to rotate them (appendix).

Correlation: The three sample groups were aggregated (N = 27) to examine correlations between the metrics using Spearman's rank correlation. A moderate positive correlation was observed between ATI and overall UEQ-S scores ($\rho = 0.54$, $p < 0.001$). This relationship appears to be primarily driven by the pragmatic quality dimension from UEQ-S, which showed a strong positive correlation with ATI ($\rho = 0.70$, $p < 0.001$).

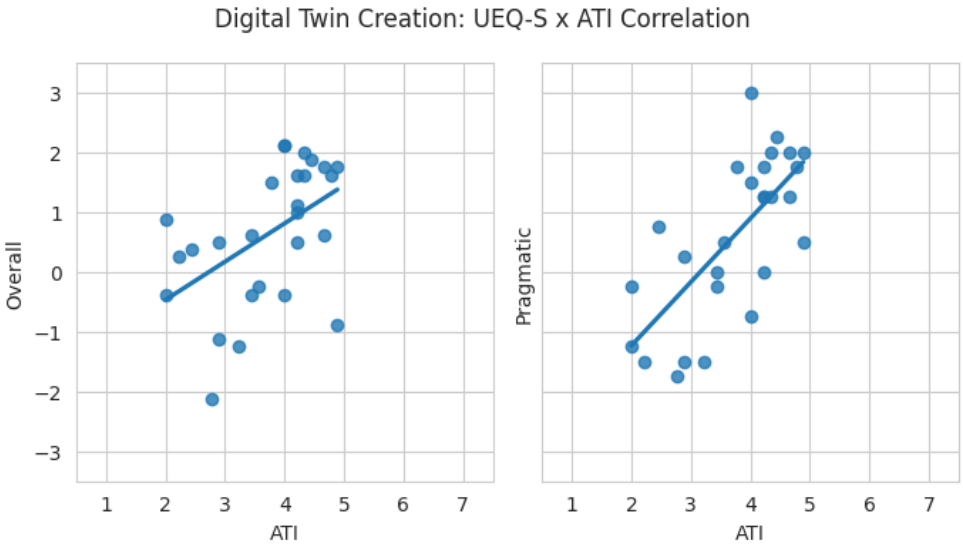

Figure 9:Correlations between ATI and UEQ-S (Overall, Hedonic, Pragmatic) for aggregated sample group (N=27).

Conclusion

The validation was conducted with a total of 27 participants distributed across three sample groups. The largest group consisted of 16 participants, while the smallest, the stakeholder board group, had only 3 participants. Given the limited sample sizes within each group, the results must be interpreted with caution, as smaller groups may not provide robust or generalizable findings. The results indicate that all three groups experienced relatively low task-related workload showing that the participants were not mentally or physically challenged. The highest task load was reported in the Etteln inhabitant's group. User experience evaluations had a great variance between the groups. Specifically, Etteln inhabitants reported a negative experience, the stakeholder board showed a below-average experience, and remote Vilanova

participants reported an above-average user experience. Most participants rated the hedonic aspects of the application higher than its pragmatic. Emotionally, the Etteln group remained neutral with low arousal and an average sense of control. The emotion assessment was then discontinued due to the feedback from the study instructors. The custom questionnaire results confirm that the specific tasks were easy to carry out. Issues were raised regarding the navigation in the scene with both mouse and trackpad and the placement of buildings, specifically rotation was mentioned as improvable. For performing a correlation analysis between the assessed metrics, the sample groups were merged to increase the data points and provide more reliable results. A notable correlation was found between ATI and user experience, particularly in the pragmatic dimension, suggesting that technology-prone participants rated usability better.

2.1.2 City Planning

Introduction

Objective: The objective of the study was to evaluate the usability of the application. To achieve this, a supervised usability test was selected as the study's methodology. A supervised usability test is a structured evaluation method designed to assess the effectiveness, efficiency, and user satisfaction of a product, system, or interface. In this approach, participants complete a series of predefined tasks in a controlled testing environment under the observation of a facilitator or moderator. The facilitator provides guidance when necessary, records observations, and may ask follow-up questions to better understand the participants' thought processes, challenges, and overall experience.

Experiment Goals: The expected outcome of the study was to identify usability issues, uncover areas for improvement, and gather insights into how users interact with the application. These findings would serve as a foundation for refining the application to better meet user needs and expectations.

Experimental Setup

Environment:

- The tests did not have any special requirements regarding the test environment, therefore the evaluation was done in multiple unrelated places (inside the Etteln community hall, remotely in VNG and in offices)

Hardware and Tools:

- The Usecase is developed using the Unity Engine as its core platform. It integrates SUMO (Simulation of Urban Mobility) and Carla, an open-source urban simulation software, to enable realistic urban modeling and simulation. The Destination Earth API is utilized for accessing geospatial data and enhancing environmental accuracy. Additionally, the Fluid Frenzy asset from the Unity Asset Store is incorporated to enrich visual and interactive elements
- Notebooks used for the evaluation did not contain any specialized components but were recent (less than 3 years old) consumer models with a stable internet connection.

Subject Groups

- **Etteln inhabitants:** The sample group is the same as described in 2.1.3. Digital Twin Creation - Subject Groups.
- **Stakeholder board:** The sample group is the same as described in 2.1.3. Digital Twin Creation - Subject Groups.
- **Vilanova remote:** The group was intended to be the same as described in Digital Twin Creation. However, one participant had to be removed due to incomplete data, resulting in a final sample size of 15 participants (8 males and 7 females) with an average ATI score of 3.95.

Experiment Design

Independent Variables:

- **Cognitive Task Load**: Measured using the NASA-TLX (Task Load Index), a widely used tool for assessing perceived workload. Participants rate their experiences on dimensions such as mental demand, physical demand, temporal demand, effort, performance, and frustration. This provides a comprehensive view of task difficulty.
- **User Experience:** Measured using the UEQ-S (User Experience Questionnaire - Short version). This tool assesses the user's perception of the application's usability and aesthetics, focusing on factors like attractiveness, efficiency, dependability, and stimulation.
- **Emotion**: Measured using the Self-Assessment Manikin (SAM), a non-verbal assessment technique. Participants rate their emotional state on dimensions of valence (positive/negative feelings), arousal (calmness/excitement), and dominance (sense of control).
- **Custom metrics:** Measured through a custom made questionnaire. It included a mix of quantitative and qualitative questions aimed at assessing enjoyment, ease of use, and preferences across different features and devices.

Dependent Variables:

- **Gender**: participants' gender is recorded as a categorical variable (e.g., male, female, non-binary, etc.). This helps in identifying potential gender-based differences in task load perception, user experience, and emotional responses.
- **Age**: the age of participants is recorded as a continuous variable or categorized into age groups (e.g., 18–25, 26–35, etc.). This allows for an analysis of how usability, workload, and emotions differ across age demographics.
- **Technological Affinity:** Measured using the Affinity for Technological Interaction (ATI) scale, understanding the participants' technological affinity helps determine whether difficulties in testing applications stem from a general lack of interaction with technical systems or issues specific to the program.

Procedure

Task Execution

Explanation of Test Steps: Participants were given a detailed explanation of the test steps using a prepared manual. The manual contained structured instructions to guide them through the validation process, ensuring consistency in the execution of tasks.

Opening the Digital Twin: Participants accessed the software and opened a previously created digital twin. This step ensured that the system correctly loaded and displayed the stored digital environment.

Accessing the City Planning Features: Participants used the "City Planning" button to open the weather control menu. This menu provided interactive options to adjust environmental factors such as lighting and precipitation.

Adjusting Weather and Light Conditions: Participants interacted with the available controls to modify the digital twin's environment. The controls included:

- Time of day: Adjusting sunlight position based on different times.
- Sun intensity: Modifying brightness and overall lighting conditions.
- Rain intensity: Controlling the amount of rainfall in the environment.
- Rain drop size: Altering the visual representation of rain effects.

This step allowed participants to explore various weather scenarios and assess their impact on the digital twin.

Completion and Feedback Collection: Following the test, participants were introduced to questionnaires designed to gather their feedback on usability, performance, and experience with the city planning features. Each participant completed the questionnaire, providing insights into potential improvements.

Data Collection and Analysis

- **Demographics**
- **Affinity for Technological Interaction (ATI)**: The ATI provides insights into the interest and experience of subject groups in interacting with technology and technical systems. Using a 9-item questionnaire with 7-point Likert scales an overall ATI score can be calculated.
- **NASA Task Load Index (NASA-TLX):** The NASA-TLX is a widely used, subjective, multidimensional assessment tool designed to evaluate perceived workload during a task. It measures six dimensions:
    - **Mental Demand:** Assesses the mental and perceptual activity required.
    - **Physical Demand:** Evaluates the physical effort needed.
    - **Temporal Demand:** Considers the time pressure felt during the task.
    - **Performance:** Reflects the individual's self-assessment of task success.
    - **Effort:** Measures the amount of work exerted to achieve performance.
    - **Frustration Level:** Captures feelings of irritation, stress, and annoyance.

  Participants rate each dimension, and these ratings are combined to produce an overall workload score. The tool is versatile and has been applied in various fields,

including aviation and healthcare. For more information and access to the questionnaire materials, visit the official NASA-TLX website.

- **User Experience Questionnaire - Short Version (UEQ-S):** The UEQ-S is a concise version of the original User Experience Questionnaire, tailored for scenarios where a quick assessment is needed. It evaluates user experience across two primary dimensions:
    - **Pragmatic Quality:** Focuses on usability aspects such as efficiency, clarity, and dependability.
    - **Hedonic Quality:** Addresses the product's ability to stimulate and engage users, including aspects like excitement and innovation.

  The UEQ-S consists of 8 items, each presented as a pair of contrasting adjectives (e.g., "complicated" vs. "easy"). Participants rate their experience on a 7-point scale between these adjectives.
- **Self-Assessment Manikin (SAM):** The Self-Assessment Manikin is a non-verbal pictorial assessment technique that measures emotional responses across three dimensions:
    - **Valence (Pleasure):** Ranges from happy to unhappy.
    - **Arousal:** Spans from excited to calm.
    - **Dominance:** Varies from controlled to in control.

  Participants select images that best represent their feelings in each dimension, providing insights into their emotional state during or after interacting with a product or system.
- **Custom made questionnaire:** The questionnaire consisted of the following sections:
    - **Overall Enjoyment:** Participants rated their overall enjoyment of the application on a **7-point Likert scale** (1 = very low enjoyment, 7 = very high enjoyment).
    - **Feature-Specific Usability Ratings:** Participants provided usability ratings for specific features of the application, such as:
        - **Placing a Frame**: Rated on a 7-point scale for ease of use.
        - **Finding Points of Interest (POIs)**: Rated on a 7-point scale for ease of discovery.
        - **Opening and Looking Through the Portal**: Rated on a 7-point scale to assess the intuitiveness and engagement of these features.
    - **Device Preferences:** Participants were asked to express their preferences for the device they would like to use with the application (e.g., smartphones or AR headsets such as the HoloLens).
    - **Open-Ended Feedback:** Participants were invited to share additional comments, providing qualitative insights into their experiences, challenges, and specific issues (e.g., difficulties in finding POI information).

Results

- **Task Load**: The NASA-TLX (Task Load Index) scores, which measure perceived workload on a scale from 1 to 7, show variations across the three groups. The Etteln inhabitants reported the highest workload with a score of 2.81, followed by the stakeholder board at 1.94, and the remote Vilanova group with the lowest workload at 1.75. These results indicate that the task load was generally low across all groups, with the Etteln inhabitants perceiving slightly more difficulty compared to the others.

- **User Experience:** The User Experience (UEQ-S) results (on a scale from -3 to 3) highlight notable differences across the three groups. For the overall experience, the Etteln inhabitants scored 0.02, the stakeholder board scored 0.29, which is considered “bad” according to the UEQ benchmark, while the remote Vilanova group achieved a score of 1.69, translating to “excellent,” the highest category (Figure 10). Regarding hedonic quality, the Etteln inhabitants scored -0.5, the stakeholder board scored 0.33 (“bad”), and the remote Vilanova group achieved a “good” score of 1.33. For pragmatic quality, the Etteln inhabitants scored 0.531, the stakeholder board scored 0.25 (“bad”), and the remote Vilanova group achieved a strong score of 2.05, which is again categorized as “excellent”. These results highlight significant differences in user experience among the three groups.

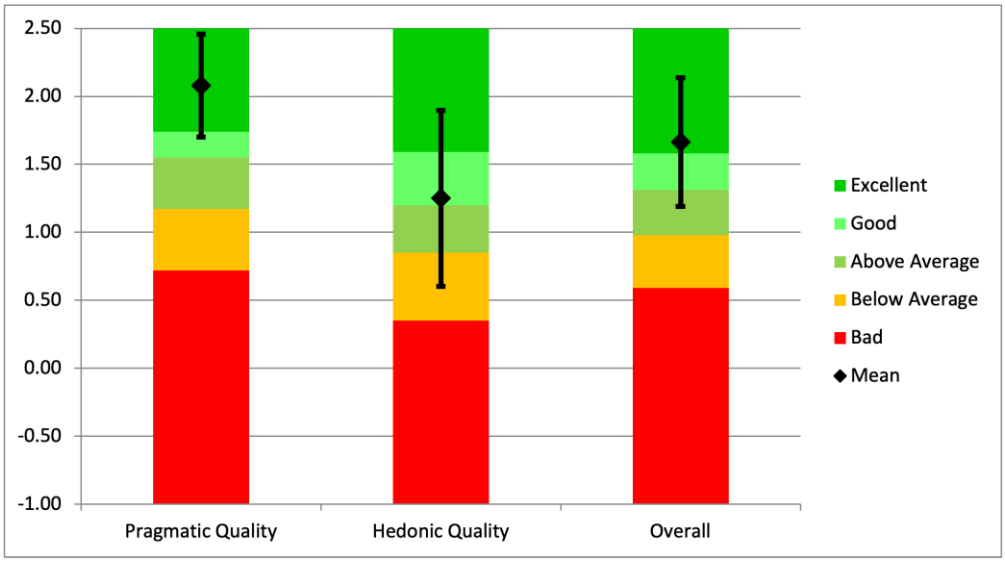

Figure 10: UEQ-S Benchmark for the Vilanova Remote Group (N=16) in City Planning.

- **Emotion:** The SAM (Self-Assessment Manikin) results for the Etteln inhabitants group (on a scale from 1 to 5) reveal a valence score of 3.0, indicating a neutral emotional state. The arousal score of 1.875 reflects a low level of excitement or activation, while the dominance score of 3.0 suggests a moderate sense of control over the situation.
- **Custom Questionnaire:** For overall enjoyment of the experience, the Etteln inhabitants rated it 3.875, the stakeholder board scored 4, and the remote Vilanova group gave the highest score of 5.53. Regarding ease of navigation, the scores were 4.63 for Etteln inhabitants, 6 for the stakeholder board, and 6.2 for remote Vilanova. The ease of changing the weather was rated highly by all groups, with scores of 6.125, 6, and 6.6, respectively. When asked to rate the usefulness of the feature, the Etteln inhabitants scored it 2.375, much lower than the stakeholder board at 6 and the remote Vilanova group at 5.74. The ease of changing rain intensity received high scores of 6.375 from Etteln inhabitants, 6 from the stakeholder board, and 6.87 from remote Vilanova. Similarly, the ease of changing sun intensity was rated 5.875, 6, and 6.87, respectively. Finally, the percentage of users who preferred the Hololens was 25% among Etteln inhabitants, 100% for the stakeholder board, and 80% for the remote Vilanova group. Additional features proposed by the participants were among others: incorporating diverse environmental effects such as snow and varying grass conditions, enabling distance measurement capabilities, integrating multiple layers for environmental impact analysis, including air pollution, noise, and traffic, simulating urban mobility patterns, integrating IoT sensors for real-time data acquisition, enabling

visualization in XR environments, developing scenarios for natural disasters such as floods and earthquakes, facilitating multi-user collaboration with annotation and comment features, and incorporating surface irregularities for more realistic visual representations. Further comments highlighted inconsistencies in the visual representation such as floating buildings and unnatural shadows.

- **Correlation**: The three sample groups were aggregated (N = 27) to examine correlations between the metrics using Spearman's rank correlation. A low to moderate positive correlation was observed between ATI and overall UEQ-S scores ($\rho = 0.38$, $p < 0.05$). This relationship appears to be primarily driven by the pragmatic quality dimension from UEQ-S, which showed also shows a low moderate positive correlation with ATI ($\rho = 0.39$, $p < 0.05$). Further a low to moderate negative correlation was found between ATI and the task load measured with NASA-TLX ($\rho = -0.40$, $p < 0.05$).

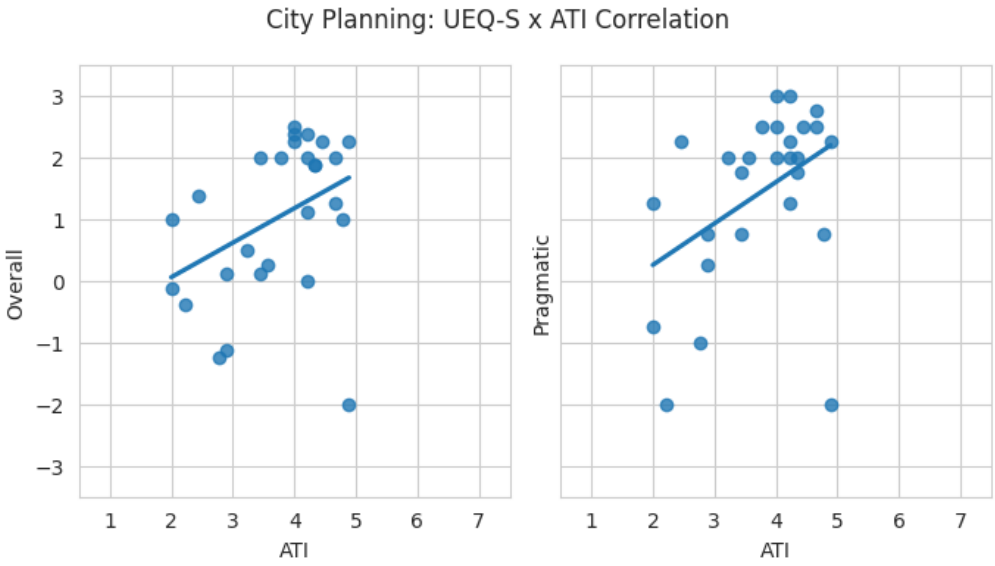

Figure 11:Correlations between ATI and significant UEQ-S values (Overall, Pragmatic) for the aggregated sample group (N=27).

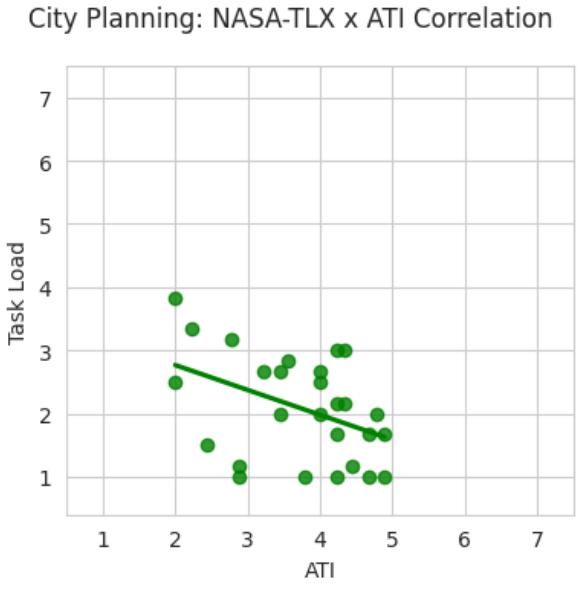

Figure 12:Correlations between ATI and the NASA-TLX score for the aggregated sample group (N=27).

Conclusion

The validation was conducted by the same participants as the Digital Twin Creation Validation which resulted in the same disclaimers for interpretation. The findings highlight notable differences in task load and user experience across the three groups, with the remote Vilanova group generally reporting the most positive outcomes. While task load scores were relatively low overall, Etteln inhabitants perceived slightly higher difficulty. User experience scores varied greatly, with the remote Vilanova group rating the experience as excellent, whereas the stakeholder board reported low scores that are considered bad experiences according to the benchmark. Regarding emotion, the Etteln group reported neutral valence with low arousal and average dominance. Similar to the digital twin creation, the emotion assessment was discontinued. In the custom questionnaire, participants rated ease of navigation and weather-related interactions highly, with the remote Vilanova group consistently providing the most positive feedback, while Etteln inhabitants rated the usefulness of features significantly lower. Suggested improvements that were mentioned in the qualitative feedback included adding environmental effects, real-time data integration, urban mobility simulations and enhanced visual accuracy to address inconsistencies like floating buildings and unnatural shadows. The correlation analysis suggests that a higher affinity for technology is associated with a better user experience, particularly in terms of pragmatic quality and a lower perceived task load.

2.1.3 VR Virtual Tour

Introduction

Objective: The objective of the study was to evaluate the usability of the application. To achieve this, a supervised usability test was selected as the study's methodology. A supervised usability test is a structured evaluation method designed to assess the effectiveness, efficiency, and user satisfaction of a product, system, or interface. In this approach, participants complete a series of predefined tasks in a controlled testing environment under the observation of a facilitator or moderator. The facilitator provides guidance when necessary, records observations, and may ask follow-up questions to better understand the participants' thought processes, challenges, and overall experience.

Experiment Goals: The expected outcome of the study was to identify usability issues, uncover areas for improvement, and gather insights into how users interact with the application. These findings would serve as a foundation for refining the application to better meet user needs and expectations.

Experimental Setup

Environment

- The VR application was tested in a spacious, unobstructed room in the community hall, ensuring ample movement space for participants. To enhance safety, no objects were placed near the testing area, preventing accidental collisions during immersive interactions.
- Each test was conducted individually, with a dedicated safety person standing next to the participant at all times. This approach eliminated parallel test runs but ensured immediate assistance in case of disorientation or imbalance.

- This controlled setup allowed participants to focus entirely on the VR experience while maintaining a safe and distraction-free environment.

Hardware and Tools

- The Usecase is built using the Unity Engine (2023 LTS) as its foundation. Navigation and interaction are handled through the XR Interaction Toolkit. Cinemachine is utilized to implement dynamic camera systems, while Shader Graph is employed to simulate weather and time changes. For multiplayer and shared experiences, Photon Unity Networking (PUN) is integrated into the project.
- Meta Quest Pro: A high-end mixed reality headset designed for immersive AR and VR experiences. It features advanced sensors for precise hand tracking, high-resolution displays for clear visuals, and a comfortable, ergonomic design. The Meta Quest Pro was used in the study to evaluate the application's performance and user interaction in a fully immersive environment, offering a seamless blend of the physical and digital worlds.

Subject Groups

- **Etteln inhabitants:** The Etteln inhabitants group is composed out of 3 subgroups: Elderly (n = 7), Youngsters (n = 5) and Middle-aged (n = 4) summed up to reach a participant count of 16 persons. The sample group had a mean age of 47.81 years ranging from 16 to 79 years. The group was evenly split between males (n = 8) and females (n = 8). Participants reported an average experience with XR technologies of 2.5 on a 7-point scale. Prior to the study, 7 persons had tried VR on a headset, while 5 had tried it on a smartphone. Regarding AR, 3 persons had experienced it through an HMD, and 5 had used a smartphone. The average ATI score was 3.23.

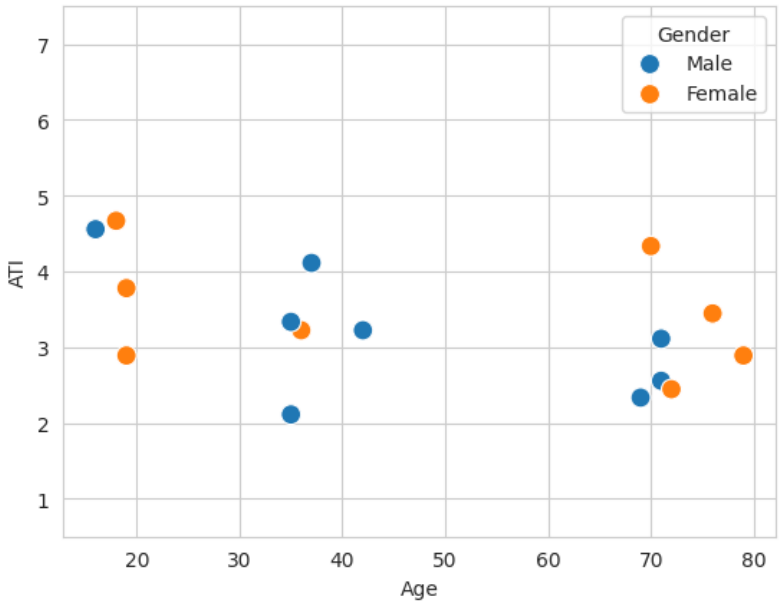

Figure 13: ATI scores and age of the sample group (N=16)

Experiment Design

- **Sense of Presence:** Measured using the Igroup Presence Questionnaire (IPQ), which evaluates participants' sense of presence in a virtual environment. It assesses three key dimensions: spatial presence (the feeling of being physically present in the virtual space), involvement (engagement with the environment), and experienced realism (the perceived authenticity of the environment).

- **Cybersickness:** Measured using the Cybersickness in Virtual Reality Questionnaire (CSQ-VR). This tool assesses symptoms of cybersickness across three categories: nausea (stomach discomfort, urge to vomit), oculomotor symptoms(eye strain, blurred vision, headaches), and disorientation (dizziness, confusion).
- **User Experience:** Measured using the UEQ-S (User Experience Questionnaire - Short version). This tool assesses the user's perception of the application's usability and aesthetics, focusing on factors like attractiveness, efficiency, dependability, and stimulation.
- **Emotion**: Measured using the Self-Assessment Manikin (SAM), a non-verbal assessment technique. Participants rate their emotional state on dimensions of valence (positive/negative feelings), arousal (calmness/excitement), and dominance (sense of control).
- **Custom metrics:** Measured through a custom made questionnaire. It included a mix of quantitative and qualitative questions aimed at assessing enjoyment, ease of use, and preferences across different features and devices.

Dependent Variables:

- **Gender**: participants' gender is recorded as a categorical variable (e.g., male, female, non-binary, etc.). This helps in identifying potential gender-based differences in task load perception, user experience, and emotional responses.
- **Age**: the age of participants is recorded as a continuous variable or categorized into age groups (e.g., 18–25, 26–35, etc.). This allows for an analysis of how usability, workload, and emotions differ across age demographics.
- **Technological Affinity:** Measured using the Affinity for Technological Interaction (ATI) scale, understanding the participants' technological affinity helps determine whether difficulties in testing applications stem from a general lack of interaction with technical systems or issues specific to the program.

Procedure

Training Phase

- Before the test execution, participants underwent a dedicated training phase to ensure comfort and familiarity with the VR system. They were given ample time to properly adjust the head-mounted display (HMD) and complete the necessary calibration process to optimize tracking and visual clarity.
- Additionally, participants had the opportunity to freely explore the headset options before the actual test run. This allowed them to familiarize themselves with the controls, navigation, and interactive elements, ensuring a smoother experience during the evaluation phase.

Task Execution

Introduction and Overview: Participants were welcomed and given a short overview of the project and the software. This introduction provided context for the test and familiarized participants with the application's purpose and features.

Explanation of Test Steps: A structured explanation of the test steps was provided to ensure that participants understood the tasks they were expected to perform.

Training Period: Before starting the main test, participants were given time to familiarize themselves with VR controls and interactions. This training period helped reduce potential usability challenges and ensured a smoother experience.

Entering the Digital Twin: The application was launched, and participants were placed within the town's digital twin. This step ensured that all participants could successfully enter and interact with the virtual environment.

Free Exploration: Participants used free movement controls to explore the digital twin. This step allowed them to navigate the environment naturally and assess the level of detail and immersion.

Teleportation to Points of Interest (POIs): Participants were encouraged to use the teleportation feature to instantly move to designated Points of Interest (POIs). This functionality provided an alternative navigation method, enhancing accessibility within the digital twin.

Adjusting Weather Conditions: Participants accessed the weather control system to modify environmental conditions, including lighting and rain. This step allowed them to experience the impact of different weather scenarios within the virtual environment.

Capturing Screenshots: Participants used the built-in screenshot functionality to capture images of the VR experience. This feature allowed for documentation and review of different perspectives within the digital twin.

Exploring the High-Resolution Community Hall Model: Participants had the opportunity to further explore a high-resolution model of the community hall, assessing the level of detail and realism provided by the application.

Completion and Feedback Collection: Following the test, participants were introduced to questionnaires designed to gather their feedback on usability, performance, and experience with the VR tourism features. Each participant completed the questionnaire, providing insights into potential improvements.

Data Collection and Analysis

- **Demographics**
- **Affinity for Technological Interaction (ATI)**: The ATI provides insights into the interest and experience of subject groups in interacting with technology and technical systems. Using a 9-item questionnaire with 7-point Likert scales an overall ATI score can be calculated.
- **Igroup Presence Questionnaire (IPQ):** The IPQ is a tool designed to measure a user's sense of presence within a virtual environment. It evaluates three primary dimensions:
    - **Spatial Presence:** The feeling of being physically present in the virtual space.
    - **Involvement:** The degree of attention and engagement with the virtual environment.
    - **Experienced Realism:** The perceived authenticity of the virtual environment.

  Participants respond to items on a 7-point Likert scale, with higher scores indicating a stronger sense of presence. The IPQ has been widely used in virtual reality research to assess how immersive and engaging a virtual experience is.

- **Cybersickness in Virtual Reality Questionnaire (CSQ-VR):** The CSQ-VR is a brief tool developed to assess symptoms of cybersickness experienced during virtual reality exposure. It evaluates three categories of symptoms:
    - **Nausea:** Symptoms like stomach discomfort and urge to vomit.
    - **Oculomotor:** Issues such as eye strain, blurred vision, and headaches.
    - **Vestibular:** Feelings of dizziness and confusion.

  Each category includes two questions, and participants rate their symptoms on a 7-point Likert scale, ranging from “1-absent feeling” to “7-extreme feeling.” The CSQ-VR provides a total score and sub-scores for each symptom category, offering a comprehensive assessment of cybersickness severity.
- **User Experience Questionnaire - Short Version (UEQ-S):** The UEQ-S is a concise version of the original User Experience Questionnaire, tailored for scenarios where a quick assessment is needed. It evaluates user experience across two primary dimensions:
    - **Pragmatic Quality:** Focuses on usability aspects such as efficiency, clarity, and dependability.
    - **Hedonic Quality:** Addresses the product’s ability to stimulate and engage users, including aspects like excitement and innovation.

  The UEQ-S consists of 8 items, each presented as a pair of contrasting adjectives (e.g., “complicated” vs. “easy”). Participants rate their experience on a 7-point scale between these adjectives.
- **Self-Assessment Manikin (SAM):** The Self-Assessment Manikin is a non-verbal pictorial assessment technique that measures emotional responses across three dimensions:
    - **Valence (Pleasure):** Ranges from happy to unhappy.
    - **Arousal:** Spans from excited to calm.
    - **Dominance:** Varies from controlled to in control.

  Participants select images that best represent their feelings in each dimension, providing insights into their emotional state during or after interacting with a product or system.
- **Custom made questionnaire:** The questionnaire consisted of the following sections:
    - **Overall Enjoyment:** Participants rated their overall enjoyment of the application on a **7-point Likert scale** (1 = very low enjoyment, 7 = very high enjoyment).
    - **Feature-Specific Usability Ratings:** Participants provided usability ratings for specific features of the application, such as:
        - **Placing a Frame**: Rated on a 7-point scale for ease of use.
        - **Finding Points of Interest (POIs)**: Rated on a 7-point scale for ease of discovery.
        - **Opening and Looking Through the Portal**: Rated on a 7-point scale to assess the intuitiveness and engagement of these features.
    - **Device Preferences:** Participants were asked to express their preferences for the device they would like to use with the application (e.g., smartphones or AR headsets such as the HoloLens).
    - **Open-Ended Feedback:** Participants were invited to share additional comments, providing qualitative insights into their experiences, challenges, and specific issues (e.g., difficulties in finding POI information).

Results

- **Presence**: For the Etteln inhabitants, the IPQ scores (on a scale of 1 to 7) highlight mixed perceptions. The overall sense of presence was rated at 3.93, while involvement scored slightly higher at 4.15. The sense of spatial presence was rated at 4.23, indicating a moderate connection to the virtual environment. However, the realness of the experience was rated notably lower at 2.40, suggesting room for improvement in the perceived authenticity of the environment.
- **Cybersickness:** The CSQ-VR results indicate a total score of 15.50 (on a scale of 6 to 42), reflecting a moderate level of cybersickness experienced. Among the subscales, the vestibular symptoms score was 5.13 (on a scale from 2 to 14), the oculomotor symptoms score was slightly higher at 5.81 (on a scale from 2 to 14), and the nausea symptoms score was 4.56 (on a scale from 2 to 14), indicating relatively lower levels of discomfort in this category. Additional comments indicated explicitly eye strain, dizziness and a fuzzy stomach.
- **User Experience:** The UEQ-S results, measured on a scale from -3 to 3, indicate suboptimal outcomes across all categories. The overall user experience score was -0.10, while the hedonic quality scored slightly higher at 0.18. However, the pragmatic quality scored the lowest at -0.39. According to the UEQ-S benchmark, these scores are categorized as "bad" in all evaluated aspects.

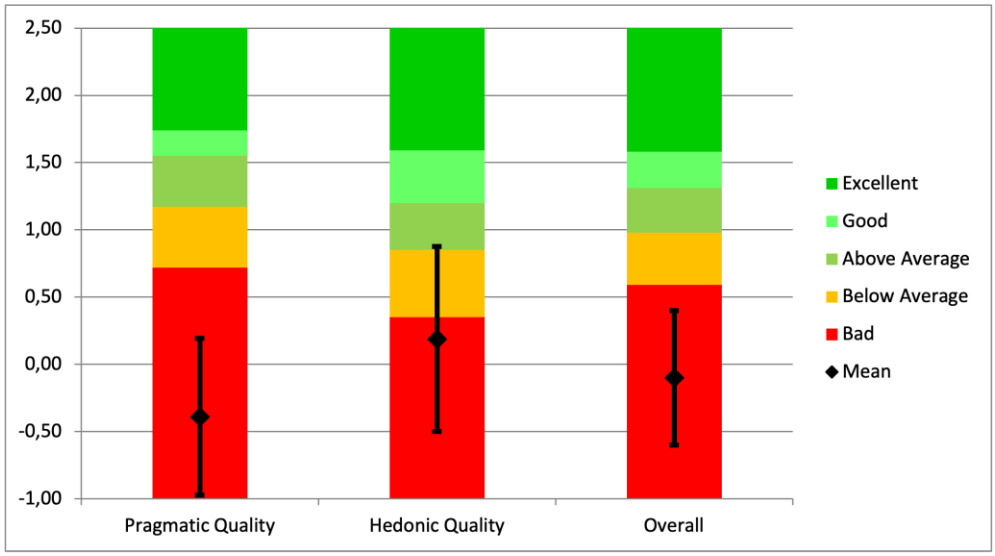

Figure 14: UEQ-S Benchmark for the Etteln inhabitants Group (N=16) in VR Virtual Tour.

- **Emotion:** The SAM results (on a scale from 1 to 5) indicate a generally neutral to slightly negative experience. The valence score of 2.37 reflects a tendency towards negative emotions. The arousal score of 3.00 suggests a moderate level of activation or alertness, while the dominance score of 3.10 indicates an average sense of control or influence over the situation.
- **Custom Questionnaire:** The results of the custom questionnaire provide insights into specific aspects of the experience (all rated on a scale of 1 to 7). Overall enjoyment received a moderate rating of 4.44, reflecting a fairly positive experience. In terms of ease of navigation, free movement scored 4.00, while teleportation scored slightly higher at 4.13. However, free movement was still preferred when explicitly asked to chose. Reaching points of interest (POI) received a high ease-of-use rating of 5.19. The feature that allows changing the weather scored 3.31 for ease of use and 3.43 for usefulness. Regarding the snapshot feature, ease of use scored 2.93, indicating some challenges, while its

usefulness was rated moderately at 3.68. The additional comments revealed difficulties in finding the weather and snapshot feature.

- **Correlation:** The analysis revealed several significant correlations between the variables using Spearman's rank test. ATI exhibited a moderate positive correlation with Pragmatic quality (ρ = 0.62, p < 0.01) from the UEQ-S. Prior XR experience was strongly positively correlated with Pragmatic quality (ρ = 0.73, p < 0.01) and showed a strong negative correlation with CSQ (ρ = -0.74, p < 0.01). Age demonstrated a moderate positive correlation with the Realness (ρ = 0.55, p < 0.05) and with the Involvement dimensions of the IPQ (ρ = 0.53, p < 0.033).

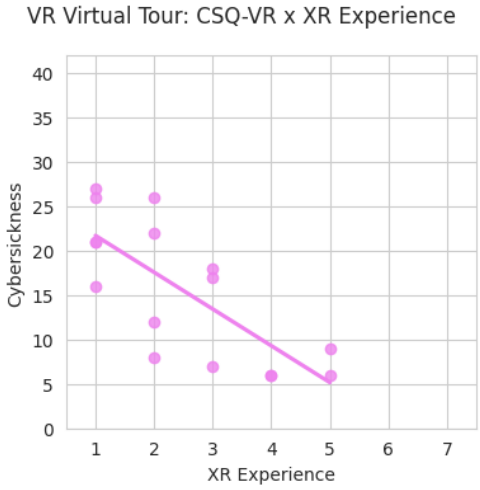

Figure 15: Correlation between prior XR Experience and Cybersickness (N=16).

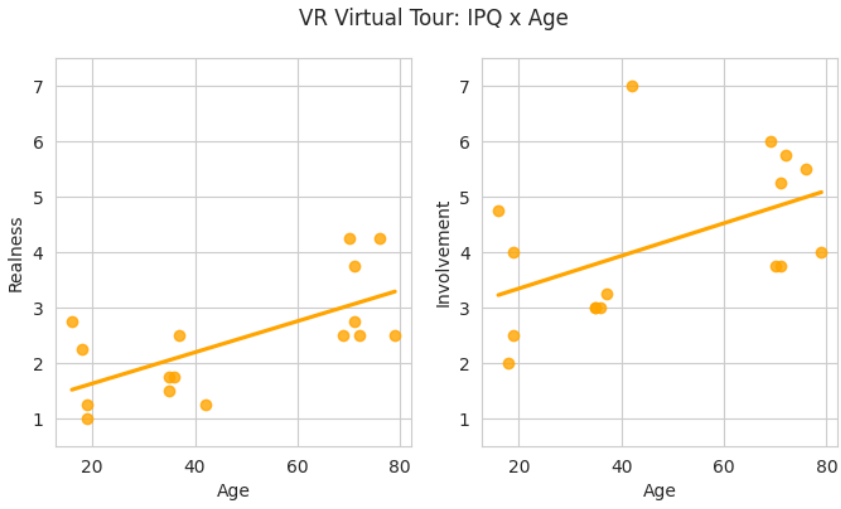

Figure 16: Correlation between IPQ dimensions (Realness, Involvement) and Cybersickness (N=16).

Conclusion

The validation was conducted with a sample group of 16 inhabitants of the village of Etteln with a wide age distribution. Due to a single sample group, the results are merely exploratory without possible comparisons. The results show that participants perceived a moderate feeling of presence in the virtual environment. However, the realness received a low score indicating that the visual quality of the 3D scene was insufficient. Participants reported a low to moderate feeling of cybersickness with further indication of symptoms like dizziness and a fuzzy stomach. The user

experience was rated bad in both hedonic and pragmatic aspects and the emotion metrics show a below-average valence, indicating an unpleasant experience. The custom questionnaire results suggest that participants generally enjoyed the experience, despite the challenges in usability. Free movement was preferred over teleportation as a method of navigation. Qualitative feedback revealed difficulties in finding the features. The correlation analysis shows that technical affinity is correlated with pragmatic quality. Further previous experience with XR seems to improve the pragmatic quality ratings and lowers the probability of getting cybersick. Age correlates with the sense of presence, meaning older participants felt more involved and rated the realness higher.

2.1.4 AR Virtual Tour

Introduction

Objective: The objective of the study was to evaluate the usability of the application. To achieve this, a supervised usability test was selected as the study's methodology. A supervised usability test is a structured evaluation method designed to assess the effectiveness, efficiency, and user satisfaction of a product, system, or interface. In this approach, participants complete a series of predefined tasks in a controlled testing environment under the observation of a facilitator or moderator. The facilitator provides guidance when necessary, records observations, and may ask follow-up questions to better understand the participants' thought processes, challenges, and overall experience.

Experiment Goals: The expected outcome of the study was to identify usability issues, uncover areas for improvement, and gather insights into how users interact with the application. These findings would serve as a foundation for refining the application to better meet user needs and expectations.

Experimental Setup

Environment:

- Community Hall (external/internal)

Hardware and Tools:

- Samsung S23 and S24: These smartphones are flagship models from Samsung, equipped with cutting-edge technology for enhanced performance and usability. Both devices feature high-resolution AMOLED displays, powerful processors(Snapdragon 8 Gen 2 for the S23 and Snapdragon 8 Gen 3 for the S24), and advanced cameras.

Subject Groups

- **Etteln inhabitants:** The sample group is the same as for VR Virtual Tour described in 2.1.3. VR Virtual Tour - Subject Groups.

Experiment Design

Independent Variables:

- **Cognitive Task Load**: Measured using the NASA-TLX (Task Load Index), a widely used tool for assessing perceived workload. Participants rate their experiences on dimensions such as mental demand, physical demand, temporal demand, effort, performance, and frustration. This provides a comprehensive view of task difficulty.
- **User Experience:** Measured using the UEQ-S (User Experience Questionnaire - Short version). This tool assesses the user's perception of the application's usability and aesthetics, focusing on factors like attractiveness, efficiency, dependability, and stimulation.
- **Emotion**: Measured using the Self-Assessment Manikin (SAM), a non-verbal assessment technique. Participants rate their emotional state on dimensions of valence (positive/negative feelings), arousal (calmness/excitement), and dominance (sense of control).
- **Custom metrics:** Measured through a custom made questionnaire. It included a mix of quantitative and qualitative questions aimed at assessing enjoyment, ease of use, and preferences across different features and devices.

Dependent Variables:

- **Gender**: participants' gender is recorded as a categorical variable (e.g., male, female, non-binary, etc.). This helps in identifying potential gender-based differences in task load perception, user experience, and emotional responses.
- **Age**: the age of participants is recorded as a continuous variable or categorized into age groups (e.g., 18–25, 26–35, etc.). This allows for an analysis of how usability, workload, and emotions differ across age demographics.
- **Technological Affinity:** Measured using the Affinity for Technological Interaction (ATI) scale, understanding the participants' technological affinity helps determine whether difficulties in testing applications stem from a general lack of interaction with technical systems or issues specific to the program.

Procedure

Task Execution

Introduction and Overview: Participants were welcomed and given a brief overview of the project and the software. This introduction provided context for the test and familiarized participants with the application's purpose and features.

Explanation of Test Steps: A structured explanation of the test steps was provided to ensure that participants understood the tasks they were expected to perform.

Device Preparation: Each participant was handed a mobile phone with the AR application already started and calibrated. This step ensured a smooth user experience without requiring additional setup.

Viewing the Community Hall in AR: Participants used the phone's screen to view the augmented reality representation of the community hall. A door frame as portal hint was displayed to guide the participants toward the interactive feature.

Opening the Virtual Portal: Participants were encouraged to press the screen to open a virtual portal within the application. This portal provided a transition from the external AR view into the building's interior.

Stepping into the AR Experience: Participants were encouraged to physically move into the projected AR view, allowing them to experience the building's interior as if they were inside. This step tested the immersive quality of the application and its ability to align virtual elements with the real-world movement.

Accessing the Event Calendar: Participants interacted with an info pillar within the AR environment, where they accessed the event calendar. This step demonstrated the application's ability to provide interactive and practical information.

Completion and Feedback Collection: Following the test, participants were introduced to questionnaires designed to gather their feedback on usability, performance, and overall experience. Each participant completed the questionnaire, providing insights for future improvements.

Data Collection and Analysis

- **Demographics**
- **Affinity for Technological Interaction (ATI)**: The ATI provides insights into the interest and experience of subject groups in interacting with technology and technical systems. Using a 9-item questionnaire with 7-point Likert scales an overall ATI score can be calculated.
- **NASA Task Load Index (NASA-TLX):** The NASA-TLX is a widely used, subjective, multidimensional assessment tool designed to evaluate perceived workload during a task. It measures six dimensions:
    - **Mental Demand:** Assesses the mental and perceptual activity required.
    - **Physical Demand:** Evaluates the physical effort needed.
    - **Temporal Demand:** Considers the time pressure felt during the task.
    - **Performance:** Reflects the individual's self-assessment of task success.
    - **Effort:** Measures the amount of work exerted to achieve performance.
    - **Frustration Level:** Captures feelings of irritation, stress, and annoyance.

  Participants rate each dimension, and these ratings are combined to produce an overall workload score. The tool is versatile and has been applied in various fields, including aviation and healthcare. For more information and access to the questionnaire materials, visit the official NASA-TLX website:
- **User Experience Questionnaire - Short Version (UEQ-S):** The UEQ-S is a concise version of the original User Experience Questionnaire, tailored for scenarios where a quick assessment is needed. It evaluates user experience across two primary dimensions:
    - **Pragmatic Quality:** Focuses on usability aspects such as efficiency, clarity, and dependability.
    - **Hedonic Quality:** Addresses the product's ability to stimulate and engage users, including aspects like excitement and innovation.

  The UEQ-S consists of 8 items, each presented as a pair of contrasting adjectives (e.g., "complicated" vs. "easy"). Participants rate their experience on a 7-point scale between these adjectives

- **Self-Assessment Manikin (SAM):** The Self-Assessment Manikin is a non-verbal pictorial assessment technique that measures emotional responses across three dimensions:
    - **Valence (Pleasure):** Ranges from happy to unhappy.
    - **Arousal:** Spans from excited to calm.
    - **Dominance:** Varies from controlled to in control.

  Participants select images that best represent their feelings in each dimension, providing insights into their emotional state during or after interacting with a product or system.
- **Custom made questionnaire:** The questionnaire consisted of the following sections:
    - **Overall Enjoyment:** Participants rated their overall enjoyment of the application on a **7-point Likert scale** (1 = very low enjoyment, 7 = very high enjoyment).
    - **Feature-Specific Usability Ratings:** Participants provided usability ratings for specific features of the application, such as:
        - **Placing a Frame**: Rated on a 7-point scale for ease of use.
        - **Finding Points of Interest (POIs)**: Rated on a 7-point scale for ease of discovery.
        - **Opening and Looking Through the Portal**: Rated on a 7-point scale to assess the intuitiveness and engagement of these features.
    - **Device Preferences:** Participants were asked to express their preferences for the device they would like to use with the application (e.g., smartphones or AR headsets such as the HoloLens).
    - **Open-Ended Feedback:** Participants were invited to share additional comments, providing qualitative insights into their experiences, challenges, and specific issues (e.g., difficulties in finding POI information).

Results

- **Task Load**: The NASA-TLX was evaluated on a scale of 1 to 7, with an overall score of 2.72, indicating a relatively low level of perceived workload for the sample group.
- **User Experience**: The user experience (on a scale from -3 to 3) shows mixed results according to the UEQ benchmark. The overall score of 0.70 and the hedonic quality score of 0.75 are both categorized as “below average”, while the pragmatic quality score of 0.65 is rated as "bad".

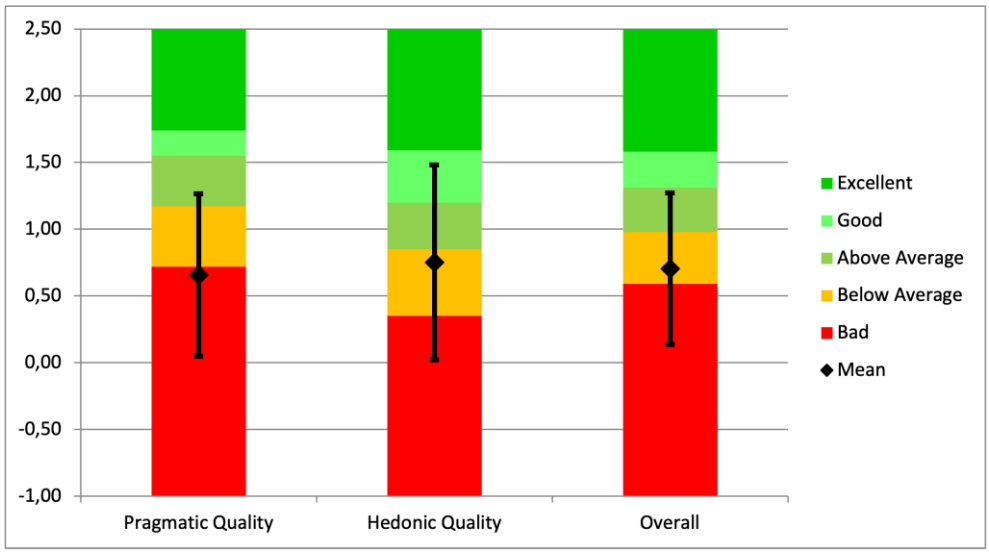

Figure 17: UEQ-S Benchmark for the Etteln inhabitants (N=16) in AR Virtual Tour.

- **Emotion**: The emotional reaction results (on a scale of 1 to 5) indicate mixed responses. The valence score of 4.0 reflects a positive emotional state, while the arousal score of 2.0 suggests a relatively low level of excitement or activation. The dominance score of 3.38 indicates a moderate sense of control over the situation.
- **Custom Questionnaire:** Overall enjoyment (on a scale of 1 to 7) of the application was rated 5.4 which shows a fairly high enjoyment. In terms of ease of use, the feature that allows to place a frame scored 4.56. Finding points of interest (6), opening the portal (6), and looking through the portal (6,18) score high and indicate an easy interaction. Additionally when asked to chose between devices, 4 out of 16 users expressed a preference for of using an AR headset such as the HoloLens. Further comments revealed difficulties in finding the POI information for one participant.
- **Correlation:** Using Spearman's rank test a statistically significant moderate negative correlation was found between the participants age and the pragmatic dimension from the UEQ-S questionnaire ($\rho = -0.499261$, $p < 0.05$).

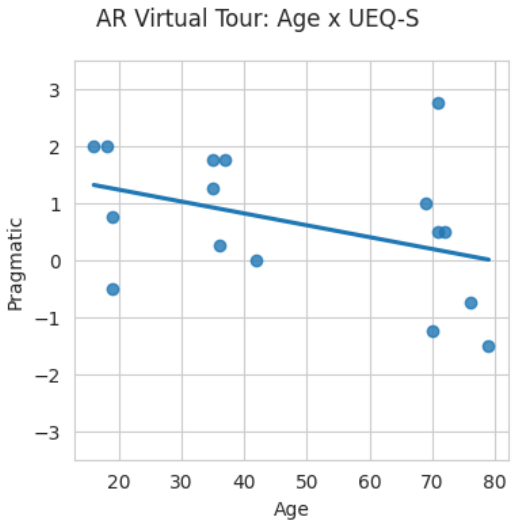

Figure 18: UEQ-S Graph AR Virtual tourism Etteln

Conclusion

The results have again to be handled with care as the validation was conducted with the same sample group as the VR Virtual Tourism in Etteln. The participants reported a low task load, suggesting the application was easy to use. The user experience received a below-average rating for hedonic and a bad rating for pragmatic quality, indicating the appeal was higher compared to its usability. The emotion analysis shows a high valence and low arousal which is equivalent to a calm and pleasant experience. The custom questionnaire confirmed this with a high score for overall enjoyment and the specific features received a high ease of use. A correlation was found between the participant's age and the pragmatic quality, suggesting that older participants had difficulties with the usability of the application.

2.2 Use case validation in Vilanova

At Vilanova only UC03C_ARVT has been tested. During all Q4 from 2024 there had been some validation activities with more than 50 people from different sources. There had been 3 main user testers sources:

Neàpolis ecosystem: The innovation local agency from Vilanova has a 5000 square meters building where more than 25 companies had their offices. At the same time, many activities are also carried out, such as startup acceleration programs or social innovation projects, among other activities. As one of the entity's specialisations are the projects related to the field of smart media, it was possible to engage with different users to carry out the first tests of the application of augmented reality tourism use case.

Tourism and cultural environment: Vilanova has a rich cultural and tourist activity, it is for this reason that there are different entities, public and private, that work in this field. It has been very valuable to have professionals from this environment, since it has allowed the consortium to have feedback from people who have to use this application once its development is over.

Vilanova's big far: La Fira de Novembre, which took place between November 8 and 10 is considered one of the most important commercial multi sectorial outdoors exhibitions in Catalonia, and aims to show the different sectors of the this spanish city economy to a wide audience. The fair had an exhibition area of 6.5 hectares, with 190 exhibitors and more than 135 activities, one of those activities was the UC03C_ARVT user validations.

In the biggest exhibitor in the main square of the city, the team managed to make validations with more than 50 people, chosen accordingly to the different profiles necessary to be able to extract data as realistic as possible.

Figure 19: Validation activity carried out at the fair

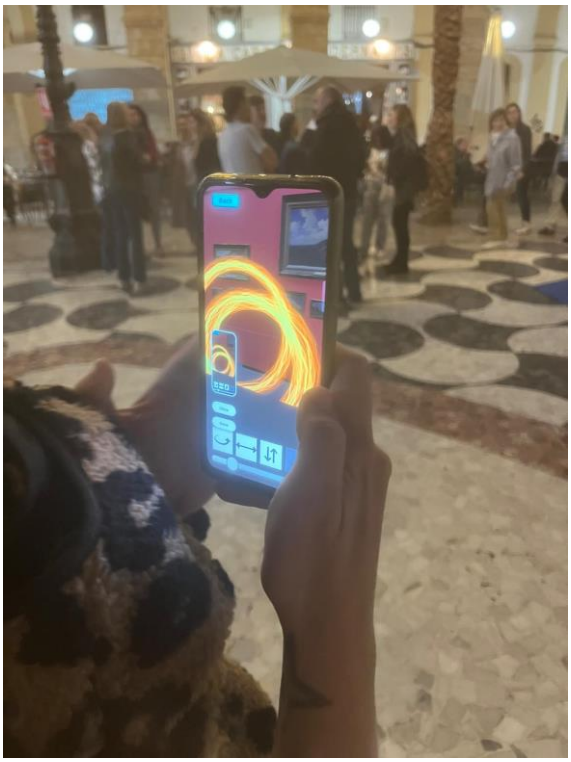

Figure 20: User tester doing a validation

These validations have been carried out in different environments, within Neàpolis, in the vicinity of the museum (the initial application is the recreation of the Victor's Balaguer museum library) and in the main square of the city, *La Plaça de la Vila*. In order to make all the use case validation more interesting, All participants were given a series of rewards for their participation in the activities. These gratifications were, free tickets to the Víctor Balaguer museum as well as exchangeable gift vouchers to Amazon.

Much to our dismay there are 3 use cases which no validation has been carried out since the applications have not yet been developed. Although there are many advances in the creation of them, the necessary applications for use cases: UC03C.VRVT, UC04CMCM and UC.02C.DTC, cannot yet be tested.

Figure 21: 3D render of Vilanova i la Geltrú. Asset needed for constructing the UC_02C_DTC and UC03C_VRVT above all.

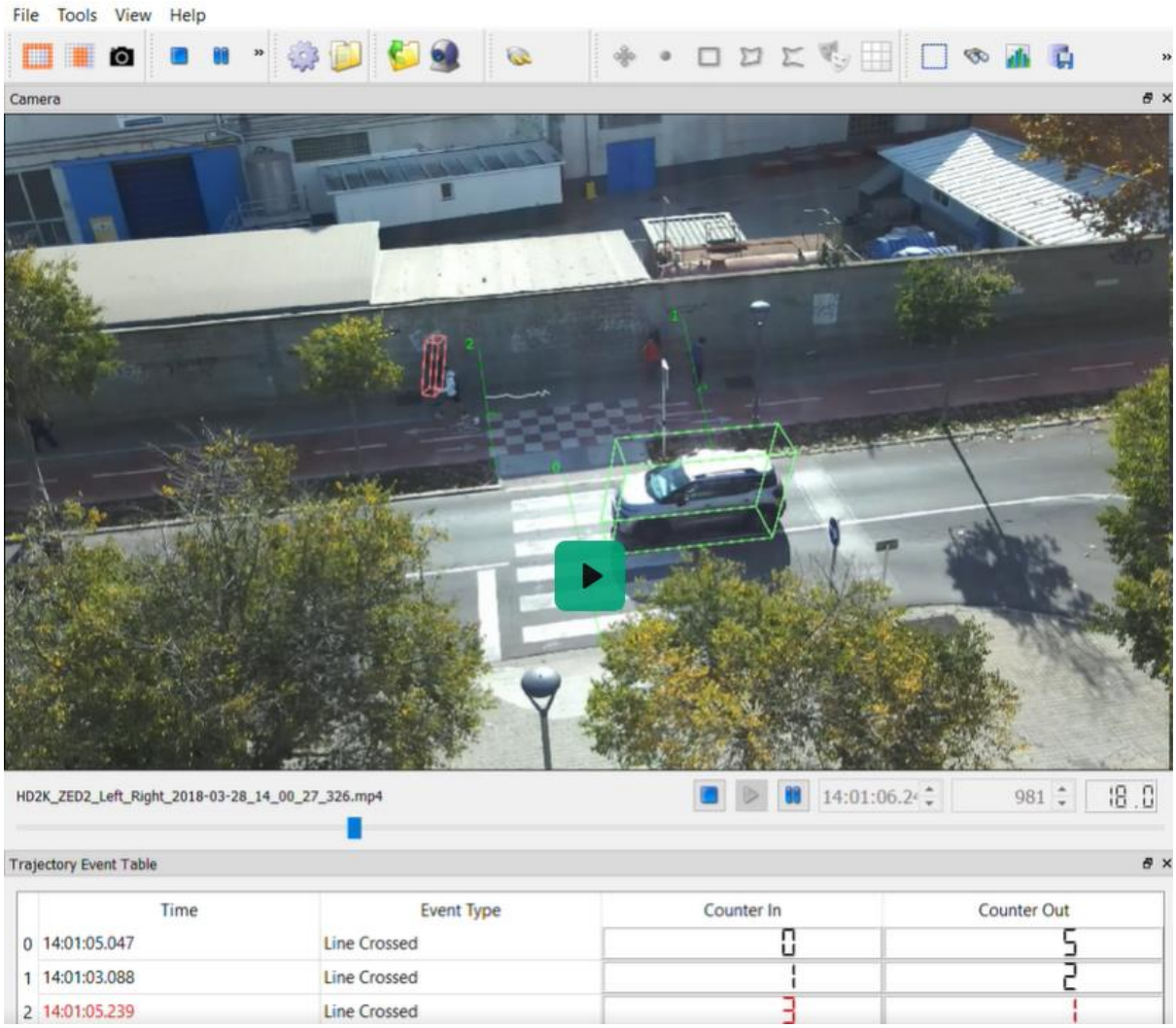

Figure 22: Development of city maintenance use case

One of the priorities of this WP5 is to start making user tester validations once the first versions of the applications are finished, which is estimated to be during the first months of 2025.

2.2.1 AR Virtual Tour

Introduction

Objective: The objective of the study was to evaluate the usability of the application. To achieve this, a supervised usability test was selected as the study's methodology. A supervised usability test is a structured evaluation method designed to assess the effectiveness, efficiency, and user satisfaction of a product, system, or interface. In this approach, participants complete a series of predefined tasks in a controlled testing environment under the observation of a facilitator or moderator. The facilitator provides guidance when necessary, records observations, and may ask follow-up questions to better understand the participants' thought processes, challenges, and overall experience.

Experiment Goals: The expected outcome of the study was to identify usability issues, uncover areas for improvement, and gather insights into how users interact with the application. These findings would serve as a foundation for refining the application to better meet user needs and expectations.

Experimental Setup

Environment:

- **Neàpolis Headquarters**: Controlled environment for usability testing.
- **Vicinity of the Victor Balaguer Museum**: Focusing on the recreation of the museum's library in the application.
- **Main Square of the City (La Plaça de la Vila)**: A central and public location for broader user interaction and validation.

Hardware and Tools:

- **Smartphones**: Modern smartphones with Android and iOS operating systems were used for testing. These devices featured high-resolution screens, fast processors, and sufficient storage to ensure smooth performance of the application. Samsung Galaxy one, Realme 5

Subject Groups:

- **Vilanova Fair:** The fair group consisted of 50 participants, with an average age of 37.76 years and an age range spanning from 18 to 67 years. The group included 30 male participants and 20 female participants. The reported experience with XR level was 3.54 on a scale from 1 to 7, indicating a moderate familiarity with the technology or domain. Additionally, the group's Affinity for Technology Interaction (ATI) score was 4.14, reflecting a fairly high interest and ease in interacting with technology. Regarding prior experience with similar technologies, 25 participants had tried AR on smartphones, 27 participants had tried VR on smartphones, 24 participants had used VR headsets, and 11 participants had used VR headsets.

- **Vilanova:** The size of the Vilanova group is 18 participants with an average age of 42.72. Among them 6 males and 12 females. Their mean experience with XR (scale 1 to 7) was 3.61 and their ATI score 3.85. In terms of prior experience with similar technologies, 10 participants had previously used AR on smartphones, 11 participants had tried VR on smartphones, 7 participants had experience with AR headsets, and 12 participants had used VR headsets.

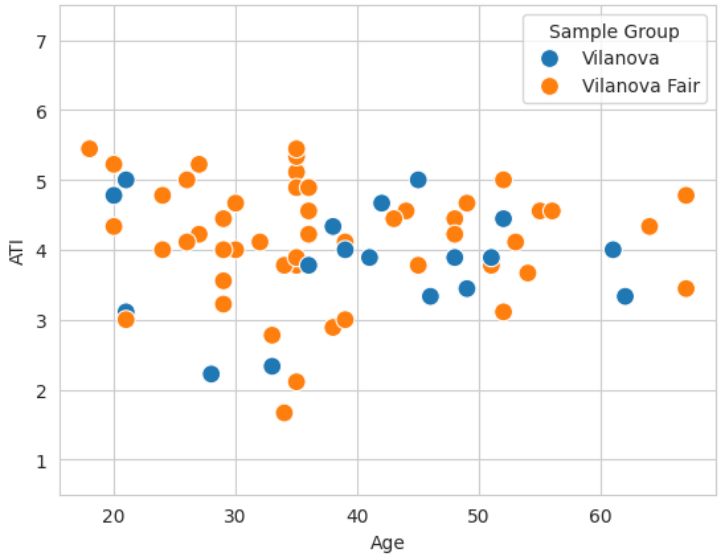

Figure 23: ATI scores and age for both subject groups (N=68)

Experiment Design

Independent Variables:

- **Cognitive Task Load**: Measured using the NASA-TLX (Task Load Index), a widely used tool for assessing perceived workload. Participants rate their experiences on dimensions such as mental demand, physical demand, temporal demand, effort, performance, and frustration. This provides a comprehensive view of task difficulty.
- **User Experience:** Measured using the UEQ-S (User Experience Questionnaire - Short version). This tool assesses the user's perception of the application's usability and aesthetics, focusing on factors like attractiveness, efficiency, dependability, and stimulation.
- **Emotion**: Measured using the Self-Assessment Manikin (SAM), a non-verbal assessment technique. Participants rate their emotional state on dimensions of valence (positive/negative feelings), arousal (calmness/excitement), and dominance (sense of control).

Dependent Variables:

- **Gender**: participants' gender is recorded as a categorical variable (e.g., male, female, non-binary, etc.). This helps in identifying potential gender-based differences in task load perception, user experience, and emotional responses.
- **Age**: the age of participants is recorded as a continuous variable or categorized into age groups (e.g., 18–25, 26–35, etc.). This allows for an analysis of how usability, workload, and emotions differ across age demographics.

- **Technological Affinity:** Measured using the Affinity for Technological Interaction (ATI) scale, understanding the participants' technological affinity helps determine whether difficulties in testing applications stem from a general lack of interaction with technical systems or issues specific to the program.

Procedure

Also here you can place text describing the entire experiment execution. From when the participants ere welcome until the filled in the last questionnaire and receive the compensation (you can also avoid the previous bullets, they were a suggestion...).

Data Collection and Analysis

- **Demographics**
- **Affinity for Technological Interaction (ATI)**: The ATI provides insights into the interest and experience of subject groups in interacting with technology and technical systems. Using a 9-item questionnaire with 7-point Likert scales an overall ATI score can be calculated.
- **NASA Task Load Index (NASA-TLX):** The NASA-TLX is a widely used, subjective, multidimensional assessment tool designed to evaluate perceived workload during a task. It measures six dimensions:
    - **Mental Demand:** Assesses the mental and perceptual activity required.
    - **Physical Demand:** Evaluates the physical effort needed.
    - **Temporal Demand:** Considers the time pressure felt during the task.
    - **Performance:** Reflects the individual's self-assessment of task success.
    - **Effort:** Measures the amount of work exerted to achieve performance.
    - **Frustration Level:** Captures feelings of irritation, stress, and annoyance.

  Participants rate each dimension, and these ratings are combined to produce an overall workload score. The tool is versatile and has been applied in various fields, including aviation and healthcare. For more information and access to the questionnaire materials, visit the official NASA-TLX website.
- **User Experience Questionnaire - Short Version (UEQ-S):** The UEQ-S is a concise version of the original User Experience Questionnaire, tailored for scenarios where a quick assessment is needed. It evaluates user experience across two primary dimensions:
    - **Pragmatic Quality:** Focuses on usability aspects such as efficiency, clarity, and dependability.
    - **Hedonic Quality:** Addresses the product's ability to stimulate and engage users, including aspects like excitement and innovation.

  The UEQ-S consists of 8 items, each presented as a pair of contrasting adjectives (e.g., "complicated" vs. "easy"). Participants rate their experience on a 7-point scale between these adjectives.
- **Self-Assessment Manikin (SAM):** The Self-Assessment Manikin is a non-verbal pictorial assessment technique that measures emotional responses across three dimensions:
    - **Valence (Pleasure):** Ranges from happy to unhappy.
    - **Arousal:** Spans from excited to calm.
    - **Dominance:** Varies from controlled to in control.

Participants select images that best represent their feelings in each dimension, providing insights into their emotional state during or after interacting with a product or system. For a comprehensive overview and access to the SAM tool, you can refer to:

Results

- **Task Load:** The NASA-TLX (Task Load Index), measured on a scale from 1 to 7, was used to evaluate the perceived workload of two groups: Vilanova Fair and Vilanova. The Vilanova Fair group reported a workload score of 2.81, while the Vilanova group scored slightly lower at 2.5. These results indicate a generally low perceived workload for both groups.
- **User Experience:** The UEQ-S results (measured on a scale from -3 to 3) for the Vilanova Fair and Vilanova groups reveal some differences across three key aspects: overall experience, hedonic quality, and pragmatic quality. The Vilanova group scored 1.11 for the overall experience, 1.05 for hedonic quality, and 1.16 for pragmatic quality. In comparison, the Vilanova Fair group scored 1.41 for the overall experience, 1.46 for hedonic quality, and 1.36 for pragmatic quality. These results suggest a generally positive experience for both groups, with Vilanova Fair showing slightly higher scores across all categories. Comparing the results to the UEQ benchmark, the Vilanova group is in the “above average” range and the Vilanova group in the “good” catergory.

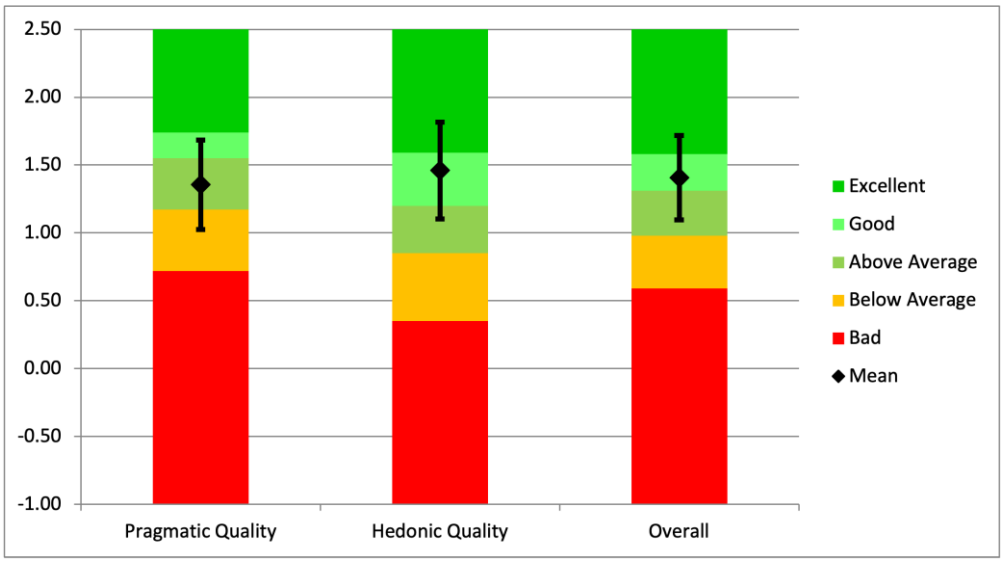

Figure 24: UEQ-S Benchmark for the Vilanova Fair Group (N=50) in AR Virtual Tour.

- **Emotion:** The emotional reaction results, measured on a scale from 1 to 5, show some differences between the Vilanova Fair and Vilanova groups. For valence, the Vilanova Fair group scored 3.89, indicating a generally positive emotional state, while the Vilanova group scored 4.3, suggesting an even more positive emotional experience. In terms of arousal, the Vilanova Fair group had a score of 2.33, reflecting a moderate level of excitement, while the Vilanova group scored 1.66, indicating a lower level of arousal. Both groups showed similar levels of dominance, with the Vilanova Fair group scoring 3.38 and the Vilanova group scoring 3.34, suggesting a moderate sense of control over the experience for both groups.

- **Qualitative Feedback:** Participants provided qualitative feedback highlighting several areas for improvement in the application. Two participants mentioned the missing ability to move within the app, while the lack of information about the art pieces was noted four times. The inability to zoom within the application was mentioned twice. Additionally, one participant noted the absence of audio, and another commented on the light not being captured well. Two participants reported an issue where a statue appeared to be cut in half.
- **Correlation:** Using spearman's rank test, two statistically significant correlations were found with the combined data from both sample groups. The participants ATI score correlates low to moderate positive with the overall user experience from the UEQ-S ($\rho = 0.28$, $p < 0.05$) and with the hedonic dimension ($\rho = 0.30$, $p < 0.05$).

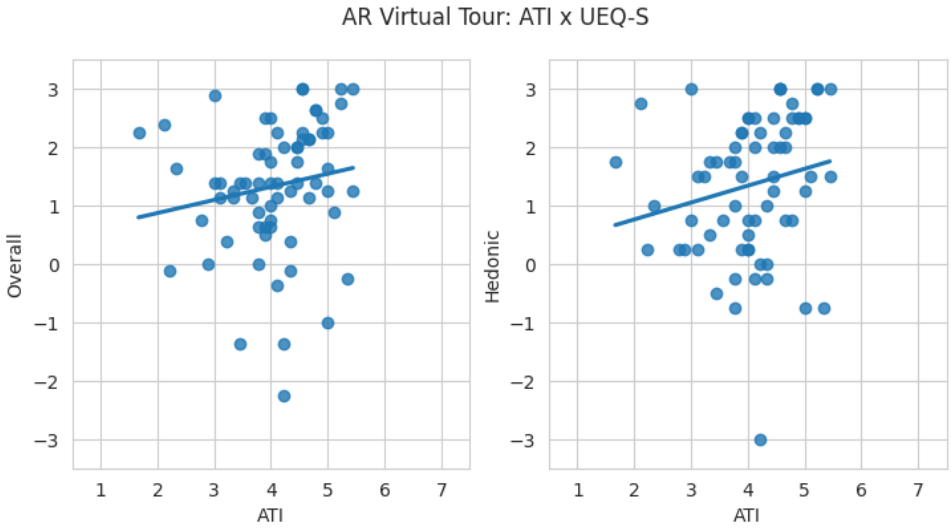

Figure 25: Correlation between UEQ-S dimensions (Overall, Hedonic) and ATI in the aggregated sample (N=68).

Conclusion

The validation in Vilanova was conducted with two sample groups, comprising a total of 68 participants. The group from the Vilanova fair included the largest number of participants, with 50 individuals, thus offering greater reliability in the results. Both groups reported low perceived workload, with the Vilanova fair group slightly higher. The user experience was above average for both groups, with slightly higher scores for the Vilanova fair group, reaching the good quality benchmark. Emotional responses were also positive, with the Vilanova group reporting even slightly higher valence and lower arousal than the Vilanova Fair group. However, qualitative feedback highlights areas for improvement, including the wish to navigate the application, insufficient information on art pieces, issues with zooming, and the lack of audio. The significant positive correlations between participants' ATI scores and the user experience, both overall and pragmatic, suggest that increased affinity for technology may enhance the experience specifically in terms of usability.

2.3 Use case validation at Ideal Works

The objective of this experiment is to compare the effectiveness and efficiency of placing Points of Interest (POIs) using a 2D webpage-based map through AnyFleet versus a 3D map viewed through HoloLens, and to evaluate the cognitive workload associated with each interface using the NASA Task Load Index (NASA-TLX).

Experiment Goals

The experiment aims to achieve the following goals:

- **Accuracy Measurement:** Evaluate the precision of POI placement in both 2D and 3D interfaces by measuring deviation from the ground truth coordinates.
- **Efficiency Analysis:** Compare the time taken to complete POI placement tasks using both interfaces.
- **Cognitive Load Assessment:** Measure the cognitive workload associated with each interface using the NASA-TLX to identify user mental effort, physical demand, and overall task difficulty.

Experimental Setup

Environment

- **Location**: IDEALworks warehouse
- **Ground Truth System:** Vicon motion capture system for precise pose tracking of POIs.

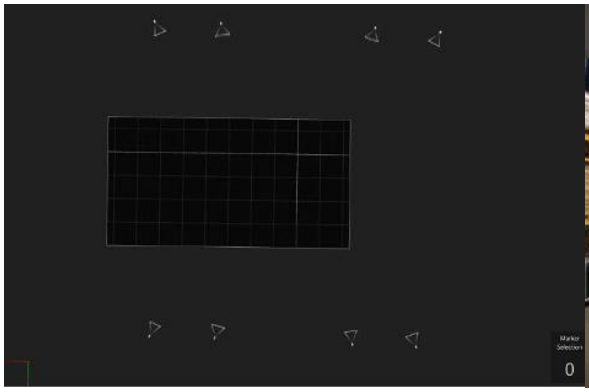

Figure 23 Vicon Layout

Figure 247: Warehouse with the Vicon System Setup

Hardware and Tools

- **2D Mapping Tool**: AnyFleet which is a Webpage-based map accessed via computer.
- **3D Mapping Tool**: Microsoft HoloLens
- **Vicon Tracking System:** Ground-truth Object Tracking System

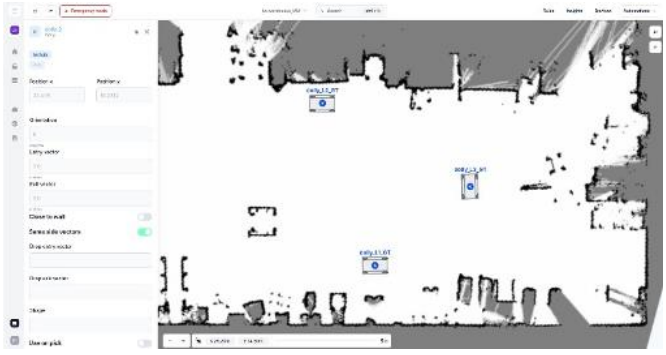

Figure 25 AnyFleet Interface

Figure 29: HoloLens

Figure 30:Vicon Tracking System

Control Variables

1. Room/environment layout
2. Number and complexity of POIs to be placed
3. Task instructions and briefing
4. Hardware specifications (computer for 2D method, HoloLens 2)

Subject Groups

- Experience Level:
  - Experienced Users: Prior experience with 2D/3D mapping tools
  - Inexperienced Users: No prior experience
- Diversity Considerations:
  - Gender diversity
  - Variety of age and professional background

Experiment Design

Independent Variables

- **Interface Type:**
  - 2D Interface: Webpage-based map
  - 3D Interface: HoloLens
- **Task Complexity:**
  - Low Complexity: Simple tasks involving placing POIs in easily identifiable locations.
  - High Complexity: Complex tasks requiring precise placement of POIs in difficult-to- identify locations

Dependent Variables

- **Accuracy of POI Placement:**
  - Measure the deviation between the intended and actual placement of the POI.
- **Time Taken to Complete the Task:**
  - Measure the time taken to complete each task until the user completes placing all POIs.

- **Cognitive Load:**
  - Assessed through the NASA-TLX.

Description of Task Complexity Levels

1. Low Complexity:
   - Description: The dolly is positioned near objects with distinctive geometry, such as the grey metallic structure and a red fire extinguisher. These objects serve as clear reference points for easy placement.
   - Reasoning: The identifiable shapes (fire extinguisher, mounted equipment, and metallic frames) make the placement straightforward and reduce the chances of misalignment.

Figure 31: Dolly Near Distinct Objects for Easy Alignment

2. Moderate Complexity:
   - Description: The dolly is placed near wooden crates and industrial shelving. The crates and shelves provide linear edges and horizontal walls, but aligning the dolly precisely requires more effort due to the complexity of the environment.
   - Reasoning: While there are reference edges (the crates and shelving), the multiple horizontal lines and cluttered space make precise placement slightly more challenging.

Figure 32: Dolly by Crates and Shelves for Moderate Alignment

3. High Complexity:
   - Description: The dolly is in an open warehouse area with no immediate or distinct objects for alignment.
   - Reasoning: The lack of nearby reference points makes accurate placement more difficult, as there are no clear geometrical markers or boundaries close to the dolly.

Figure 33: Dolly In Open Space With No Clear References

Procedure

Training Phase:

Provide a brief training session for both interfaces to all participants to minimize the learning curve effect.

Task Execution Order:

Interface Randomization:

- Participants will be randomly assigned to start with either the AnyFleet or the HoloLens interface, by choosing a random ID from a bowl.
- The sequence of tasks will be varied, and the order of task complexity (low (1), moderate (2), and high (3)) will be randomized across participants to reduce the effect of task familiarity and learning bias.

Task Execution:

- Level 1: Dolly placed near distinct objects for easy alignment. (Low Complexity)
- Level 2: Dolly positioned near shelving, using horizontal edges as placement references. (Moderate Complexity)
- Level 3: Dolly in open warehouse space with minimal nearby reference points for placement. (High Complexity)

Figure 34: Experimental Setup

Data Collection and Analysis

Data Collection from Participants' Input

To ensure accurate and systematic data collection, a custom application was developed to streamline the process. This application captures all necessary participant data, task inputs, and transformation matrix details, allowing for efficient logging and exporting of information.

Participant Details and Interface Selection

The data collection application records essential information such as:

1. Participant ID
2. Date of Participation
3. Selected Interface (AnyFleet or HoloLens)

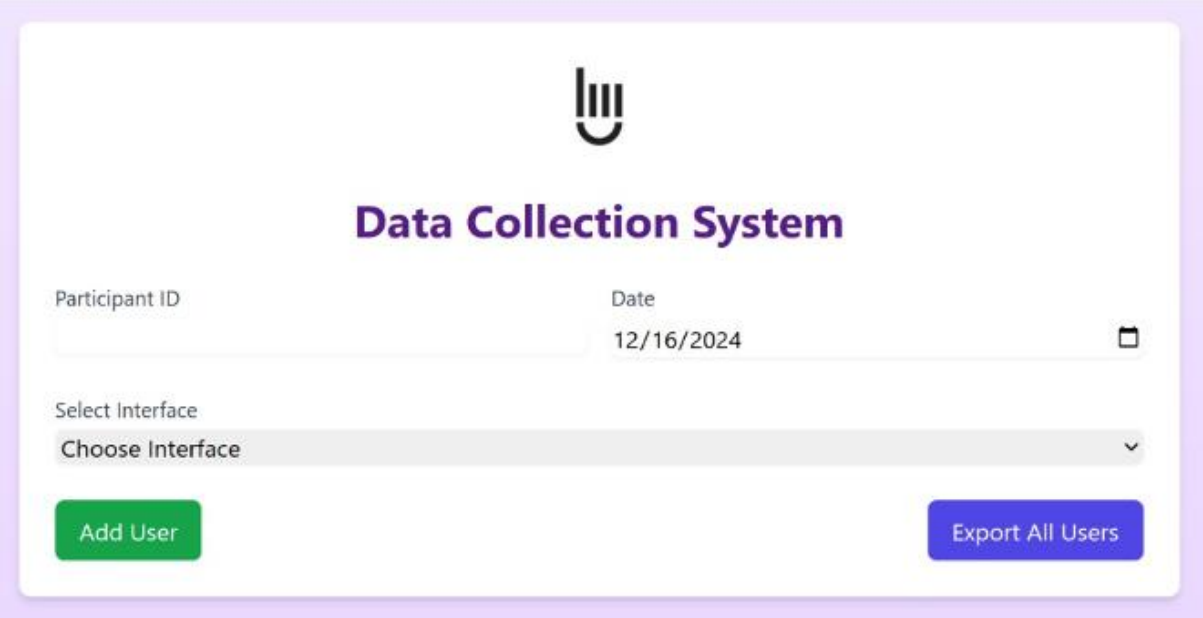

Figura 35: 26 Data Collection App

Task Execution and POI Input

For each participant and interface, task execution data is recorded, including:

1. Experience Level (using a slider to gauge familiarity with the interface)
2. Task Level Order: Level 1, 2, and 3
3. Task Completion Time: Recorded in seconds
4. POI Placement Coordinates: 2D for AnyFleet and 3D for HoloLens

This interface provides input fields for multiple tasks, ensuring flexibility to record data across different task complexities.

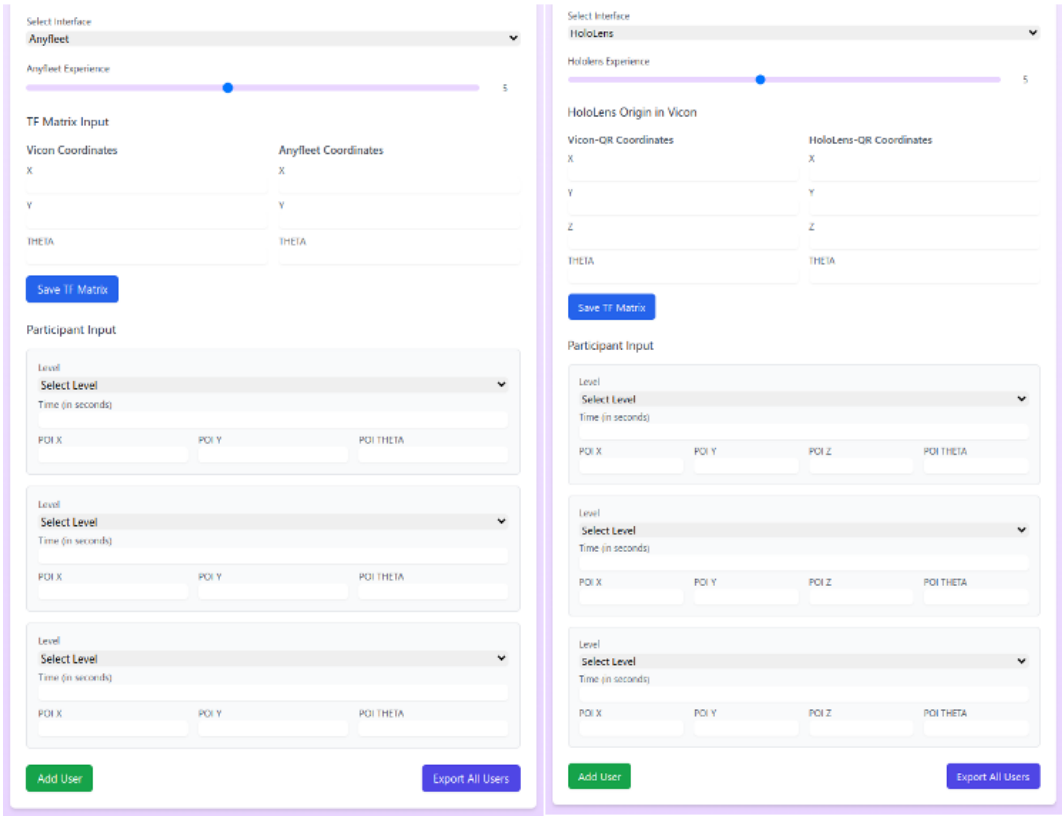

Figure 36: Data Collection System interface for entering task-level details, completion time, and POI placement coordinates

Preparation for Transformation Between Planes Calculation

The application facilitates the collection of transformation matrix inputs needed for aligning the recorded coordinates with the Vicon ground truth system. This preprocessing step ensures accurate transformation between coordinate planes.

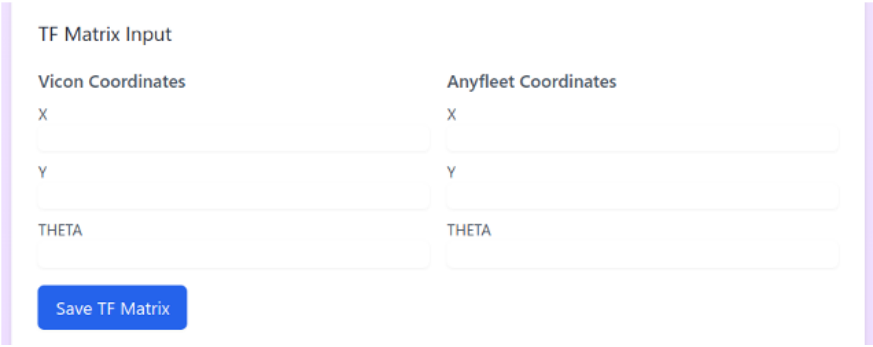

Figure 37: Coordinates for Transformation Calculation

Transformation Calculations

HoloLens to Vicon Transformation

1. **Setup :**
   - Attach the QR code to a clearly visible surface within the Vicon system's capture area.
   - Carefully install reflectors around the QR code's borders to define its origin point in the Vicon system.

- Use the Vicon system to record the pose of the QR code.
- Launch the HoloLens application and allow the user to scan the QR code.
- Capture the QR code's pose relative to the HoloLens system.
- Calculate the transformation matrix by comparing the recorded poses from the Vicon system and the HoloLens.

2. **Transformation Matrix Calculation:**
Given:
   - Vicon System (V):
     i. Coordinates: (xV, yV, zV))
     ii. Orientation: (θVz), representing rotations around the Z axis.
   - HoloLens System (H):
     iii. Coordinates: (xH, yH, zH)
     iv. Orientation: (θHz), representing rotations around the Z axis.

Step 1: Convert units if needed between the 2 systems

Step 2: Calculate the Translation Vector

The translation vector $\vec{T}\rightarrow$ is the difference between the corresponding coordinates of the point in each system:
$\vec{T}\rightarrow = (xV - xH, yv - yH, zv - zH)$

Step 3: Determine Rotation

1. Calculate the Rotation Differences. In 3D, the orientation difference is represented by the relative rotation angles around the Z axis:
$\Delta\theta z = \theta Vz - \theta Hz$
2. Compute Individual Rotation Matrices. For each axis, calculate the rotation matrices:
   a. Rotation around the Z-axis ($\Delta\theta z$)

$$M = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Step 4: Combine Translation and Rotation into Full Transformation Matrix Combine the rotation and translation into a single transformation matrix M using homogeneous coordinates:

$$R_Z = \begin{bmatrix} \cos(\Delta\theta_z) & -\sin(\Delta\theta_z) & 0 \\ \sin(\Delta\theta_z) & \cos(\Delta\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Step 5: Applying the Transformation

- Point Transformation: $\boldsymbol{Pv = M \,.\, PH}$
- Orientation Transformation : $\theta Vz = \theta Hz + \Delta\theta z$

AnyFleet to Vicon Transformation

1. **Setup:**
   - Direct the AnyFleet-controlled robot to a predefined POI within the warehouse.
   - Equip the robot with reflectors to be visible with the Vicon system and set its origin confounded over its center in AnyFleet.
   - Simultaneously record the robot's endpoint as per AnyFleet's system and the actual physical position as captured by the Vicon system.
   - Derive the transformation matrix by comparing these recorded positions, accounting for any deviations or errors between the perceived endpoint by AnyFleet and the actual position recorded by Vicon.
   - This transformation is the same for each setup.

Figure 38: AnyFleet Predefined POIs

2. **AnyFleet to Vicon Transforation:**
   Given:
   - Vicon System (V):
     i. Coordinates: (xV, yV)
     ii. Orientation around Z-axis: θV
   - AnyFleet System (A):
     i. Coordinates: (xA, yA)
     ii. Orientation around Z-axis: θA

   Step 1: Convert Units if needed between the 2 systems

   Step 2: Calculate the Translation Vector
   The translation vector $\vec{T}\rightarrow$ is the difference between the corresponding coordinates of the point in each system: $\vec{T}\rightarrow = (xV - xA, yv - yA)$

   Step 3: Determine Rotation
   Calculate the difference in orientation around the Z-axis between the two systems: $\Delta\theta = \theta V - \theta A$
   The rotation matrix R for a rotation around the Z-axis by Δθ is:

$$R = \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) \\ \sin(\Delta\theta) & \cos(\Delta\theta) \end{bmatrix}$$

Step 4: Combine Translation and Rotation into Full Transformation Matrix
Combine the rotation and translation into a single transformation matrix M using homogeneous coordinates:

$$M = \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & x_V - x_A \\ \sin(\Delta\theta) & \cos(\Delta\theta) & y_v - y_A \\ 0 & 0 & 1 \end{bmatrix}$$

Step 5: Applying the Transformation
This matrix M can be used to transform any point from System A to System V. For any point PA in System A:

$$P_{vicon} = M \cdot \begin{bmatrix} x_A \\ y_A \\ 1 \end{bmatrix} = \begin{bmatrix} \cos(\Delta\theta).x_A - \sin(\Delta\theta).y_A + (x_V - x_A) \\ \sin(\Delta\theta).x_A + \cos(\Delta\theta).y_A + (y_v - y_A) \\ 1 \end{bmatrix}$$

$$\theta_{vicon} = \theta_A + \Delta\theta$$

Error Calculation

Accurate error analysis is critical for evaluating the effectiveness of POI placement using the AnyFleet (2D) and HoloLens (3D) interfaces. Errors are calculated by comparing the recorded POI placements to the ground truth coordinates provided by the Vicon motion capture system.

Euclidean Distance Error

The deviation between the recorded POI placement and the ground truth is calculated using the **Euclidean distance** formula:
$Epos$ = √($xmeasured$ − $xground$)^2 + ($ymeasured$ − $yground$)^2 + ($zmeasured$ − $zground$)^2

Where:

- $xmeasured$, $ymeasured$, $zmeasured$ are the coordinates of the POI placement recorded by the participant.
- $xground$, $yground$, $zground$ are the ground truth coordinates provided by the Vicon system.

Orientation Error (Theta) Calculation

The error in orientation (Theta) measures the difference in the angle between the recorded and ground truth orientations. Theta error is expressed in degrees: $E\theta$ = |$\theta measured$ − $\theta ground$|

Where:

- $\theta measured$ is the orientation recorded by the participant.
- $\theta ground$ is the ground truth orientation provided by the Vicon system.
- If $E\theta$> 180∘, use 360∘− $E\theta$ to get the minimal angle difference.

Statistical Measures

To analyze overall performance, the following statistics are computed for both pose and orientation errors:

1. Mean Error (μ):
   - Represents the average deviation.
2. Standard Deviation (SD):
   - Measures the variability or spread of errors.

Percentage Error Calculation

Percentage Error Per Level:

$Percentage\ Error$ = ($Mean/Maximum\ Possible\ Error$) × 100

Time Tracking

Task execution time is a key metric in assessing the efficiency of each interface. The time taken for each task is recorded using a stopwatch to ensure precise and consistent timing across all participants.

NASA-TLX Survey

The NASA Task Load Index (NASA-TLX) survey is used to measure the cognitive workload experienced by participants when performing tasks with each interface (AnyFleet or HoloLens). The survey assesses six dimensions of workload, providing insight into the mental and physical demands of each interface.

Six Dimensions of NASA-TLX

1. **Mental Demand:** How mentally challenging was the task?
2. **Physical Demand:** How physically demanding was the task?
3. **Temporal Demand:** How hurried or rushed did the task make you feel?
4. **Performance**: How successful did you feel in accomplishing the task?
5. **Effort**: How much effort did you put into completing the task?
6. **Frustration**: How irritated, stressed, or annoyed were you during the task?

Statistics to be Calculated

1. **Average (Mean):**
   - **Equation**:
     $Mean(\overline{x})$ = $Sum\ of\ all\ participant\ scores\ for\ a\ dimension/\ Total\ number\ of\ participants\ (N)$
   - **Purpose**: Represents the central tendency or overall perception of workload for a given dimension across all participants for each interface.
   - **Interpretation**: Higher mean scores indicate greater workload for that dimension.

2. **Standard Deviation (SD):**
   - **Equation**:
     $SD = \sqrt{Sum\ of\ squared\ difference\ between\ each\ score\ and\ the\ mean / Total\ number\ of\ participants(N) - 1}$
   - **Purpose**: Measures the variability or spread of participant ratings for a given dimension across participants for each interface.
   - **Interpretation**: A low SD indicates consistent experience, while a high SD suggests variability in workload perception for the hardware interface.
3. **Comparative Analysis:**
   - Compare mean scores and standard deviations between the AnyFleet (2D interface) and HoloLens (3D interface) for each workload dimension.
   - Identify which interface imposes a higher cognitive load based on these comparisons.
4. **Visual Representations:**
   - Bar Charts and Box Plots to visually compare the scores for each dimension across interfaces.
   - Highlight significant differences and trends in workload perceptions

3. Results

3.1 Demographics of the Study

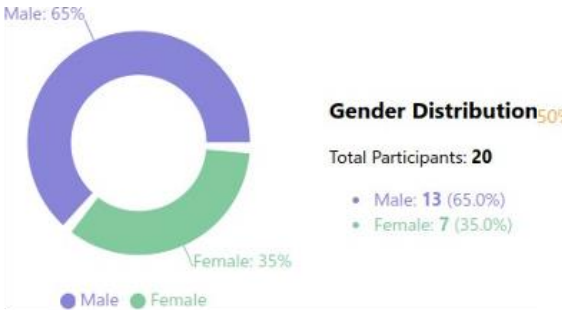

Figure 39: Gender Distribution

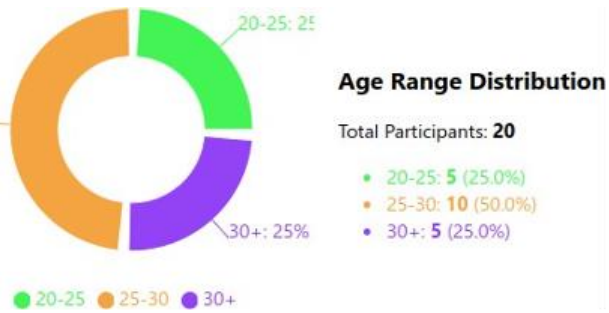

Figure 40: Age Range Distribution

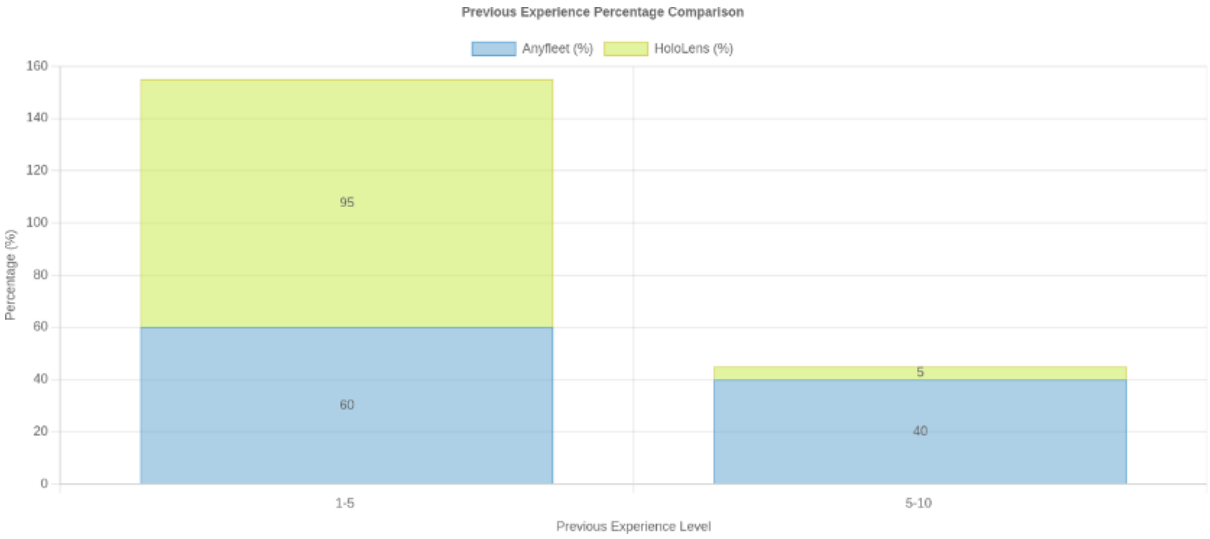

Figure 41: Previous Experience Percentage Comparison Between AnyFleet and HoloLens

3.2 Error Calculation

AnyFleet Distance Error

Figure 42: AnyFleet Distance Error in Level 1(meters)

Figure 43: AnyFleet Distance Error in Level 2(meters)

Figure 44: AnyFleet Distance Error in Level 3(meters)

AnyFleet Orientation Error

Figure 45: AnyFleet Orientation Error in Level 1

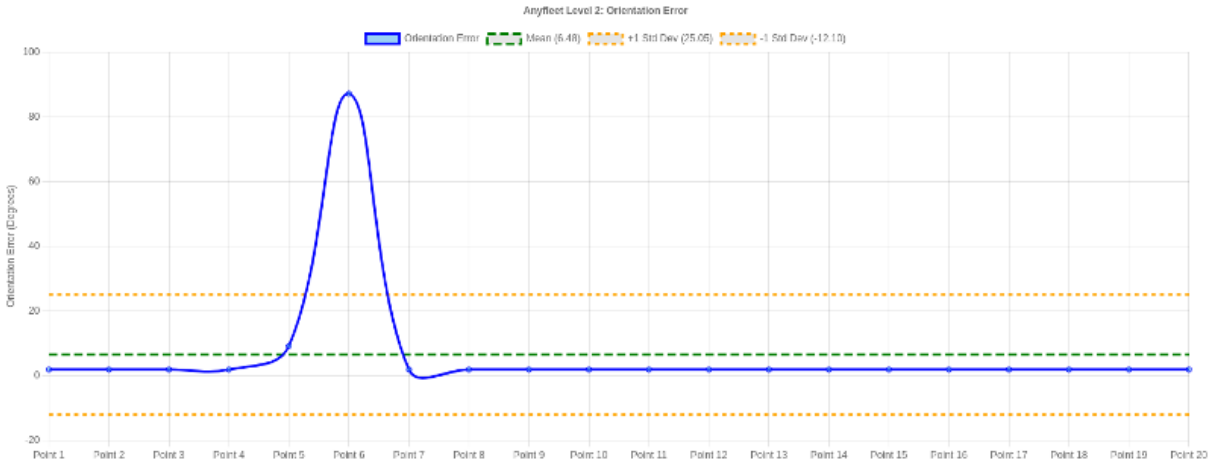

Figure 46: AnyFleet Orientation Error in Level 2

Figure 47: AnyFleet Orientation Error in Level 3

HoloLens Distance Error

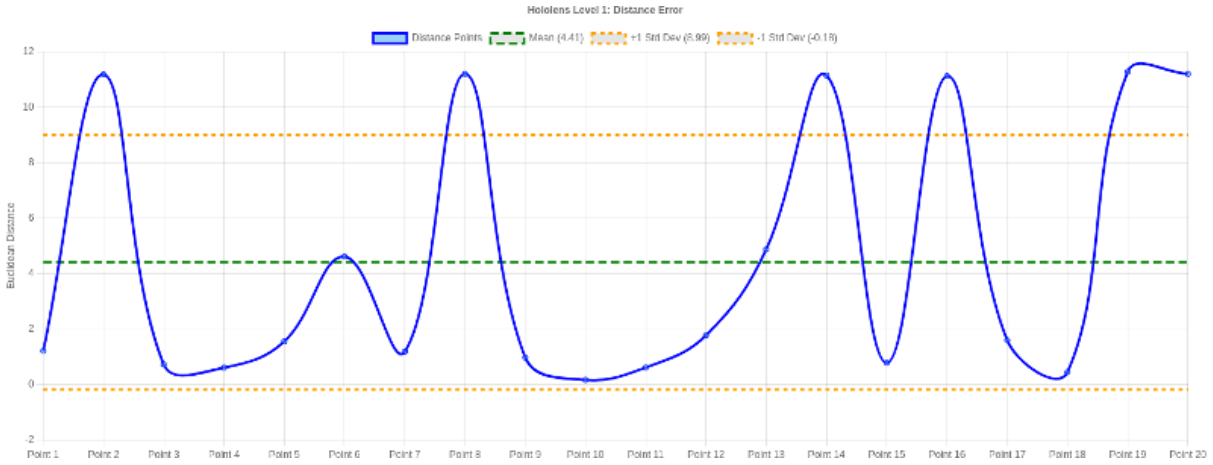

Figure 48: HoloLens Distance Error in Level 1 (meters)

Figure 49: HoloLens Distance Error in Level 2 (meters)

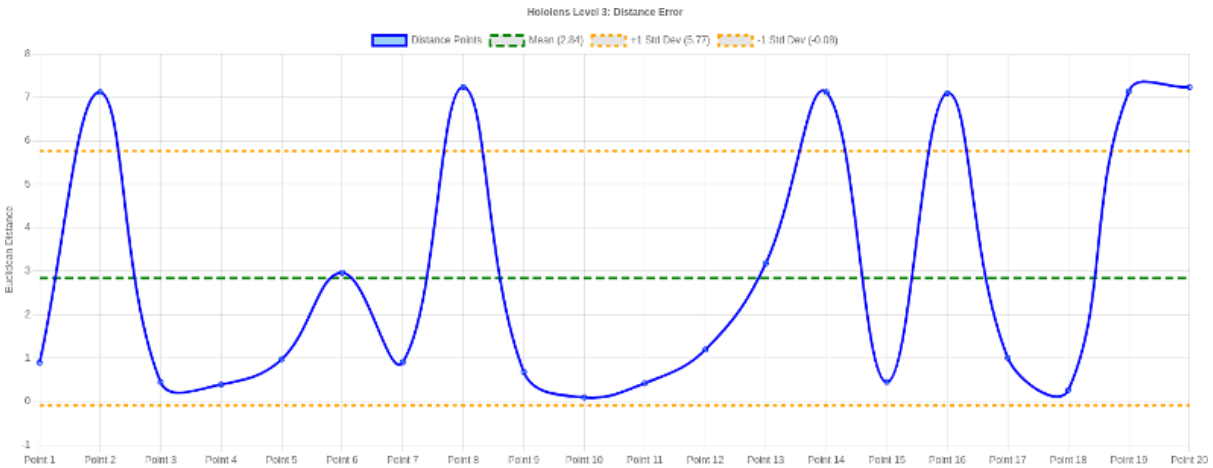

Figure 50: HoloLens Distance Error in Level 3 (meters)

HoloLens Orientation Error

Figure 51: HoloLens Orientation Error in Level 1

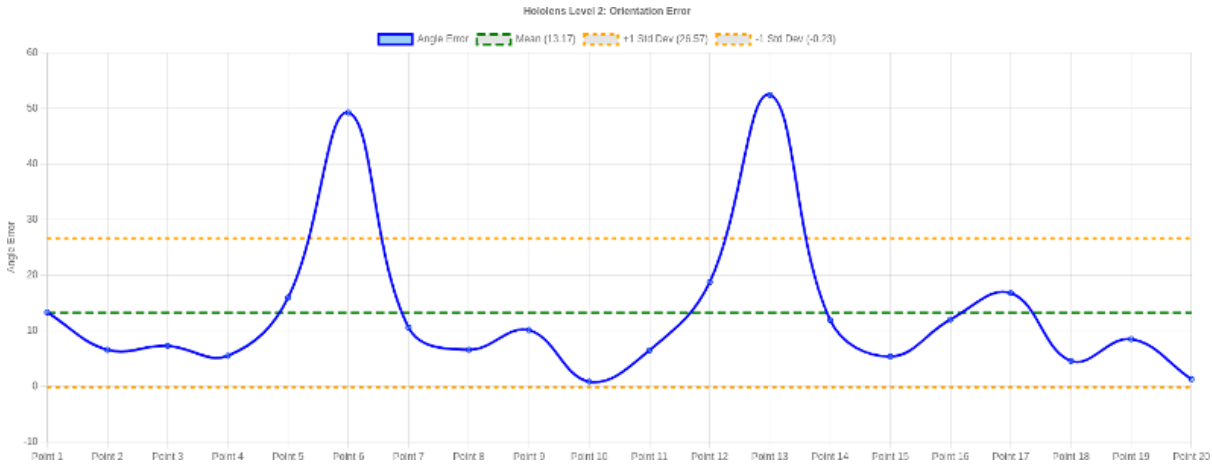

Figure 52: HoloLens Orientation Error in Level 2

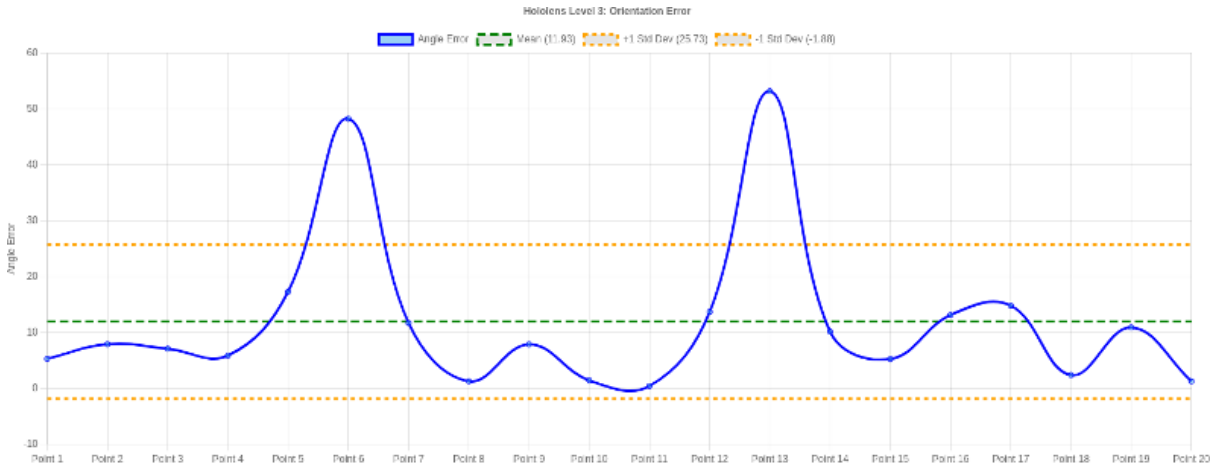

Figure 27: HoloLens Orientation Error in Level 3

Percentage Error in AnyFleet and HoloLens

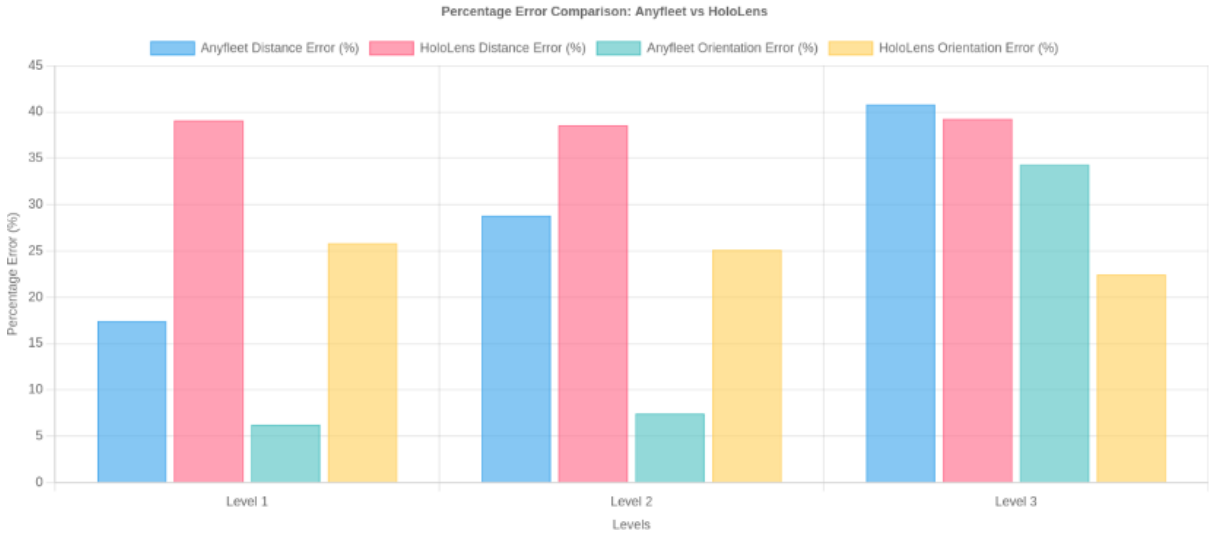

Figure 54: Percentage Error Comparison in Position and Orientation for AnyFleet and HoloLens

Time Analysis

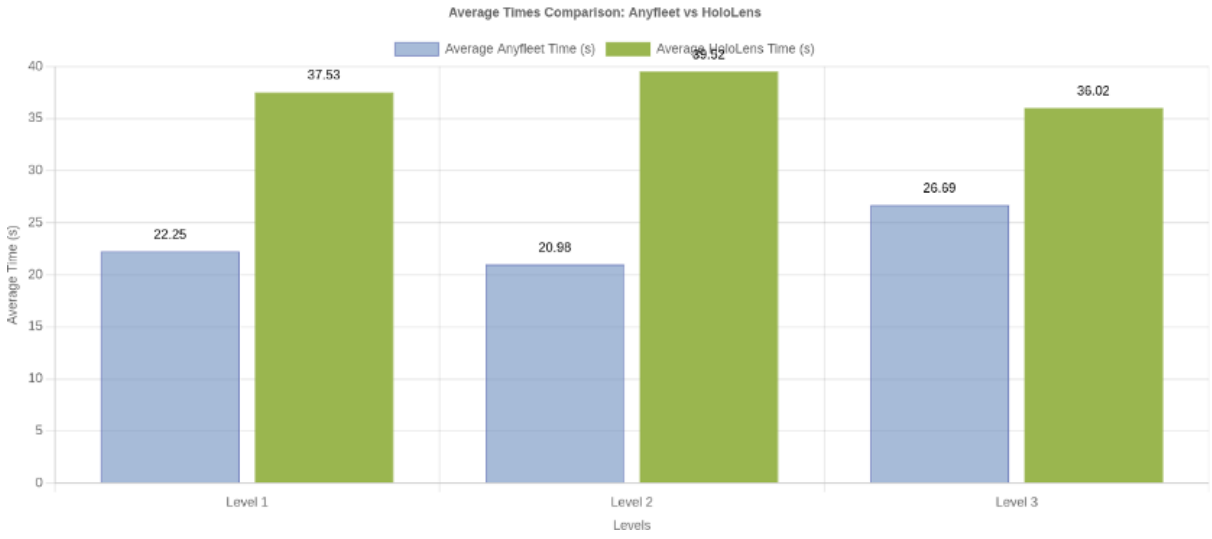

Figure 28 Average Time Comparison Between AnyFleet and HoloLens for Each Level (seconds)

NASA TLX

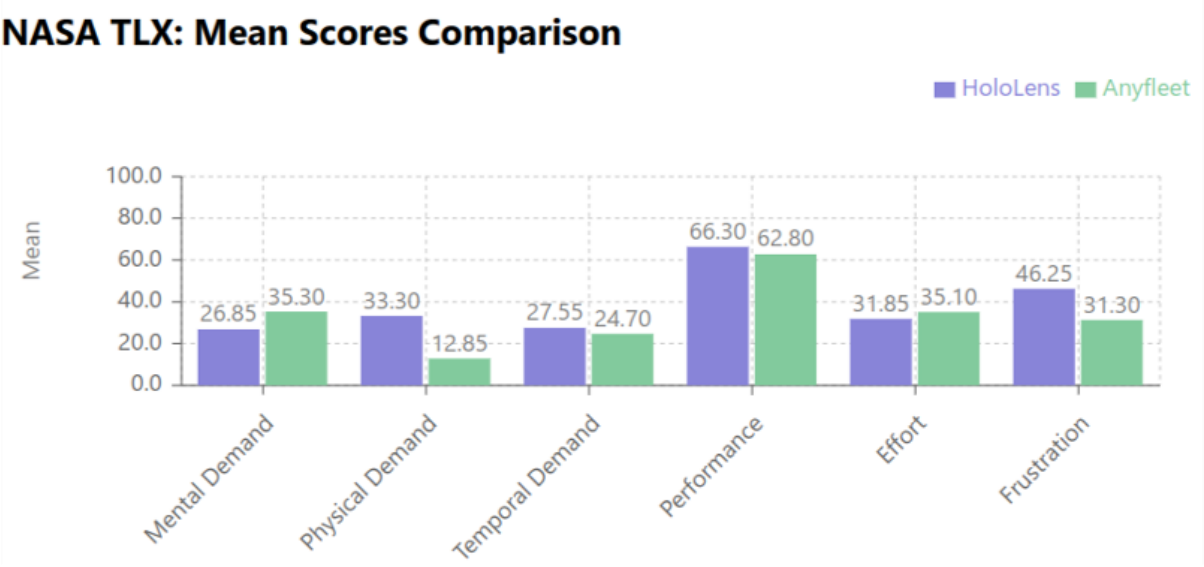

Figure 29 Mean Scores of NASA TLX Dimensions Across Hololens and Anyfleet

HoloLens required significantly more physical demand (33.3) compared to Anyfleet (12.85), while Anyfleet imposed a higher mental demand (35.3) than HoloLens (26.85). Temporal demand was slightly higher for HoloLens (27.55) compared to Anyfleet (24.7), indicating marginally greater time pressure with HoloLens. In terms of performance, HoloLens scored slightly higher (66.3) than Anyfleet (62.8), reflecting a better-perceived performance. However, Anyfleet demanded more effort (35.1) compared to HoloLens (31.85). Frustration levels were notably higher for HoloLens (46.25) than for Anyfleet (31.3), suggesting that users experienced more frustration when using HoloLens. Overall, while HoloLens demands more physical and temporal effort and causes greater frustration, it delivers slightly better performance and requires less mental effort compared to Anyfleet, which, in turn, causes less physical strain and frustration.

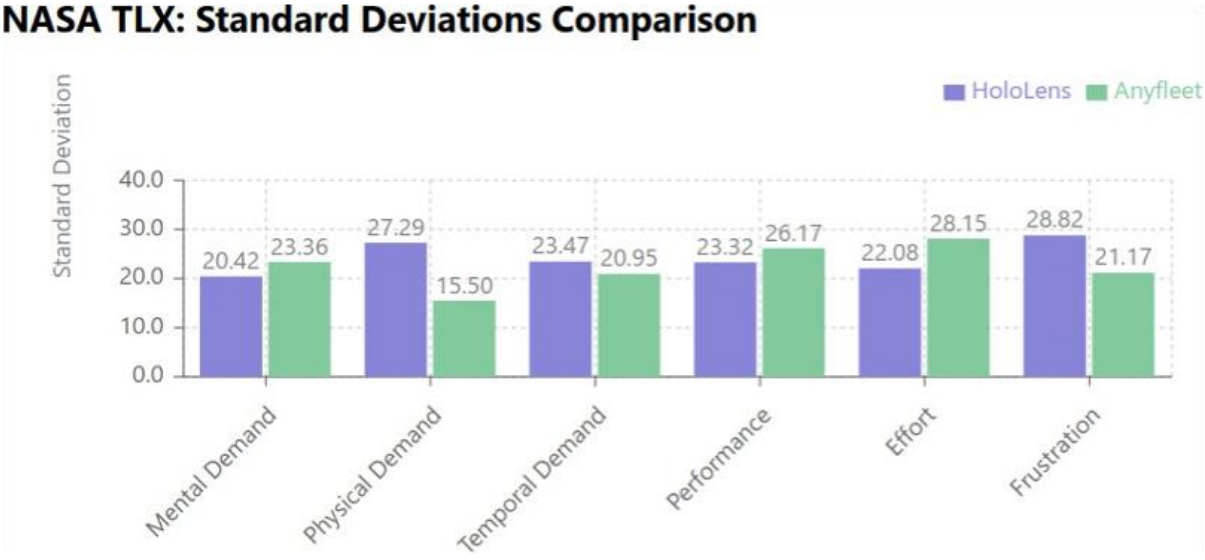

Figure 57: Standard Deviations of NASA TLX dimensions across HoloLens and Anyfleet

HoloLens exhibited a standard deviation of 20.74 for mental demand compared to 23.97 for Anyfleet, indicating slightly more consistent mental demand experiences with HoloLens. In terms of physical demand, HoloLens showed a higher standard deviation of 27.29 compared to Anyfleet's 15.50, suggesting greater variability in physical effort for HoloLens users. For temporal demand, HoloLens had a standard deviation of 23.47, slightly higher than Anyfleet's 20.95, reflecting marginally more variability in perceived time pressure. Regarding performance, HoloLens had a standard deviation of 23.32, close to Anyfleet's 26.17, showing comparable consistency in performance perception. Effort showed slightly less variability with HoloLens (22.08) compared to Anyfleet (28.15. Finally, frustration levels exhibited higher variability for HoloLens (28.82) than for Anyfleet (21.17), indicating more inconsistent frustration experiences with HoloLens. Overall, Anyfleet shows more consistent results in physical and temporal demand as well as frustration, while HoloLens demonstrates greater consistency in mental demand and effort, with both systems displaying comparable consistency in performance.

4. Conclusion

This section presents a consolidated analysis of user evaluations for a specific application, incorporating feedback from three distinct user groups: Vilanova, Etteln, and Ideal Works . The studies investigated task load, user experience, emotional response, and feature-specific feedback to identify areas of improvement and inform future development.

Perceived task load was generally low across all participating groups, suggesting that the assigned tasks were not perceived as overly demanding. However, a notable distinction emerged, with the Etteln group consistently reporting the highest perceived workload compared to both the Vilanova and Ideal Works groups. This discrepancy points to potential challenges specific to the Etteln user experience that warrant further investigation.

User experience exhibited significant variation across the different user groups. The Vilanova group consistently reported the most positive experiences, with evaluations ranging from "above average" to "excellent." The Ideal Works group presented a mixed profile, potentially influenced by its limited sample size, with user experiences ranging from below average to positive. In stark contrast, the Etteln group consistently reported negative to below-average user experiences, highlighting the presence of substantial usability issues that require immediate attention. This clear divergence in user experience underscores the significant impact of user group affiliation on overall satisfaction with the application.

Emotional responses, assessed in select studies, further differentiated the user groups. The Etteln group consistently exhibited neutral emotional states, characterized by low arousal and a moderate sense of control. This pattern of neutral emotional responses was observed across multiple evaluations, suggesting a consistent emotional impact associated with their interaction with the application.

A number of recurring issues were identified across all user groups. Foremost among these were concerns regarding navigation (both mouse and trackpad) and building placement, particularly the rotation functionality. These usability problems represent critical areas requiring immediate attention and resolution. Specific feedback also included requests for enhanced navigation tools, the incorporation of zoom

functionality, the provision of additional information about displayed content, the integration of audio elements, and improvements to the overall visual representation. The Etteln group additionally reported specific difficulties at some parts of the validations.

Correlation analysis revealed a consistent, though low to moderate, positive correlation between Attitude Towards Interacting with Technology (ATI) and overall user experience, particularly within the pragmatic dimension. This suggests that users with more positive predispositions towards technology tended to rate the application's usability more favourably. In one study, a negative correlation was observed between ATI and task load, indicating that users with higher ATI scores perceived a lower task load.

A comparative analysis of HoloLens and Anyfleet revealed inherent trade-offs between the two platforms. HoloLens was associated with higher physical and temporal demand, as well as increased user frustration, but also demonstrated better perceived performance and a lower mental demand. Conversely, Anyfleet was less physically demanding and frustrating, but required more mental effort and resulted in lower perceived performance. Anyfleet also exhibited greater consistency across user experience metrics.

In conclusion, while the limited sample size of the groups warrants careful interpretation of its results, the studies provide valuable insights into the user experience across different user groups. Furthermore, the correlation between ATI and user experience suggests that considering users' attitudes towards technology is an important factor. Addressing the identified usability issues, incorporating some suggested features, and considering the specific needs of different user groups will be essential for enhancing the application's overall usability and user satisfaction. Future research aims to have larger and more representative samples to validate these findings and further explore the complexities of user experience.

References

Andrews, R., & Entwistle, T. (2010). Does Cross-Sectoral Partnership Deliver? An Empirical Exploration of Public Service Effectiveness, Efficiency, and Equity. *Journal of Public Administration Research and Theory*, 20(3), 679-701. Retrieved from http://dx.doi.org/10.1093/jopart/mup045

Brink, T., & Madsen, S. O. (2016). The triple helix frame for small- and medium-sized enterprises for innovation and development of offshore wind energy. *Triple Helix*, 3(1), 4. doi:10.1186/s40604-016-0035-8

Carayannis, E., & Campbell, D. (2014). Developed democracies versus emerging autocracies: arts, democracy, and innovation in Quadruple Helix innovation systems. *Journal of Innovation and Entrepreneurship*, 3(1), 1-23.

Funded by the European Union

Funded by the European Union under grant number 101092875. UK based consortium partner Trilateral Research Limited, UK is funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant number 10069394).

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or UKRI. Neither the European Union nor the granting authority can be held responsible for them.