UNIVERSITY OF CRETE
DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF SCIENCES AND ENGINEERING

The ARgus Designer: An XR evaluation tool for ubiquitous applications in Intelligent Environments

by

Eleni Stefanidi

MSc dissertation submitted in partial fulfillment for the
Master of Science degree in Computer Science
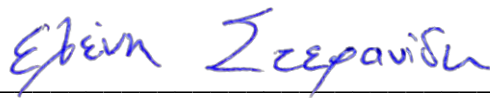
Heraklion, September 2022

UNIVERSITY OF CRETE

DEPARTMENT OF COMPUTER SCIENCE

The ARgus Designer: An XR evaluation tool for ubiquitous applications in Intelligent Environments

by **Eleni Stefanidi**

in partial fulfillment of the requirements for the
Master of Science degree in Computer Science

APPROVED BY:
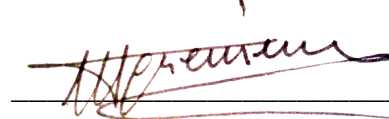
_____

Author: **Eleni Stefanidi**

_____

Supervisor: **Georgios Papagiannakis**, Professor, University of Crete

_____

Committee Member: **Panagiotis Tsakalides**, Professor, University of Crete

_____

Committee Member: **Margherita Antona**, Principal Researcher, FORTH-ICS

_____

Director of Graduate Studies: **Polyvios Pratikakis**, Associate Professor, University of Crete

Heraklion, September 2022

iii

To my beloved family…

Acknowledgements

First of all, I would like to express my sincere appreciation to my Supervisor, Professor George Papagiannakis, for his guidance, support and for believing in me.

Then, I would like to thank my advisors Dr. Asterios Leonidis and Dr. Maria Korozi for their insights, guidance and encouragement in the course of this work. I also acknowledge the valuable input of Dr. Margherita Antona.

I am indebted to the Department of Computer Science (CSD) of the University of Crete, and the Human Computer Interaction (HCI) Laboratory of the Institute of Computer Science (ICS) of the Foundation for Research and Technology - Hellas (FORTH) for creating a highly motivational working environment that facilitated the achievement of my research goals. Moreover, I would like to acknowledge the scholarship of the Institute of Computer Science of FORTH during my postgraduate studies.

Last but not least, I would like to express my gratitude to my family for their eternal love and unwithering support.

Abstract

Augmented Reality (AR) and especially Mixed Reality (MR) applications have gained momentum over the past couple of years among researchers, developers, enterprises and consumers. With new MR devices and accessories being released in the market at a constantly increasing pace, from HoloLens, Magic Leap and Varjo, to the upcoming Meta and Apple devices, the AR/MR field is set to revolutionize computing. As this field matures, it has progressively cultivated an imperative need for evaluating such applications. Even though many approaches have addressed the challenges of evaluating AR/MR applications, with some even using these technologies as evaluation methods, only a few attempts have been made in creating tools to support their overall evaluation process.

To this end, this Thesis aims to contribute filling this gap by proposing the 'ARgus Designer System' whose objective is to assist evaluators in conducting user studies for assessing the usability and User Experience (UX) of AR/MR applications. In the proposed evaluation tool, the information provided to experts' is enhanced in real-time during testing with various metrics and features, while wearing an MR device (specifically the HoloLens 2), in an effort to facilitate the evaluation process. Focusing primarily on the domain of prototyping for Intelligent Environments (IEs), the first instantiation of the envisioned system augments the user's surroundings by overlaying virtual interactive artifacts, 3D objects and digital user interfaces (UIs) on top of the real-world environment. In particular, the 'ARgus Designer' system enables users to create their own virtual experiment scene, by placing the desired

digital objects into the MR world. Users can additionally select the metrics and features they want to include in the mixed-reality environment during the evaluation process. Moreover, the suggested evaluation tool supports multiple users and facilitates multi-user collaboration, enabling experts to conduct user-based studies completely immersed into the Mixed Reality space.

At runtime, apart from the different metrics and features being visualized in AR, evaluators have access to a virtual workstation, called the 'ARgus Workstation', through which they can control the flow of the experiment and collect various runtime measures and data (e.g. performance indicators, events of interest). Another noteworthy feature of the system is the inclusion of virtual questionnaires, which not only enhance experts' supervision, but most importantly minimize the "breaks in presence" of the immersive experience.

Finally, the introduced tool is envisioned to help with post-study analysis, permitting experts to overview raw data and replay an evaluation experiment into the MR world. Through all the aforementioned processes, the 'ARgus Designer' system aims to deliver an all-round efficient approach of conducting user-based evaluation with the help of AR/MR technologies.

In this context, this Thesis presents the 'ARgus Designer' system, describes the design process followed and its functionality, elaborates on implementation details and outlines the results of the expert-based evaluation to assess the validity of the concept from an expert's point-of-view.

Keywords: User Studies, Evaluation, Augmented Reality, Design, Intelligent Environments, Interactive Prototyping

Περίληψη

Οι εφαρμογές Επαυξημένης Πραγματικότητας (Augmented Reality - AR) και ιδιαίτερα οι εφαρμογές Μικτής Πραγματικότητας (Mixed Reality - MR) έχουν σημειώσει ραγδαία ανάπτυξη τα τελευταία χρόνια μεταξύ ερευνητών, προγραμματιστών, επιχειρήσεων και καταναλωτών. Με νέες συσκευές και εξαρτήματα MR να κυκλοφορούν στην αγορά με συνεχώς αυξανόμενο ρυθμό, όπως τα HoloLens, Magic Leap, Varjo, αλλά και τις επερχόμενες συσκευές των εταιριών Meta και Apple, ο τομέας AR/MR αναμένεται να φέρει μια νέα επανάσταση στον τομέα της Πληροφορικής. Καθώς η περιοχή αυτή ωριμάζει, έχει σταδιακά δημιουργηθεί η επιτακτική ανάγκη για την αξιολόγηση τέτοιων εφαρμογών. Παρόλο που πολλές προσεγγίσεις έχουν ασχοληθεί με τις προκλήσεις που συνδέονται με την αξιολόγηση εφαρμογών AR/MR, με ορισμένες μάλιστα να χρησιμοποιούν αυτές τις τεχνολογίες ως μεθόδους αξιολόγησης, λίγες μόνο προσπάθειες έχουν γίνει για τη δημιουργία εργαλείων που υποστηρίζουν τη συνολική διαδικασία αξιολόγησής τους.

Για το σκοπό αυτό, η παρούσα εργασία στοχεύει να συμβάλει στην αντιμετώπιση αυτού του κενού, προτείνοντας το 'ARgus Designer System', στόχος του οποίου είναι να βοηθήσει τους αξιολογητές (evaluators) στην εκτέλεση μελετών με τελικούς χρήστες (user studies) για την αξιολόγηση της ευχρηστίας και της εμπειρίας χρήστη (User Experience - UX) των εφαρμογών AR/MR. Το προτεινόμενο εργαλείο αξιολόγησης ενισχύει σε πραγματικό χρόνο την διαθέσιμη πληροφορία μέσω πλήθους μετρήσεων και λειτουργιών, κατά τη διάρκεια διεξαγωγής μελετών με τελικούς χρήστες, μέσω μιας φορετής συσκευής MR (συγκεκριμένα το HoloLens 2), διευκολύνοντας κατά αυτόν τον τρόπο τη διαδικασία της αξιολόγησης. Εστιάζοντας κυρίως στον τομέα της

δημιουργίας πρωτοτύπων (prototypes) για Ευφυή Περιβάλλοντα (Intelligent Environments - IEs), η πρώτη εφαρμογή του εν λόγω συστήματος επαυξάνει το περιβάλλον του χρήστη με την τοποθέτηση εικονικών διαδραστικών τεχνουργημάτων (artifacts), 3D αντικειμένων και ψηφιακών διεπαφών χρήστη (User Interfaces - UIs) μέσα στο πραγματικό περιβάλλον. Συγκεκριμένα, το σύστημα 'ARgus Designer' επιτρέπει στους χρήστες του να δημιουργήσουν το δικό τους τεχνολογικά επαυξημένο πείραμα, τοποθετώντας τα ψηφιακά αντικείμενα που επιθυμούν στον MR κόσμο. Οι χρήστες μπορούν επιπλέον να επιλέξουν τις μετρήσεις και τις λειτουργίες που θέλουν να συμπεριλάβουν στο περιβάλλον μικτής πραγματικότητας κατά τη διάρκεια της διαδικασίας αξιολόγησης. Επιπλέον, το προτεινόμενο εργαλείο αξιολόγησης μπορεί να υποστηρίξει πολλαπλούς χρήστες και να διευκολύνει τη συνεργασία τους, επιτρέποντας στους ειδικούς να διεξάγουν αξιολογήσεις πλήρους εμβύθισης (immersed) στον χώρο μικτής πραγματικότητας.

Κατά τη διάρκεια της εκτέλεσης μιας αξιολόγησης, εκτός από τις διάφορες μετρήσεις και τις λειτουργίες που απεικονίζονται στην επαυξημένη πραγματικότητα, οι εμπειρογνώμονες - αξιολογητές έχουν πρόσβαση σε έναν εικονικό σταθμό εργασίας (workstation), που ονομάζεται 'ARgus Workstation', μέσω του οποίου μπορούν να ελέγχουν τη ροή του πειράματος και να συλλέγουν διάφορες μετρήσεις (π.χ. δείκτες απόδοσης, γεγονότα ενδιαφέροντος). Ένα άλλο αξιοσημείωτο χαρακτηριστικό του συστήματος είναι η ενσωμάτωση εικονικών ερωτηματολογίων (virtual questionnaires), τα οποία όχι μόνο ενισχύουν την εποπτεία των εμπειρογνωμόνων, αλλά επιπλέον ελαχιστοποιούν τα 'breaks in presence' της εμβυθιστικής εμπειρίας (immersive experience).

Τέλος, το εργαλείο αυτό προβλέπεται να βοηθήσει στην ανάλυση των αποτελεσμάτων μετά το πείραμα, επιτρέποντας στους εμπειρογνώμονες – αξιολογητές να κάνουν επισκόπηση των δεδομένων και να αναπαράγουν ένα πείραμα αξιολόγησης σε έναν κόσμο Μικτής Πραγματικότητας. Μέσα από τις παραπάνω διαδικασίες, το σύστημα 'ARgus Designer' στοχεύει να προσφέρει μια ολοκληρωμένη και αποτελεσματική μέθοδο, με τη βοήθεια τεχνολογιών AR/MR, για την διεξαγωγή μελετών αξιολόγησης ευχρηστίας με τελικούς χρήστες.

Σε αυτό το πλαίσιο, η παρούσα εργασία παρουσιάζει το σύστημα 'ARgus Designer', περιγράφει τη διαδικασία σχεδίασής του και τη λειτουργικότητά του, αναλύει τις τεχνικές λεπτομέρειες της υλοποίησής του και περιγράφει τα αποτελέσματα αξιολόγησης απο εμπειρογνώμονες (expert-based evaluation) που διενεργήθηκε για τη διερεύνηση της εγκυρότητας της ιδέας (concept) του συστήματος από την οπτική γωνία των εμπειρογνωμόνων.

Λέξεις κλειδιά: Μελέτες με συμμετοχή χρηστών, Αξιολόγηση, Επαυξημένη Πραγματικότητα, Σχεδίαση, Ευφυή Περιβάλλοντα, Αλληλεπιδραστικά Πρωτότυπα

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Augmented reality (AR) and Mixed Reality (MR) headsets and applications are becoming increasingly accessible to consumers, boosting research interest in AR/MR technologies [48]. Even though such applications have been ranked among the Top-10 ICT technologies for the past couple of years [48, 49], and despite their accelerating appeal, there is still a paucity of evaluation techniques and tools used for their assessment in contrast to "conventional" applications (e.g. desktop, smartphones and tablets, etc.) [38]. In particular, from an UX perspective, only a few common evaluation tools exist, with the majority of the studies utilizing video analysis or building additional ad-hoc solutions on top of research prototypes, so as to capture quantitative data [14].

From a user perspective, AR applications are not so widely adopted and used mainly due to high costs and complex operations [67]. In particular,

their low traction may be attributed to various usability issues that have been identified in AR applications, such as their lack of consistency due to the AR interaction paradigm, [41], [13] that might make them less desirable to many end-users. These usability problems might be attributed to the lack of methodologies, frameworks and tools to properly support the evaluation process. Considering that the evaluation of interactive systems has always been central to Human-Computer Interaction (HCI) and one of the main topics (along with design and implementation) of the HCI curriculum [24], [34], it is crucial to support evaluators and the evaluation process of AR and MR-based applications as well.

The problem is further magnified considering that AR/MR applications often inter-operate with and are fueled by Intelligent Environments [58], [5]. Applications for such environments present unique characteristics that differentiate them from traditional applications. Traditional applications focus on one device (desktop or mobile) while applications for Intelligent Environments are distributed across various devices in the surrounding space, which may include multiple displays that need to be synchronized based on the user context, meaning that they should recognize and respond to user presence and location. Moreover, an application may include interactive artifacts that are not displays, such as smart lamps and smart blinds. Interaction within such environments is greatly influenced by numerous factors that extend beyond the scope of the application itself, such as the overall state of the space (e.g. its smart objects and devices), the position of the users, as well as any direct or indirect (user) actions. Additionally, given the plethora of interpolated devices, with or without User Interfaces (UIs), a user may interact with a single UI, but other interfaces or devices in the space

might get affected as well; this can also occur if voice-based interaction is used instead. When designing for such a complex environment, the placement (e.g. position, orientation) and even the physical appearance of an artifact can greatly affect the User Experience. As a consequence, it would be useful to have the opportunity to quickly test alternative solutions (e.g. is the placement of a UI presenting a list of the recipe steps to be followed on the counter-top better than projecting it on the kitchen wall?). Furthermore, it is rather common not to have all the equipment from the beginning of the design phase; screens, projectors and devices might be missing. Therefore, it would be useful to be able to visualize alternative layouts via AR/MR. These aspects highlight the necessity for testing user scenarios early in the design phase, so that the overall User Experience, including the shortcomings of the application, are recognized and comprehended by the designer.

This work presents 'ARgus Designer', a tool that uses Augmented Reality to assist evaluators in conducting user studies for assessing the usability and User Experience (UX) of AR/MR applications. Additionally, acknowledging the significant challenges faced by the designers of systems for Intelligent Environments (e.g. missing equipment, lack of visualization techniques, sparsity of user monitoring tools, etc.), ARgus Designer offers a tool enabling rapid prototyping and testing.

1.2 Objectives

ARgus Designer aims to assist evaluators in conducting user studies for assessing the usability and User Experience of AR/MR applications. In particular, through Mixed Reality glasses (i.e. Microsoft HoloLens 2), evaluators can see the digitally enhanced environment that the evaluation participants interact with, permitting them to observe their operation first hand and thus facilitating the evaluation process.

3

Additionally, the evaluator's view is augmented with valuable information during testing (e.g. time elapsed, the scenario that is being followed) making the supervision of the user's actions in the intelligent environment more efficient and effective. Following the experiment, the system permits evaluators to relive an evaluation experiment in extended reality environment by displaying the snapshots, the video-recordings and experiment data which are important for data (e.g. times, errors, hints), and by replaying the participants' actions.

Considering that interactive prototyping while designing an application for Intelligent Environments can help designers understand better the User Experience of their designs and capture usability issues early in the design process, ARgus Designer enables users to setup an AR evaluation scene by integrating interactive prototypes (e.g. mockups) and artificial 3D objects (e.g. a smart fridge, a tablet, a display, a shelf, a microwave oven). This feature is particularly helpful since it permits evaluators to observe participants interacting with a working prototype without the need to develop it first.

The key objectives of this research endeavor are:

- Enable evaluators to configure an evaluation scene (i.e. add/remove digital artifacts, prototypes) a-priori and load this configuration during testing.
- Augment the evaluators' view with useful information during user-based testing.
- Permit users to interact with virtual interactive artifacts superimposed through AR in their surrounding environment.
- Support simultaneous use by multiple users, i.e. multiple users interacting with the environment and an evaluator/observer monitoring their interaction unobtrusively.

- Support post-evaluation activities (e.g. reliving the experiment) in a Mixed Reality environment.

1.3 Thesis structure

This Master's Thesis comprises six Chapters as indicated in the Table of Contents, with the following structure:

- Chapter 2 reviews related work regarding evaluation systems and approaches that utilize XR for the evaluation process.
- Chapter 3 describes the design methodology that was followed, introduces scenarios that motivated this work, and outlines the functional requirements of the proposed system.
- Chapter 4 describes the functionality and user interface of the system.
- Chapter 5 describes the cognitive walkthrough evaluation study that was conducted with the participation of six (6) User Experience (UX) experts, and presents the findings in detail.
- Chapter 6 concludes with a summary of the results and a discussion of possible future directions.

Chapter 2

Background Theory and Related Work

2.1 Background Theory

This section discusses the evaluation of XR systems, with an emphasis on AR/MR, which has been recently the subject of much research work. Firstly, the findings of a literature review regarding usability in AR applications are presented. Secondly, the conceptual model and framework that this work was inspired by is introduced. Then, some common challenges that researchers and developers face when designing and evaluating AR/MR applications are mentioned. Existing methods to evaluate prototypes are finally addressed with an emphasis on approaches that utilize XR along with the most widely-used practices and performance measures in XR evaluations.

2.1.1 Previous Research & Challenges of Evaluating XR Apps

In recent years, an abundance of research endeavors has dealt with the evaluation of AR/MR systems [15], [21], [16], [56] and their usability [14], [7]. Even though these research efforts have established some common ground regarding the evaluation methods and study design, the tools and heuristics utilized in the studies mostly differ [38]. In relation to usability evaluation, one of the most cited works regarding user-based experimentation in Augmented Reality [16] established three types of experiments that are relevant to AR research: 'Human perception and cognition', 'User task performance and interaction techniques' and 'User interaction and communication between collaborating users'. In addition, Bai et al. [7] conducted an analytic review of papers published in International Symposium of Mixed and Augmented Reality (ISMAR) proceedings from 2001 to 2010. Moreover, Dey et al. [14] conducted a systematic literature review of the most influential AR user studies, from 2005 to 2014, presenting the broad landscape of user-based AR research, and providing a high-level view of its changes, such as the shift towards controlled laboratory experiments and the largest application areas being perception and interaction.

Taking the latter literature review also into account, Marques et al. [36] initially proposed a conceptual model for multi-user data collection and analysis which included four modules: data gathering, analysis, inspection and report, and then created a conceptual framework to assist researchers in conducting AR evaluations in a more structured manner, with a focus on collaboration [35]. In this framework, the contextualized data gathering step of the evaluation process includes: pre-task measures, runtime measures and post-task measures. The pre-task measures can comprise demographic questionnaires and participants'

background information, while the post-task measures might include tools' usability, collaboration metrics and feedback from interviews. Special attention is given to the runtime measures which may encompass performance metrics, behavior metrics, collaboration metrics and interesting events (audio, video, requested assistance, etc.).

Other works ([48], [57], [4]) dealt with the challenges researchers and developers face when designing and evaluating AR/MR applications such as problems with cross-device communication, environmental mapping, tracking, the obtrusiveness of the applications and the hardware. In particular, Rokhsaritalemi et al. [48] identified two major challenges for creating a platform in an MR environment: display technology and tracking. Appropriate display technology needs to be used for the MR platform to produce a reasonable output with appropriate resolution and contrast, while precise methods are required to track the interaction between virtual and real objects.

Speicher et al. [57] focused on cross-device interaction challenges of AR platforms. Amongst their key findings, they highlight that: a) the AR content should be made available on various everyday devices, displays and output modalities; b) the AR content should be synchronized between multiple users using different devices in real time; c) a global coordinate system is needed to ensure correct placement of the digital objects; and d) the AR devices should be less obtrusive.

Additionally, Ashtari et al. [4] mentioned that AR/VR creators face many challenges and difficulties while building their projects such as: a) they have to carefully choose the appropriate hardware to use due to significant differences amongst the various available headsets, lengthy installation procedures specific to headsets and other peripherals, etc.; b) they have difficulties in planning targeted experiences in AR, since

9

they cannot forecast users' actions/movements; c) viable testing methods are lacking; d) users' knowledge of tool usage and access to AR/VR devices are inadequate; and e) concrete design guidelines and examples are required.

2.1.2 Existing Evaluation Methods, Practices & Performance Measures

Various methods that evaluate prototypes exist with the most common being online surveys, lab studies, and in-situ studies [62]. Additional methods include expert evaluation through cognitive walkthrough and heuristic evaluation [44]. Technological advancements have enabled for new methods to emerge such as evaluation through VR [30] and AR [43], [27]. Particularly, Knierim et al. [30] showed the potential of VR for a vast variety of use cases by facilitating natural generic text input on a physical keyboard while being immersed in a VE with an apparatus that tracks the user's hands and visualizes them in VR. An indicative evaluation that was conducted through the use of AR technology is reported in [43], where the authors used Spatial Augmented Reality (SAR) in the prototyping of new human-machine interfaces (e.g. control panels, car dashboards), by developing a SAR system to make prototypes with interactive virtual components. Kun et al. [27] followed a different approach and conducted a study using the AR device HoloLens to investigate the visual distraction of drivers when they use an AR device for video calling while driving.

The most commonly adopted practices to collect objective and subjective data for the evaluation of XR applications, include: interviews [6], scale-based questionnaires such as System Usability Scale (SUS) and NASA Task Load Index (TLX), as well as various presence and immersion questionnaires [7]. In contemporary research, administering

questionnaires in XR, especially VR (INVRQS), is becoming more and more common, as it is believed that INVRQS may ease participation, reduce the Break in Presence (BIP) and avoid biases [45].

With respect to integrating user questionnaires inside the augmented environment, research has shown that participants prefer using questionnaires inside the virtual environment (VE) and it is recommended for researchers to apply them for their user studies. Embedding question-items in VEs, in contrast to, for example, paper-based questions, offers an opportunity to stay closer to the context of an ongoing exposure than out-of-VR research setups and avoid BIPs [1]. BIPs can be problematic for researchers and evaluators as they can lead to temporal disorientation and loss of sense of control [28] making questionnaire results possibly biased to a degree that is difficult to measure and might vary in different cases [1].

Finally, concerning the performance measures that are mostly adopted in XR evaluations, time and errors/accuracy were the most prevalent ones [7], [14], with some studies taking into account additional factors relevant to task performance like: cognitive load [60], distraction rate [61], cognitive support [18] and navigation behavior [39]. In general, the most broadly used dependent measure is subjective ratings, followed by error/accuracy measures and task completion time [14]. Furthermore, a limited number of research efforts ([2], [36]) mention a few additional useful performance metrics such as number and type of interactions, frequency of using each feature along with screenshots of the augmented content for post-study analysis. They also distinguished a number of behavior measures, namely the physical movement around the environment, the number and type of hand gestures, the physiological variables and emotions, the eye gaze as well as video and

audio recording for post-study analysis. Other studies considered physiological measurements of task strain as well, along with heart rate, electrocardiogram (ECG), galvanic skin response (GSR) and skin temperature [52], or head movement as a measure of musculoskeletal strain [23].

2.2 Related Work Overview

This section first introduces some state-of-the-art systems that use XR technologies to support the evaluation process and the visualization of data, and then present various research approaches aiming to facilitate the evaluation of XR applications.

2.2.1 Systems that Use XR to Support the Evaluation Process

Concerning existing tools that utilize Extended Reality to assist in evaluations and data visualizations, Nebeling et al. [38] designed the Mixed Reality Analytics Toolkit (MRAT). MRAT presents concepts and techniques to support usability evaluations of MR apps, through plug-in-like instrumentation of the apps in Unity, the use of virtual or physical markers to define user tasks, and a set of heuristics and metrics that can feed visualizations for analysis. A case study was conducted which demonstrated how to use MRAT to instrument an MR crisis simulation and triage training app (Figure 1). In this case study, relevant metrics and visualizations were produced, enabling instructors to remotely observe student teams collaboratively solving a crisis scenario and understand the team's different behaviors and performance. Even though the scope of this research paper is limited to a specific context and scenario (i.e. disaster management) and its generalizability still needs to be established, the authors elicited requirements for future tools, through extensive interviews, to inform the design of MRAT and in an effort to make it applicable to many use cases. Some key findings

were the need for quantitative and spatial data collection within the MR app about interactions that are difficult to capture through video (e.g. where users walked, what users looked at and clicked), the need for key performance indicators such as time spent looking in a certain direction or standing at a certain location, as well as trends and patterns between groups of users including frequent locations or targets.
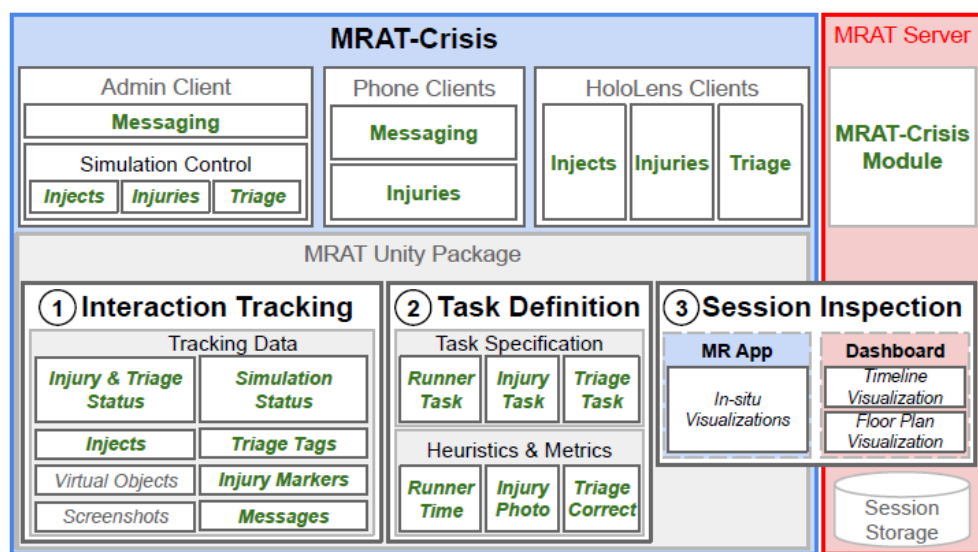
Figure 1: Architecture of the MRAT-Crisis app after instrumentation with MRAT.

Additionally, Büschel et al. [9] created MIRIA (Figure 2), a Mixed Reality Interaction Analysis toolkit designed to support the in-situ visual analysis of user interaction in MR and multi-display environments by embedding AR visualizations of spatial interaction data into the physical locations where it was originally recorded. The work focuses on the analysis of user movement, spatial interaction, and event data. Unlike other existing solutions the toolkit is multi-user capable and thus allows for a co-located, collaborative visual analysis. Specifically, the toolkit operates on one or more Microsoft HoloLens Augmented Reality head-mounted displays and is independent of any external instrumentation. Their

13

approach includes 3D views which can be integrated with videos or images and 2D visualizations (e.g. heatmaps, scatterplots, event timelines), all of which are virtually placed in the physical world.

Figure 2: The MIRIA toolkit supports the co-located, in-situ analysis of spatial interaction data by multiple users in Augmented Reality.

ReLive is another tool that uses XR technologies for exploring and analyzing Mixed Reality user studies [25] (Figure 3). This mixed-immersion visual analytics framework combines an in-situ virtual reality view with a complementary ex-situ desktop view. The in-situ virtual reality view enables users to relive interactive spatial recordings, replicating the original study and providing the possibility for in-situ analysis of data. The ex-situ desktop view provides a familiar interface for a holistic overview and analysis of aggregated study data. Specifically, in the in-situ VR view, the evaluators can walk through the scene, create visualizations based on entities and events within the scene, view the study data within its original environmental context, similarly to the two aforementioned systems [9] [38], but utilizing VR instead of AR. The authors evaluated ReLive through a two-step evaluation process consisting of a guided design walkthrough to analytically investigate the system in a formative evaluation, and an

empirical expert user study to gather real world insights into the systems applicability.
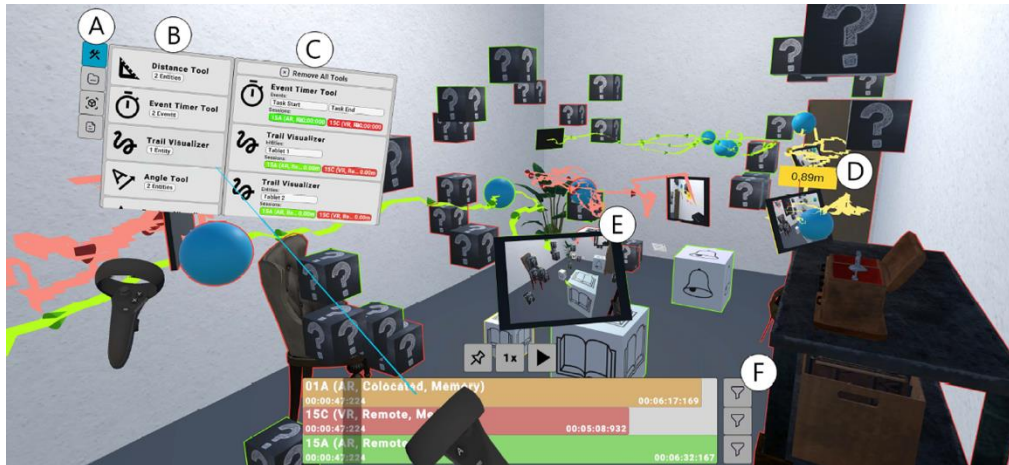
Figure 3: ReLive VR view. This view immerses analysts in an interactive reconstruction of the original study. A user interface is affixed to the analyst's left controller and a timeline at the bottom allows analysts to control the simulation.

Another research effort aiming to help AR user studies is AvatAR [46] (Figure 4), an immersive analysis environment for the in-situ visualization of human motion data, which combines 3D trajectories with virtual avatars showing people's detailed movement and posture. This work operates with an AR HMD and a tablet. The interface allows for precise input as well as a concise overview of the current scene. Additional visualizations in the AR environment, such as what people looked at, their footprints on the ground, and where they interacted with the environment by touching it, make the data experienceable from the evaluator's point of view.

Figure 4: The AvatAR analysis environment. (a) Ghost Preview technique with two pinned hand ghosts touching the display. (b) Specter Visualizations technique of just the avatar's hands, showing past and future time frames as semi-transparent instances of the hands. (c) Gaze Visualization for two persons, one looking at the display and one looking at the other person.

Moreover, the work by Lobo et al [33] introduces Flex-ER (Figure 5), an environment designed to ease sharing immersive analytics interaction techniques and user studies by enabling users to prototype, test and conduct user studies for immersive data analytics. Its goal is to create a common ground to evaluate immersive interactive visualizations across devices without having to design new experimental platforms. Flex-ER relies on a JSON specification that describes the data, the visualization, the interaction techniques and the devices and inputs for each user study. The interaction techniques are defined using a state machine and are easily reused since they are independent of the device that created them.
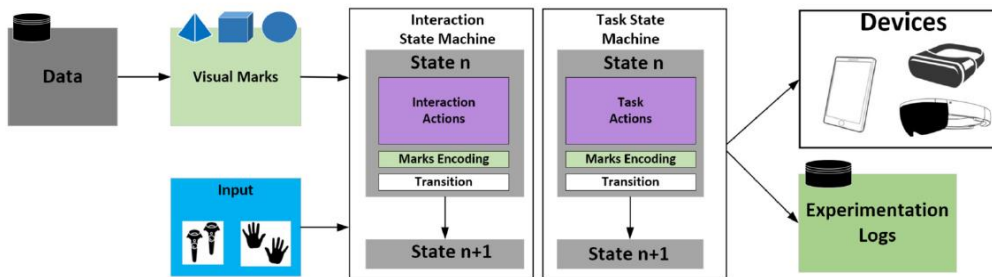
Figure 5: Schematic pipeline of Flex-ER depicting the components considered.

A number of toolkits designed specifically for visualizing data in extended reality are IATK [12], DXR [55] and VRIA [10] (Figure 6), which are based on grammars of graphics to define immersive visualizations, using either

Unity or web-based technologies. The first and second toolkit are based on the Unity development platform and allow interactive authoring and exploration of data visualization in immersive environments (AR and VR). They achieve this by a grammar of graphics that a user can configure in an in-situ GUI to create novel immersive visualization designs and interactions, with the first one also including a low-level API. The third tool is a Web-based framework for creating Immersive Analytics (IA) experiences in VR and offers a visualization creation workflow which enables users to rapidly develop Immersive Analytics experiences for the Web. VRIA is ubiquitous and platform-independent, built upon WebVR, A-Frame, React and D3.js, using a declarative format for specifying visualization types through simple configuration files, simplifying visualization prototyping, data binding and interaction configuration.
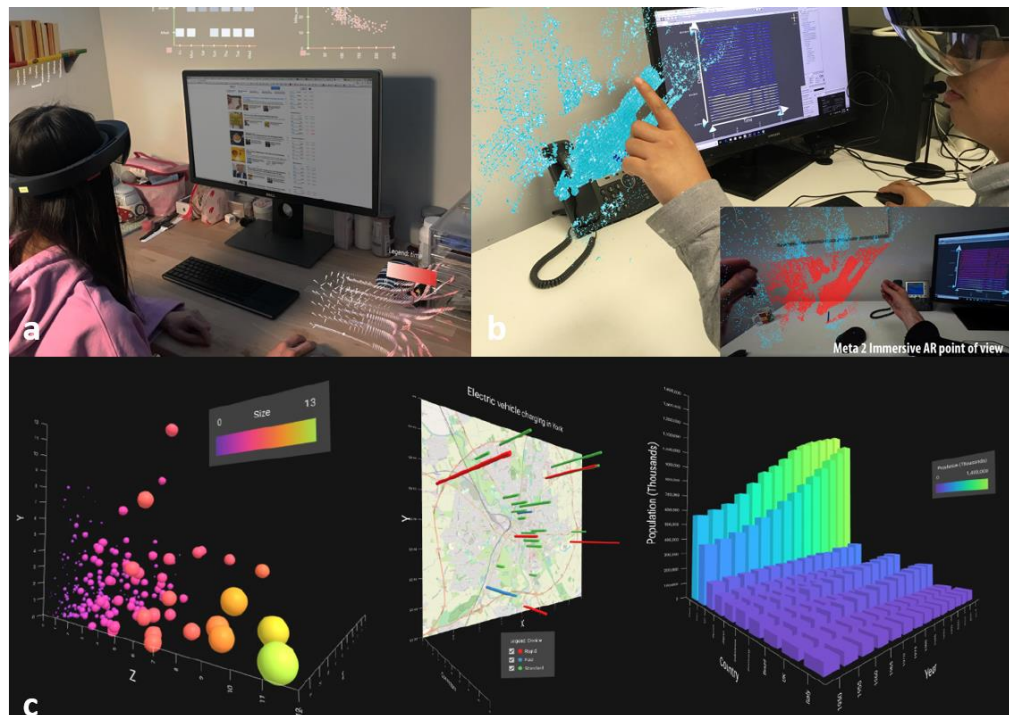
Figure 6: Examples of immersive visualizations built using (a) IATK (b) DXR (c) VRIA.

17
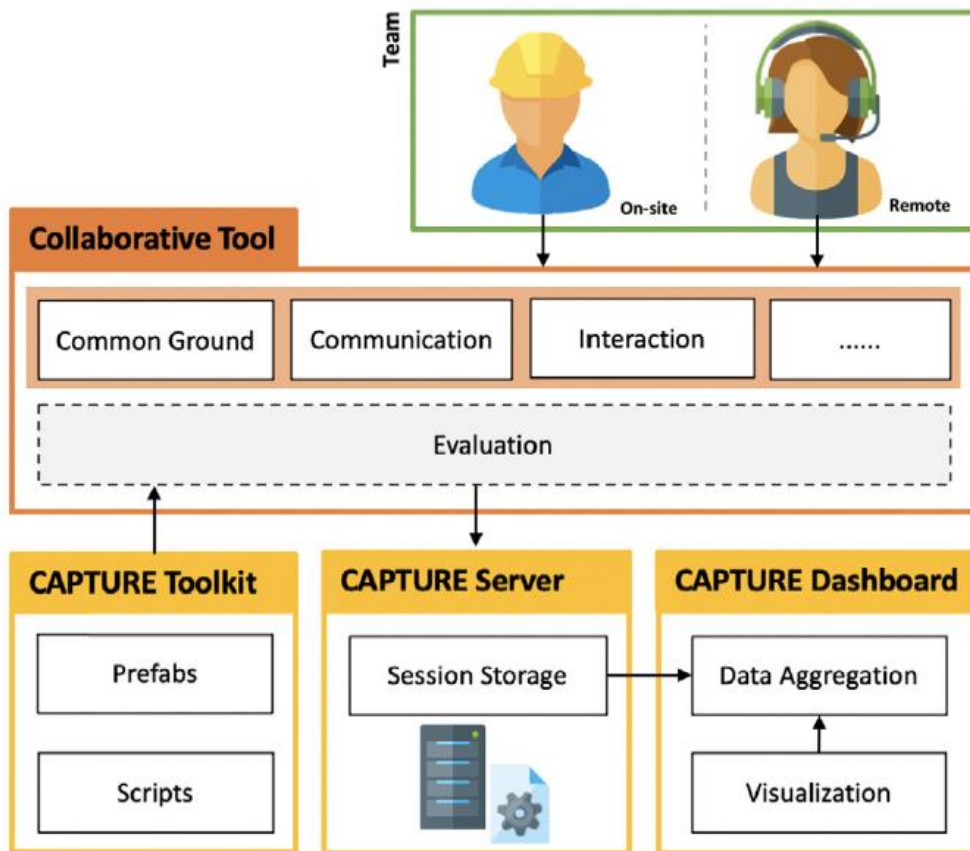
Figure 7: CAPTURE architecture. CAPTURE can be integrated into a collaborative tool via visual editor. All data collected during collaboration is stored in a central server and are analyzed through visualization dashboard.

A different approach is introduced by the CAPTURE toolkit [35], which addresses the need to include more contextual data in the evaluation process by focusing on the direction of remote collaboration using AR. Specifically, this evaluation toolkit consists of a Unity Package that can easily be added to Unity and supports contextual data collection and analysis of collaborative processes among a team, a set of teams or different tools. CAPTURE supports explicit input and data gathering regarding individual and team profile, collaborative context (e.g. number of completion stages, the number of persons), list of events (e.g. task duration, augmented content), pre-defined measures (e.g. spatial

presence, enjoyment) and interaction with the collaborative tool (e.g. number/type of interactions, frequency of using each feature). All data gathered during collaboration sessions is stored in a central server for post-evaluation analysis through a visualization dashboard (Figure 7). To demonstrate the usability and versatility of the tool the authors conducted a user study in a remote maintenance scenario, comparing two methods: Video Chat and AR-based annotations.
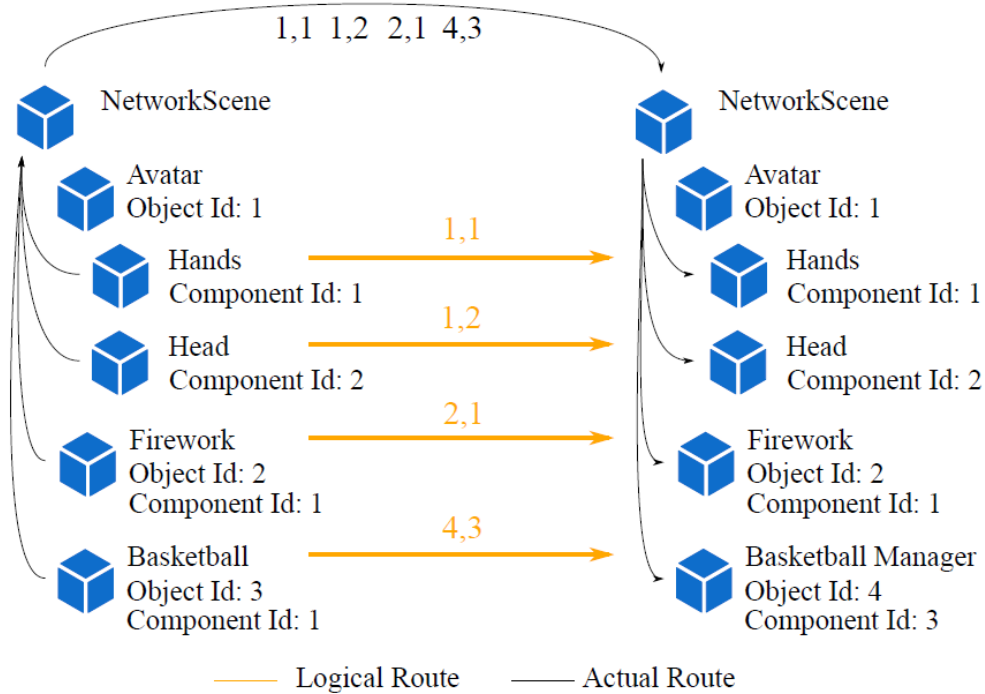
Figure 8: Diagram showing how Component instances may address each other, with the actual route taken by the messages.

Ubiq [19] is a system also built on the Unity platform that helps developers to build cross-platform social virtual reality (SVR) applications and experiments. The toolkit's expected use case is for small-scale teams to employ it for building their own applications. While supporting typical features of such systems including avatar selection, voice communication and shared synchronization of objects, it has a

19

number of services that other systems are missing or do not expose, such as logging and instrumentation. Its architecture is based around the exchange of discrete messages directly between Components, instances of classes, enforcing a method to receive messages, with users implementing networked behavior in them (Figure 8). This system's usability has been tested in a classroom setting, where the system was successfully operated by students without prior networking experience.
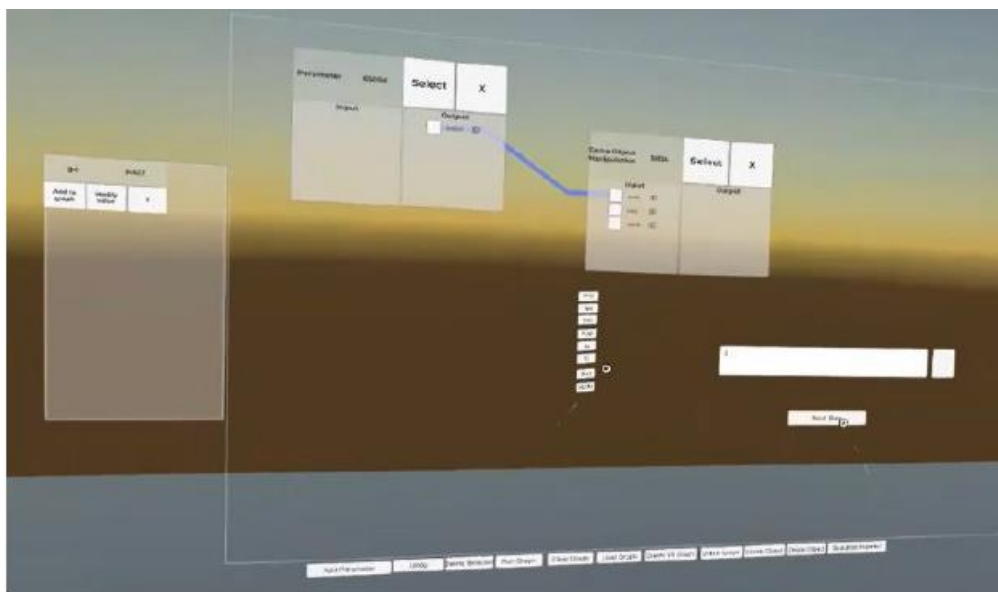
Figure 9: RealityFlow Prototype Whiteboard Interface. Parameters - input (left), nodes with input and output ports (center) and graph system commands (bottom).

Furthermore, RealityFlow [37] is an open-source immersive authoring tool which has been implemented on top of the Unity engine. The goal of this platform is to enable its users to compare different immersive visual and interaction choices using programming languages. Particularly, RealityFlow implements a dataflow-based VPL immersive editor, a multi-user real-time server for synchronous editing, and cross-platform support across mobile devices and immersive headsets. The

system consists of three main components: a Unity plugin supporting different platforms (Unity, Mobile and VR currently), a Node.js server handling real-time updates and commands and a GraphQL endpoint. The interactions, including selections and movement of nodes and manipulations of the whiteboard the system features (Figure 9), are updated live. Nevertheless, the currently available solution of the platform lacks sophistication in its interface and functionality, so it simply demonstrates the possibilities of the envisioned tool.
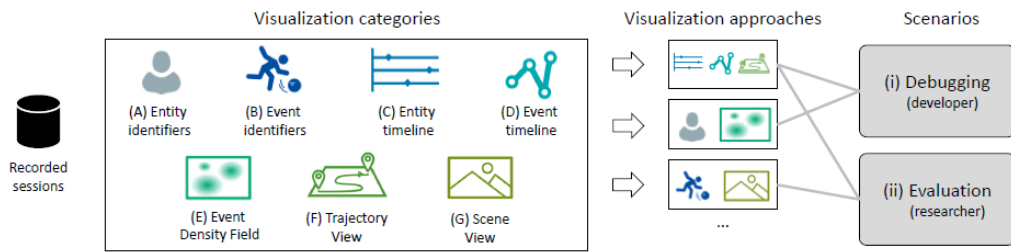
Figure 10: Design and application space for visualizations of recorded virtual or Mixed Reality sessions; the seven categories of visual encodings (A–G) provide the building blocks of specific visualization approaches, which can be used in two scenarios: (i) debugging the environment and (ii) analyzing data from user studies in a research context.

Agarwal et al. [31] proposed a design and application space for visualizing data from mixed and virtual reality user sessions (Figure 10). By classifying existing visualizations into seven categories and combining them into specific visualization systems, the authors illustrate the potential of creating visualizations to support researchers in gaining insights on user behavior within mixed and virtual reality environments. This work focuses on two application examples: evaluating user interaction with a tangible virtual object and a collaborative problem–solving scenario to showcase how visualizations from the different categories could be combined and used in one debugging and one evaluation scenario. In the evaluation scenario, when a mistake occurs,

the developers can listen to the audio and playback the scene view to gain insight on what went wrong. However, these visualizations are not necessarily part of the Mixed Reality scene, but are used in a separate analysis interface contrary to immersive analytics, which leverage visual data representations in a 3D immersive environment.

2.2.2 Methods that Support the Evaluation of XR Applications

To evaluate the usability of the plethora of XR systems and applications that have been created during the previous years, researchers have conducted both expert and user-based evaluations. Cognitive walkthrough and heuristic evaluation are two characteristic methods that have been used in expert-based evaluations, with heuristic evaluation being one of the majorly adopted techniques. In this inspection method, an expert is asked to inspect the usability of the system based on their observation.

Indicatively, Zainuddin et al. [66] conducted a heuristic evaluation using thirteen evaluators to determine the design and usability issues of PekAR-Mikroorganisma, a courseware for the deaf in learning science by using Augmented Technology. Several experts from different fields were involved in this evaluation, considering the interface, pedagogy, AR environments and video sign language of the system.

On the other hand, in [29] cognitive walkthrough was used as a technique to oversee usability studies in a laboratory environment where the goal was to teach human body anatomy. Cognitive walkthrough identified a couple of design pitfalls and revealed many errors which were later addressed.

Some examples of diverse approaches that have been previously followed to evaluate the usability of XR applications in user-based

evaluations are interviews, NASA-TLX, the Think-Aloud technique and questionnaires. Interviews and questionnaires both fall under the category of enquiry methods, which are very subjective in nature. Olsson et al. [40] organized 16 semi-structured interview sessions with 28 participants in shopping centers to produce knowledge and requirements about potential end users' expectations of mobile augmented reality (MAR) services. The user experience categories and user requirements that were identified can serve as targets for the design of user experience of future MAR services.

Other research efforts [64] used the NASA Task Load Index version 1.0 (NASA TLX 1.0) [22] to measure the mental workload perceived by the subjects during two experiments. These experiments were testing the capabilities of two MR systems in a realistic environment and collaborative tasks against prevalent methods. The participants rated each of the six categories (mental demand, physical demand, temporal demand, effort, performance, and frustration level) based on their experience in the experiment using a 20-point scale and then performed pairwise comparisons.

Following a different technique, the work in [26] utilized the Think-Aloud protocol to create a remote ideation and usability method that establishes a surrogate relationship between participants and a facilitating researcher, who wears a VR headset, via video conferencing called Surrogate-Aloud. In this approach, the Think-Aloud protocol is applied by the participant to express movement and the interaction commands to be executed by the researcher, along with their thought process while they interact with virtual prototypes or scenarios.

In addition to the Think-Aloud technique, questionnaires are a dominant method in research to collect qualitative user data. Characteristically, the

authors of [50] conducted a study to design a usability questionnaire to assess the usability of handheld augmented reality systems reliability. They called the questionnaire scale as Handheld Augmented Reality Usability Scale (HARUS). HARUS consists of two sub questionnaires which were used for studying the comprehensibility and manipulability of handheld augmented reality (HAR) [14]. They analyzed the issues reported by previous studies and classified them as perceptual and ergonomic. This scale differs from the System Usability Scale (SUS) in that it is specific to handheld AR [54].

With questionnaires being a primary tool to evaluate the usability of an XR application, many researchers integrated their post-experiment questionnaires into the virtual world inside AR/VR headsets in order to provide a more immersive experience to their subjects. For instance, Schwind et al. [53] investigated the effect of filling out a questionnaire directly in VR. The participants were asked to fill out presence questionnaires in physical reality and in VR after exposing them to a VE at varying degrees of realism. However, no significant differences regarding presence between the 2 questionnaire modalities were found. Nevertheless, the data displayed higher consistency and lower variance when the questions were answered in the VE.

On the other hand, a significant effect on presence when questionnaires are integrated in the game context has been observed by other research efforts [20]. Characteristically, the authors propose the mapping of questionnaire elements to game elements, while maintaining consistency regarding both presentation and input. This concept was investigated with a prototypical serious game, and a user study was conducted comparing two variants of questionnaires during game play; an overlay questionnaire screen and the questionnaire being integrated

as a game element. Their results demonstrate that embedding question-items in the VE offers an opportunity to stay closer to the context of an ongoing experience, enhancing the levels of presence and the experience of the participants.

A systematic investigation into the effects of interrupting the VR experience through questionnaires was also performed, using physiological data as a continuous and objective measure of presence [45]. Similarly to the work of Schwind et al. [53], they created for their user study an immersive VE in which participants engaged in playful tasks at different levels of realism and responded to questionnaires inside and outside VR rating their player experience. At the same time, researchers recorded their physiological signals to assess the "breaks in presence" (BIPs). The outcome of the study suggested that questionnaires inside VR are less invasive than the ones outside and provide more reliable self-reports.

Furthermore, Wagner et al. [63] investigated how to best embed questionnaires within the VE and compared them with classical post-test evaluations regarding preference, presence, and questionnaire completion time. Their findings indicate a user preference as well as a higher interactive participant engagement with the questions' content for questionnaires inside the virtual environment, as opposed to those that are outside, suggesting an increase of accuracy in the questionnaires.

2.3 Discussion

A number of features that were extracted from the aforementioned tools and systems in the above literature review could be classified into two groups (general & specific), namely:

General features regarding employing XR technology:

A. Evaluation of XR applications.
B. XR in the pre-study phase.
C. XR to augment evaluator's view at runtime.
D. XR in the post-study phase (post-analysis).

Specific features regarding experiment functionality:

E. Interaction with virtual objects.
F. Multi-User: can be used by multiple XR users simultaneously.
G. Capture the experiment: audio, screenshots, video.
H. Replay experiment: virtual avatars, hands, objects, 3D trajectories.
I. Display participant's behavior measures: includes physiological variables, eye gaze, gestures, questions, movement in the environment.
J. Display participant's performance metrics: includes task duration, number of errors.

In Table 1 the features that each of the related systems includes are presented. From a brief glance on existing literature relevant to the adoption of XR technologies in evaluations, as well as the evaluation of XR applications, it is apparent that the majority of the research efforts concentrate on the visualization of various evaluation metrics and data inside the VE through XR technologies. Immersive analytics solutions have been the focal point of recent scientific endeavors, offering tools for in-situ and ex-situ analysis using AR/MR or VR devices alone or with complementary non-immersive means (e.g. desktop, tablet) as overview devices.

Some of the most frequently analyzed performance metrics include error/accuracy measures and task completion time, with most systems dealing with users' movement in the environment from the behavior metrics. The favorable virtual visualization method for position and distance data has been 3D trajectories [36], with works combining them with virtual avatars to show people's precise movements [37], or videos and 2D visualizations (e.g., heatmaps, scatterplots) in the VE [35]. Still, all these visualization approaches relate to the post-analysis process of an experiment (feature D).

Table 1: Comparison Matrix of systems using XR to support the evaluation process

	A	B	C	D	E	F	G	H	I	J
MRAT	✓			✓	✓	✓	✓	✓	✓	✓
MIRIA				✓	✓	✓	✓	✓	✓	✓
ReLive	✓			✓	✓		✓	✓		✓
AvatAR				✓	✓			✓	✓	
Flex-ER				✓						
IATK				✓		✓				
DXR				✓	✓	✓				
VRIA				✓	✓	✓				
CAPTURE	✓					✓	✓		✓	✓
Ubiq	✓	✓			✓	✓				
RealityFlow	✓	✓				✓				
ARgus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

As seen in Table 1, MRAT seems to have the majority of the described features in comparison to 'ARgus Designer', and supports the overall evaluation process of MR Apps. Nonetheless, it does not use MR in all the stages of the evaluation, neither in the pre-study phase nor at runtime. With respect to the remaining tools (i.e. MIRIA, ReLive, AvatAR, Flex-ER, IATK, DXR and VRIA) they are immersive analytics tools which employ XR technologies to support primarily the post-analysis phase an evaluation and only ReLive seems to be targeted to the evaluation of XR applications. CAPTURE supports collaboration using AR, as well as the evaluation of the collaborative process, but the evaluation phase does not look to be based on XR technologies. Furthermore, Ubiq and RealityFlow assist collaborative immersive authoring and both of them can be applied during the pre-experiment phase of the evaluation process.

The presented tools, depending on their characteristics and the evaluation phase they deal with, can effectively contribute to the evaluation of XR applications with their in-depth and analytical approach. However, none of these tools provides a virtual workstation in the augmented environment, offering the option to modify and set the experiments setting in the XR environment, without prior configuration in another device. This is important, as evaluators are able to conduct experiments and gather information important for user-based evaluations without having to switch the medium in which the experiment unfolds and simplifying the whole evaluation process (e.g. automatically record times). Furthermore, this liberates evaluators from transcribing raw data while synchronizing them to the experiment timeline. Our work aims to contribute filling this gap, by presenting a comprehensive approach in the evaluation process, as there is no AR/MR tool to our

knowledge that assists specifically AR/MR evaluations while augmenting the evaluators' view. By virtually displaying contextual information such as performance and physiological metrics of the participant, evaluators can have a clearer picture of the experiment at runtime.

Chapter 3

Design Process

This Chapter presents the design methodology process of ARgus Designer and outlines the functional requirements of the system along with the motivational scenario that inspired this work.

3.1 Methodology

An iterative design process was followed while designing the system, based on the Design Thinking methodology [42]. This methodology consists of five phases, namely 'Empathize', 'Define', 'Ideate', 'Prototype' and 'Test', which are not always sequential, and teams often run them in parallel, in varying orders and repeat them iteratively (Figure 11). In a series of meetings with a number of potential end-users (researchers and developers of AR/MR applications, and applications for intelligent environments), a set of scenarios and personas (Section 3.3) were selected for the 'Empathize' and 'Define' steps of Design Thinking. For the 'Ideate' step, an extensive literature review was conducted and multiple brainstorming sessions were organized with AR/MR application

experts, interaction designers, UX designers and evaluation experts. The gathered ideas were then filtered, in view of their feasibility, through interviews with a different group of domain experts. Experienced interaction designers and developers additionally reviewed the ideas and offered important feedback such as insights, comments, and preferences regarding which ideas would be the most innovative, scientifically interesting, useful and accepted by potential end-users of the system. A set of final system requirements (Section 3.2) was then generated, constituting the final document of high-level functional requirements for the ARgus Designer tool.
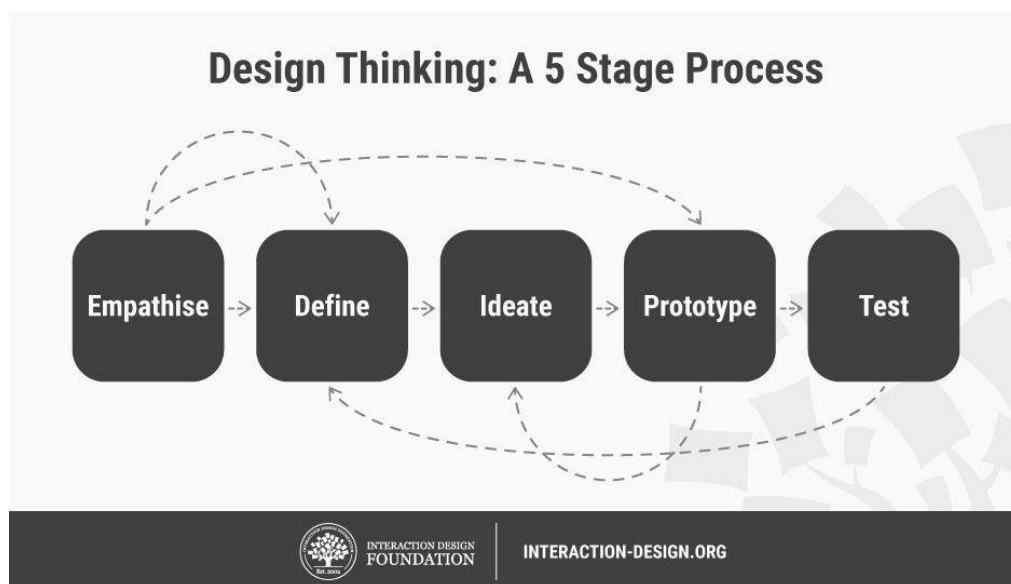
Figure 11: Design Thinking Process [59]

The 'Prototype' phase began by creating a low-fidelity representation (Figure 12) of the overview of the ARgus Workstation and its functionality. Higher-fidelity prototypes followed (Figure 13), conforming to the most promising ideas from the 'Ideate' phase. Finally, an expert-based evaluation with the participation of six (6) User Experience (UX) experts was conducted, aiming to assess the overall concept of the

system, detect valuable unsupported features as well as uncover possible usability errors.

Figure 12: Low-fidelity prototype of the ARgus Workstation
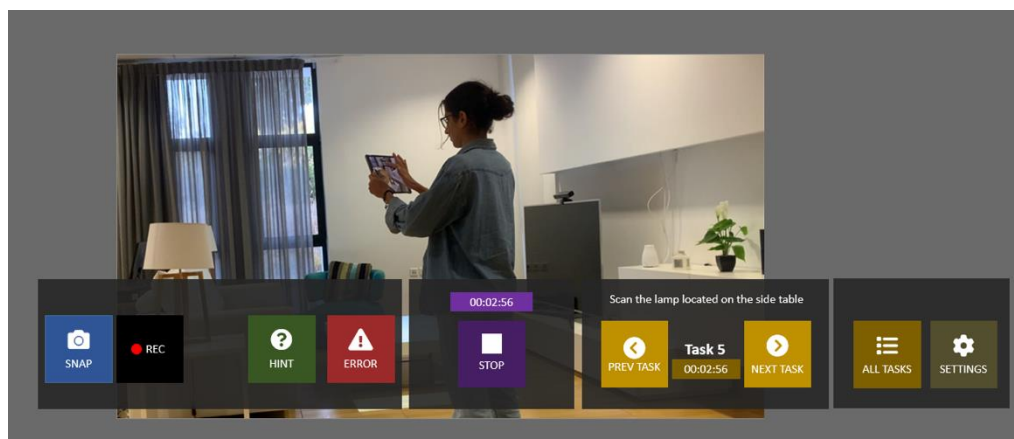
Figure 13: High-fidelity prototype of the ARgus Workstation

3.2 Requirements Elicitation

This section presents the functional requirements that the system needs to satisfy. Such requirements were collected through firstly conducting an extensive literature review (Ideate Phase of the Design Thinking

Methodology). By reviewing recent literature, which aimed to assess the way usability could be measured in Augmented Reality applications [14] [7] [16], some common themes and requirements for conducting such studies have emerged. Specifically, one of the most cited works regarding user-based experimentation in Augmented Reality [16] established three types of user-based experiments that are relevant to AR research:

- Human perception and cognition.
- User task performance and interaction techniques.
- User interaction and communication between collaborating users.

Based on the above and by further examining the framework described in [35], it was revealed that in order to assist researchers in conducting user studies, the evaluation tools should provide appropriate visualization of performance measures, behavior measures and post-analysis features (Table 2). The performance measures include the overall time and duration of specific events, the time when a task is started or completed, the frequency of using a specific feature and the time spent on a task. The behavior measures comprise the physical movement of the user around the environment, the number and type of hand gestures and the physiological variables and emotions of the user, while the post analysis features contain the video and audio recordings along with screenshots of the augmented content of specific events and moments. Consequently, in order to augment the evaluators' view and relieve them from the confinement of the limited human information processing capacity, such information must be properly presented to evaluators during the study. This will help them acquire a comprehensive view of as well as optimize their performance in the evaluation process.

Table 2: Data that must be available to evaluators during and after a user study. Figure inspired by [17] [18]

Performance measures	Behavior measures	Post-analysis
Overall time & duration of specific events	Physical movement around the environment	Video recording
When a task is started/completed	Number & type of hand gestures	Audio recording
Frequency of using a feature	Physiological variables & emotions	Screenshots of the augments content
Time spent on a task		

The literature review was followed by an iterative process (Ideate and Define Phases of the Design Thinking Methodology) based on multiple collection methods such as brainstorming, focus groups, observation and scenario building, which resulted in the functional requirements for ARgus Designer described in Sections 3.2.1 and 3.2.2.

3.2.1 Requirements for Evaluators

From the aforementioned process, eleven (11) requirements arose for evaluators using the system. These requirements span through all the phases of the evaluation process, pre-study, during runtime and post-study.

Requirement 1: Configure the evaluation scene a priori

The evaluator can digitally augment the physical environment with interactive 3D objects a-priori and load the configuration at run-time.

- The evaluator should be able to add/remove digital artifacts.

- The evaluator should be able to add/remove mockups.

- The evaluator should be able to add/remove prototypes.

Requirement 2: Set experiment settings

The evaluator can set the experiment's settings and choose the information they want to be displayed before the beginning of the experiment.

- The evaluator should be able to set the information they want to be displayed during an experiment regarding the participant

    - Show/hide the participant's bio-signals (heart-rate).

    - Show/hide the participant's behavior measures.

    - Show/hide the participant's field of view.

    - Show/hide the participant's finger touch point.

    - Show/hide the participant's real-time comments.

    - Show/hide the participant's emotions.

Requirement 3: Capture the experiment

The evaluator can capture certain events and moments of the experiment by either taking screenshots or video recordings.

- The evaluator should be able to make a video recording at run-time.

- The evaluator should be able to take a snapshot at run-time.

Requirement 4: View & Manipulate Quantitative Data

The evaluator can view and manipulate quantitative data during an experiment.

- The evaluator should be able to start/stop a virtual timer.

- The evaluator should be able to see the time elapsed of the overall experiment.

- The evaluator should be able to see the errors made by the participant.

- The evaluator can provide hints to the participants and increase the respective metric.

- Show/hide the current scenario steps.

Requirement 5: View Qualitative Data

The evaluator can view qualitative data from the participant during an experiment.

- The evaluator should be able to see the participant's comments during an experiment.

Requirement 6: View and Navigate to the Scenario Steps

The evaluator can view useful information regarding the flow of the experiment and navigate to the scenario steps.

- The evaluator should be able to look at the current script.

- The evaluator should be able to see the time elapsed of the task.

- The evaluator should be able to see the list of all the tasks the participants have to complete.

- The evaluator should be able to see the next steps of the participant.

Requirement 7: Observe the participant's interactions

The evaluator can observe participants' interactions with the environment as well as behavior measures during those interactions.

- The evaluator can monitor unobtrusively the interactions of one or more participants.

- The evaluator should be able to see behavior measures of a participant (e.g. gestures, eye gaze, movement and virtual objects).

- The evaluator should be able to see the participant's field of view.

- The evaluator should be able to see where the participant is pointing.

- The evaluator should be able to see the participant's touch position.

Requirement 8: View participant's Bio-signals

The evaluator can view in real-time the participant's bio-signals and stress levels during the experiment.

- The evaluator should be able to see the participant's bio-signals.

- The evaluator should be able to see the stress levels.

Requirement 9: Interact with the XR environment

The evaluator can interact with the virtual environment in order to provide hints and help participants during an experiment.

- The evaluator should be able to interact with/manipulate virtual interactive artifacts during an experiment (e.g. move, highlight, interact).

- The evaluator should be able to provide hints by manipulating virtual artifacts.

Requirement 10: Replay an evaluation experiment in XR

The goal is to offer rich information to evaluators after an experiment. The evaluator can choose to see useful and helpful information after experiments.

- The evaluator should be able to see all video recordings post-experiment.

- The evaluator should be able to see all snapshots post-experiment.

- The evaluator should be able to see all the information regarding a user experiment post-evaluation.

- o The duration of the overall experiment and of each scenario task.

- o All scenario tasks.

- o Participant's errors on each task.

- o Peculiar bio-signals (e.g. when participants were stressed).

- The evaluator should be able to replay another user's actions

Requirement 11: Observe participants filling XR questionnaires

The evaluator can observe the answers of the participant in real-time while they are filling post-evaluation questionnaires through their MR glasses.

- The evaluator should be able to view in real-time the participants' responses.

- The evaluator should be able to mark certain questions to be discussed during debriefing.

3.2.2 Requirements for Participants

For the participants using the system, three (3) requirements were elicited. These requirements span only during the evaluation, as no participant input is required in the other two phases.

Requirement 1: Interaction with virtual 3D objects

The participant can interact with virtual interactive artifacts & digital UIs in the AR/MR world.

- The participant should be able to interact with virtual interactive artifacts.

- The participant should be able to interact with digital UIs.

Requirement 2: Simultaneous use from multiple users

The system can be simultaneously used by multiple users, providing means for user collaboration.

- The participant should be able to ask for hints.

- The participant should be able to receive help.

Requirement 3: Fill-in XR questionnaires

The participant can answer post-evaluation questionnaires through their MR glasses and evaluators can observe their answers in real-time.

- System Usability Scale (SUS).

- Short User Experience Questionnaire (S-UEQ).

- NASA Task Load Index Questionnaire (NASA-TLX).

3.3 Motivating Scenario

Scenario building is a broadly-used requirements elicitation method [11], evoking reflection-in-action [51], which can assist the process of developing requirements. A scenario is a detailed description of an event [65], and in our case of the user and their tasks in a specified context, offering a representation of the user interacting with the system to achieve a specific objective. They are primarily used in the early phases of design, aiming to elicit end-user requirements and usability goals. This section introduces a scenario that motivated the design of the system, where an evaluator interacts with the ARgus Designer tool to conduct a user-based evaluation of an intelligent environment.

Scenario

Anna is a UX designer who wants to evaluate a new MR system that she just created for the intelligent home. In particular, she wants to virtually assess the placement and design of the digital artifacts she developed, along with their respective interactive prototypes, before proceeding with their implementation and physical installation. Consequently, she decides to utilize the ARgus Designer System to assist her in the execution of a user-study, where the participants will have to interact with the virtual objects that she developed in order to complete some tasks, while following a predefined scenario. In this study, the digital artifacts to be evaluated are the intelligent fridge, the smart blinds, the main display of the living room and the microwave oven. The procedure she will have to follow is to: firstly, prepare the environment and set the scene in which the participants will follow the scenarios; then, run the study session with the participants; and, finally, process and analyze the data that the tool accumulates.

3.3.1 Pre-study, Preparatory Activities

Anna launches the application in the "Editor" mode, and selects to create a new project from the projects window. She then proceeds to place the virtual artifacts and objects that will be evaluated in the augmented environment. Afterwards, she selects the 'Bio-Signals' and the 'Touch-Point' from the study settings as supplementary information about the participant to be virtually displayed in real-time during the study session. Having prepared the scene and settings of the study session, Anna saves the configuration, names it 'Smart Home' and closes the application.

3.3.2 During-study, Main Experiment

The day of the study, Anna has prepared the space of the intelligent living room where the study will take place and has both her device and the device that the user will be wearing standby. After the participant, Andy, arrives, Anna starts explaining the procedure and they both wear the HoloLens device. As soon as everything is ready, Anna launches the study session on her device and informs the participant that he can begin and press the start button whenever he is ready. A few moments later, Andy presses the button and the study session officially starts, with him having to follow the scenario tasks; at the same time, the system automatically initiates the global study timer as well as the task timer. While the participant tries to complete the first task, Anna decides that it would be also useful to see the field of view of the participant, as he interacts with the environment, so she updates the study session settings. Shortly after, Anna wants to recall the upcoming tasks of the scenario, so she decides to have a look at the script's tasks from the ARgus "workspace", a virtual toolbox from where the evaluator can control the available supporting functionalities.

During the study session execution, Anna monitors in real-time what the user is doing, his progress in the script, the amount of time spent on every task, what has been done wrong and how stressed the user feels (from his heart-rate changes). When the participant completes a script task, Anna proceeds to the next task in the study session from the ARgus "workspace". The system automatically records all the necessary performance and behavioral measures, but Anna observes that the user is feeling stressed because of the peculiar interaction paradigm used in that particular part of the scenario, so she decides to take a screenshot of that moment to help her later during post-examination. As this event

has affected the flow of the experiment, Andy asks for a hint to complete the following task and Anna provides him with one. She then, additionally, starts recording the participant trying to complete that specific task with the extra help she provided.

At the end of the study session, Anna observes the user answering the two post-study virtual questionnaires, a SUS and a short-UEQ questionnaire, and gets notified about any odd responses. When Andy completes the questionnaires, Anna follows with an additional short interview about the usability and the UX of the intelligent environment, with the system recording and transcribing their interaction.

3.3.3 Post-study

After the study, the examination process begins with Anna analyzing the accumulated data from all the participants. While she is inspecting Andy's results, she observes the screenshot of the peculiar interaction paradigm Andy used so she decides to delve deeper and replay the event using AR to relive that moment, so as to better understand what happened. She selects the timeframe of the snapshot and examines the "virtual user's" interactions with the environment along with other physiological measures (e.g. bio-signals, emotions). By reliving that event, Anna derives that Andy became particularly stressed when he could not perform complicated gestures while interacting with the environment. Consequently, she decides to enrich the UI with controls that the user can easily use instead of gestures.

Completing the post-study analysis process, Anna was able to compare the interactions of the study participants with the display of the fridge by reliving that event for each participant. She noticed that all participants struggled with that specific interaction, so she infers that the UI of the fridge display need to be modified in order to simplify the interaction.

Chapter 4

The System

4.1 Description

The system's functionality is classified into three temporal categories: pre-study, during-study and post-study (Figure 14). In particular, the system enables evaluators to configure an evaluation scene ahead of a user study, by virtually augmenting the physical environment with interactive 3D artifacts and mockups a-priori and load the configuration at run-time. Specifically, the system can superimpose virtual artifacts in the physical environment, along with their respective digital interactive UIs, as well as digital 3D objects with which the user can interact and trigger various events within the Intelligent Environment. That way, ARgus Designer (Figure 15) can be considered as a meta-design tool [17], since it facilitates the a-priori configuration of scenes that an expert might want to evaluate during a study. Additionally, during the pre-study phase, evaluators can also set the experiment settings and select the dynamic contextual information (e.g. bio-signals, behavior measures,

touch point, user's field of view, comments) about the participant that is going to be displayed in real-time.
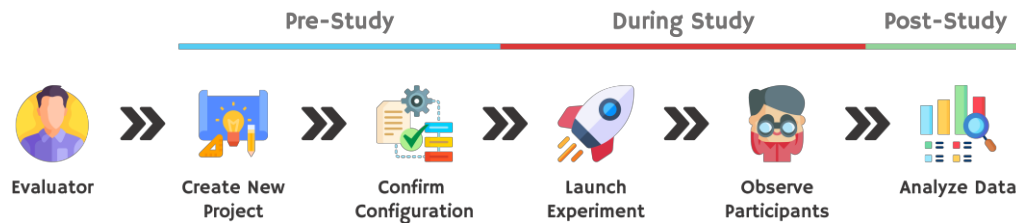
Figure 14: ARgus Designer concept visualization

With respect to UX evaluation, during the study, the system augments the evaluators' view by virtually displaying various information that the user can enable/disable via the settings menu. In particular, evaluators can view and manipulate quantitative data, such as number of errors and hints, experiment duration, task duration, etc. Additionally, around the participant's body, ARgus can superimpose their bio-signals, various psychophysiological and behavior measures (e.g. stress levels) and even interaction related data (e.g. performed gesture). A control panel (Figure 23) placed at the bottom of the user's field of view, contains simple color-coded buttons that permit the recording of critical events (e.g. error occurred, hint provided), and the initiation/termination of various functions (e.g. start timer). Moreover, supplementary functionalities include video recording, taking snapshots and browsing through the list of the available evaluation tasks.

Regarding user participants, the system enables them to interact with virtual interactive artifacts and digital UIs in the augmented world, while it also permits them to answer the post-evaluation questionnaires directly through their Mixed Reality glasses by virtually displaying them at the end of an experiment; at the same time the evaluators are able to observe their responses in real-time.
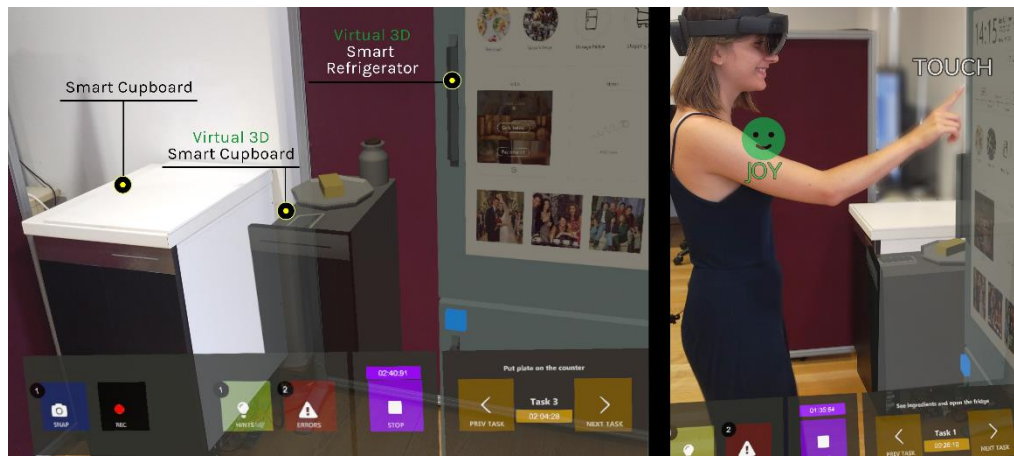
Figure 15: Snapshots of the ARgus Designer System. On the left, the ARgus Workstation is visible as well as the virtual artifacts integrated into the physical environment; on the right, some features that augment the view of the evaluator regarding the participant are displayed, i.e. the feelings of the user and the recognized interaction.

Following the study, during the post-study phase, the system permits evaluators to relive an evaluation study in a Mixed Reality environment by displaying the snapshots, the video-recordings and important study data (e.g. times, errors, hints), and by replaying the participants' actions.

The current status of the system's implementation is outlined: in Table 3 with respect to the requirements for the evaluators, and in Table 4 concerning the participants of a user study.

Table 3: Current implementation status of the requirements elicited for evaluators

	Fully Implemented	Partially Implemented	Future Work
Requirement 1	✓		
Requirement 2	✓		
Requirement 3	✓		
Requirement 4	✓		
Requirement 5			✓
Requirement 6	✓		
Requirement 7		✓	
Requirement 8	✓		
Requirement 9	✓		
Requirement 10		✓	
Requirement 11	✓		

Table 4: Current implementation status of the requirements elicited for the participants

	Fully Implemented	Partially Implemented	Future Work
Requirement 1	✓		
Requirement 2		✓	
Requirement 3		✓	

4.2 System Architecture

Figure 16 presents the Architecture of ARgus Designer. The components belonging to the frontend (Section 4.2.5) and backend (Section 4.2.6) of the ARgus Designer constitute the core of this work. Through the AmI Solertis Framework (Section 4.2.1) the system is able to communicate with Intelligent Environments (Section 4.2.3) and utilize existing intelligent services (Section 4.2.4) so as to collect the required information for improving the evaluator's view. Additionally, the Wizard of AmI tool (Section 4.2.2), empowers ARgus Designer by providing access to a pool of available mockups for building the study scene, and to the list of tasks that constitute the scenario of the evaluation study.
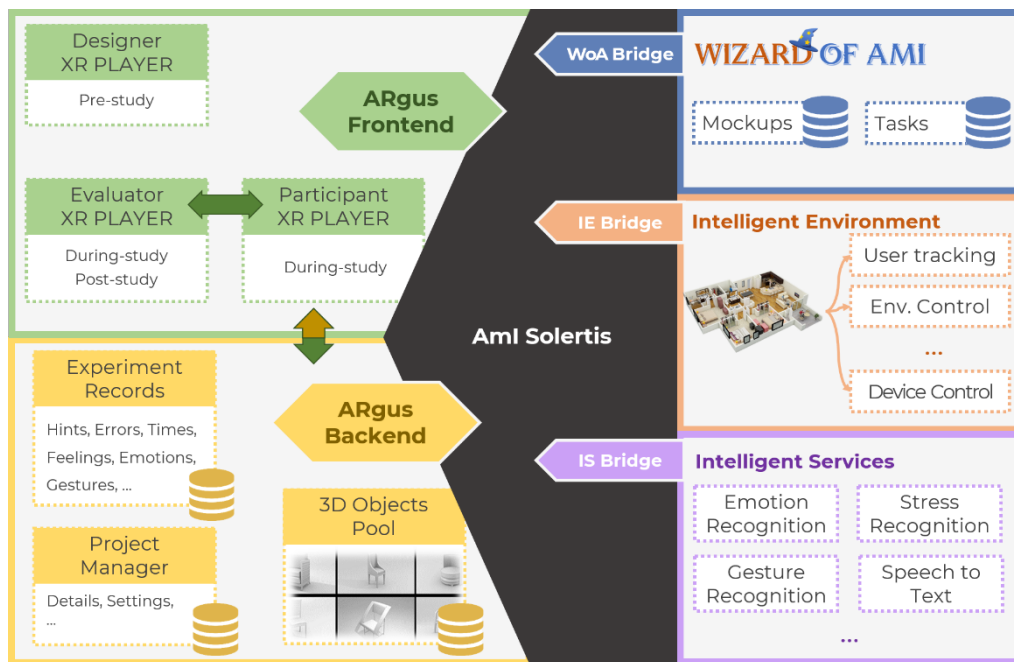

Figure 16: Architecture of the ARgus Designer system.

4.2.1 AmI Solertis

The AmI-Solertis framework [32]: (i) introduces a unified Hybrid Communication protocol which supports synchronous, asynchronous and event-based communication; (ii) unifies the definition and introduction of new devices, services and software components; (iii) facilitates the integration and usage of heterogeneous services in a standardized - yet agnostic – manner; (iv) delivers a scripting mechanism that can dynamically adapt the execution flow and govern the behavior of the entire intelligent environment; and (v) offers a standard library of tools (i.e., Analytics and History, Fault Tolerance, Storage Management, Common Utilities) that developers can use. AmI-Solertis empowers ARgus Designer to interoperate with the digital services and technologically enhanced physical infrastructure of intelligent environments.

4.2.2 Wizard of AmI

'Wizard of AmI' [3] is an online platform for desktops and mobile devices, which supports the design process of applications for Intelligent Environments by enabling the creation of interactive, prototype-based, scenarios. It targets designers of interactive context-sensitive (e.g. sensors, smart services) applications for Intelligent Environments, which can be used via multiple displays at the same time. Through the system, the designers create the user scenarios by mapping a series of mockups to the environment (i.e. select what is to be presented and where). Next, through a novel graph-inspired interface, they define the flow and interactivity of the scenario (i.e. when it is to be presented). During on-site scenario enactment, following the established 'Wizard of Oz' technique, the mobile companion application allows the designers to control its execution flow on demand.

4.2.3 Intelligent Environment

The ARgus Designer system can exploit the various services of the intelligent environment where it is deployed. In particular, it can make use of a user tracking service for locating one or multiple users (e.g. display user's heart rate next to his/her body) in the room, or a home control service to set devices' state as required (e.g. turn on the virtual coffee maker when the user opens the physical smart kitchen cupboard). Therefore, ARgus designer can manipulate the Mixed Reality environment and subsequently affect the overall experience.

4.2.4 Intelligent Services

External intelligent services can be utilized to gather input regarding the participant to the system. Characteristically, to measure participants' bio-signals a heart rate monitor (HRM) can be allocated to them to record their electrical activity of their heart (ECG), as well as a stress monitor that measure electrodermal activity (EDA) or heart rate variability (HRV). Another service regarding a participant's physiological measures include emotion recognition devices which aim to identify how the participant of the study is feeling based on a number of metrics (e.g. facial expressions, brain activity, bodily cues, etc.). Finally, the services that can be exploited are: a) gesture recognition technologies which interpret the participants' position, orientation and hand movements and eye tracking devices, and b) Speech Recognition and Speech-to-Text services for appropriately recording user comments.

4.2.5 ARgus Frontend

The frontend part of the ARgus Designer system was built using Unity and provides a user interface, which has been implemented as a single application. It supports three modes: the 'Designer XR Player', the 'Evaluator XR Player' and the 'Participant XR Player'. Via the interface

51

for designers, users can: (i) create a new MR scene, (ii) construct a configuration, and (iii) select the contextual information to be virtually displayed at runtime (Section 4.3.1). The user interface for evaluators provides features that cover the runtime of an experiment along with post-study phase of the evaluation process. More specifically, the interface includes a virtual workstation during the study phase which: (i) accommodates experiment data acquisition; (ii) manages the experiment's flow; (iii) presents an overview of the scenario steps; and (iv) enables the alteration of the augmented contextual information displayed and the existing scene configuration (Section 4.3.2). For the post-study phase the interface features: (i) an overview of the gathered data, metrics and captured media, and (ii) the option to relive an experiment or a specific event.

On the other hand, the user interface for the participants that use the system during a study, consists of: (i) a starting waiting scene, and (ii) the option to begin the study. The interfaces for evaluators and participants are interconnected, as actions in one of them directly affect the other in real-time. In particular, the capabilities of the Vuforia Engine library are utilized for the AR tracking to create a shared reference point among the evaluator and participants and anchor virtual game objects in the MR world.

4.2.6 ARgus Backend

The backend part consists of three distinct components, the 'Experiment Records', the 'Project Manager' and the '3DObjects Pool'. These backend components communicate through Web-based services with the frontend of the system. The first one is a database where all raw study data are stored during an evaluation such as performance metrics (e.g. errors, hints, times), physiological metrics (e.g. bio-signals,

emotions, gestures), as well as the captured photos and videos. The 'Project Manager' component is a database where all the details about a particular project are collected. The augmented contextual information selected, the configuration of the virtual artifacts in the scene and the AR questionnaires to be administered are some characteristic examples. Finally, the last backend component is the database where the pool of the 3D objects that are designated for each project is saved.

4.3 Design Rationale

The ARgus Designer system provides an environment for evaluators to create MR scenes, conduct user-based evaluations in them, while augmenting their view, along with a space for analyzing useful experiment data and metrics. Regarding its context of use, the system provides an immersive experience to its users with both evaluators and participants needing to wear MR glasses, specifically the HoloLens 2, to interact with it. The expected interactions among its users, while using the system, are the provided augmented tools (e.g. the ARgus Workstation) and the virtual artifacts it contains.
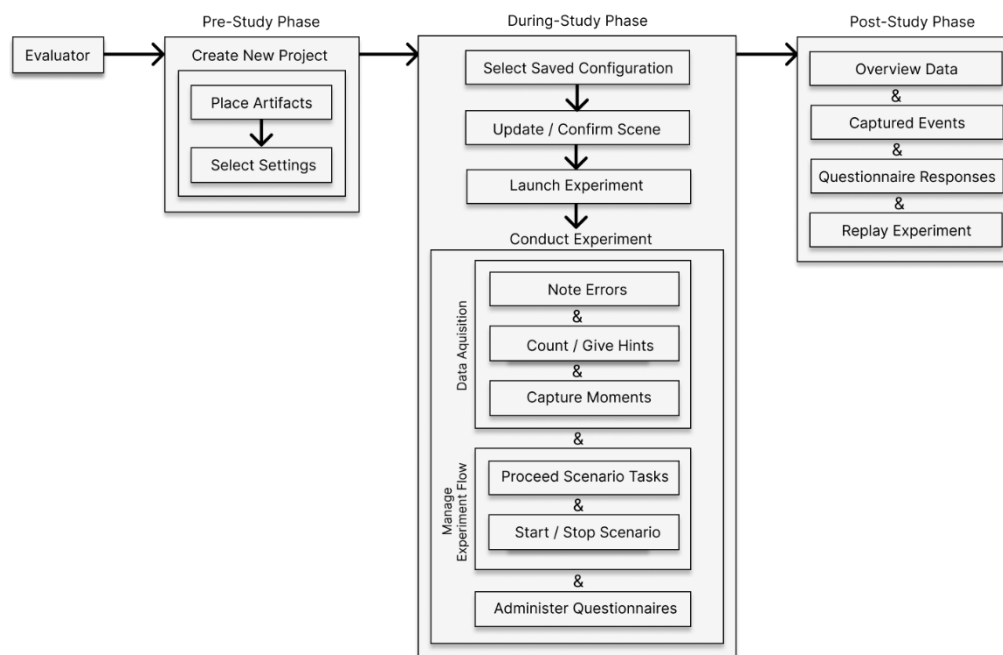
Figure 17: Flowchart of the UI Design process.

This evaluation tool aims to be a holistic solution in the evaluation process catering to all its phases: pre-study, during study and post-study. Particularly, evaluators before conducting a study session, they can select to create a project where they will initialize their augmented scene constructing their own mixed-reality IE, with the desired digital artifacts, interactive mockups and questionnaires to be administered. The additional metrics and features that will augment evaluators' in the course of the study session can also be selected at this stage. Right ahead of a study session, evaluators can choose the saved configuration and confirm the settings and structure of the scene to be evaluated. Participants wear their MR headset and wait for the evaluator to launch the study session. During the evaluation, the evaluator can observe the participant interact with the augmented environment, while the system augments their view with the selected information. Evaluators can additionally utilize the workstation that is supplied by the system to

gather important data. When the study's scenario steps have been completed, AR questionnaires can be administered to the participants with the evaluators' supervision. After the conclusion of the study, evaluators can analyze the study sessions having an overview of the data/metrics, a collection of the captured events, the potential AR questionnaire responses and the possibility to replay a study session (Figure 17)

4.3.1 Pre-Study
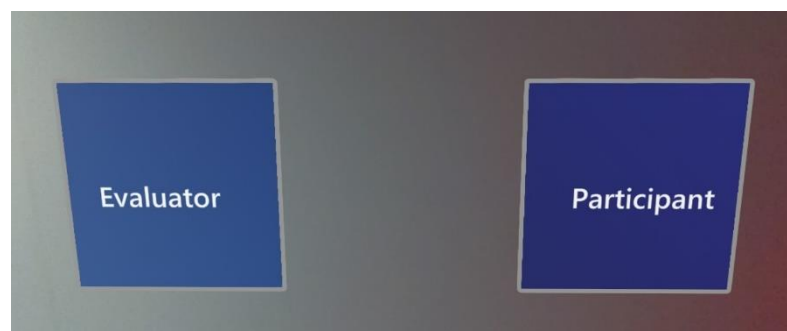
Figure 18: Starting scene of the application

The first UI when the ARgus designer application is launched is the one in Figure 18, where the user has to select from two buttons their respective role. With the first button (left) the user assumes the role of the evaluator, while with the second one (right) the user obtains the role of the participant in the study session. The role assignment process was chosen as the initial UI to minimize the steps the evaluator has to take during setup as no prior preparation is needed on the device of the user before conducting a study session. The majority of the UIs of ARgus designer concern the evaluator and aim to augment their view and provide valuable information during a user-study.

If the user is the evaluator, the second UI displayed to them is a 'Projects' window as shown in Figure 19. An evaluator might use the application to create a new project for a future study (1) or choose to load an existing

configuration/project (2) either before a study to make alterations and update the configuration or to conduct an experiment.
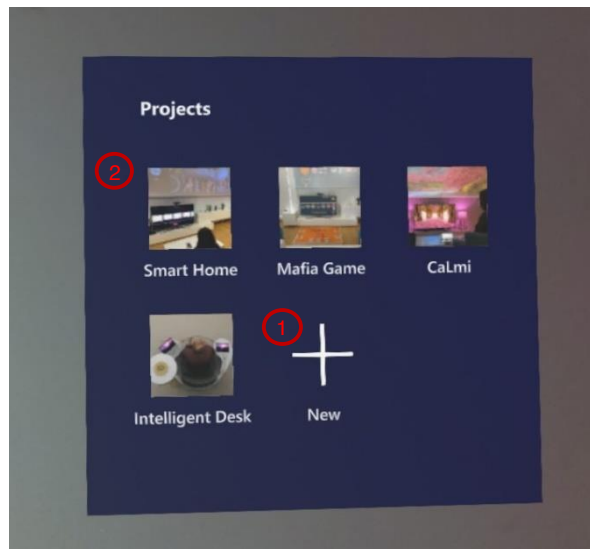
Figure 19: Projects Window displayed to evaluators

When the user selects to create a new project, they are presented with the study settings (Figure 20) and the available artifacts (Figure 21) to place in the augmented environment. The study settings contain the six features that augment evaluators' view during a study, a 'RESET SCENE' button that resets the augmented environment of any augmented artifact and 3D object configuration and a 'LAUNCH' button that commences the study process with immobilizing the virtual objects into their set positions and saving/setting the selected features.

In particular, the six features (from left to right) are, namely the 'Bio-Signals', 'Behavior Measure', 'Touch-Point', 'Field of View', 'Emotions' and 'Comments'. First, the bio-signals feature displays the heart-rate of the current participant and can help predict their stress levels. Second, the behavior measure feature shows extra information regarding the participant's gestures, eye gaze and movement. Third, the touch-point

feature annotates where the participant has touched, which can be extremely useful when they are interacting with UIs. Then, the field of view feature indicates where the participant is looking when they are interacting with an object of interest (e.g. virtual artifact, interactive UI, 3D object). Finally, the emotions feature displays the current emotional state of the participant and the comments feature shows the participant's remarks while they are executing a task.
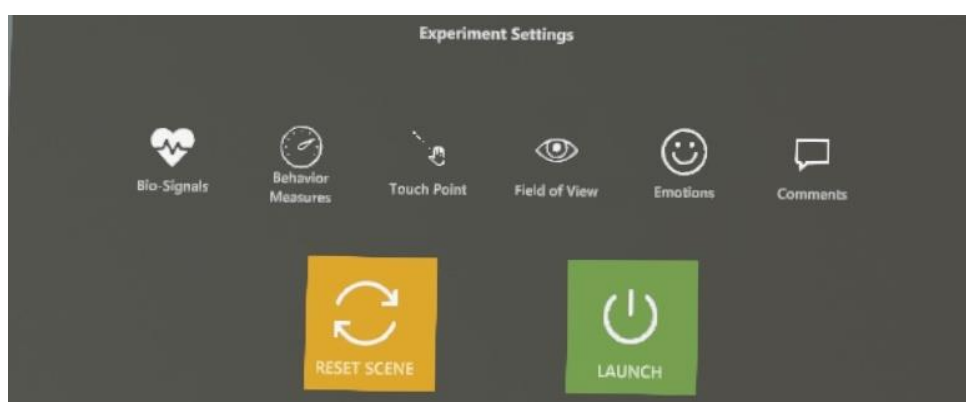
Figure 20: Default experiment settings on a new project

Regarding the artifacts list (Figure 21), the evaluator can explore, with the arrow buttons (1, 2), and select which of the available virtual artifacts they wish to place into the augmented environment for the current project.

During the creation or modification of a project, the six features that augment evaluators' view can be toggled (Figure 22 (1)), with the system saving their state for the execution of the study session or the next time the project is launched. In addition, the evaluator can see the times each virtual artifact (Figure 22 (3)) has been used in the augmented scene (Figure 22 (2)). Given the fact that ARgus designer can support and augment large spaces, it is useful to know the number of object instances in the environment.
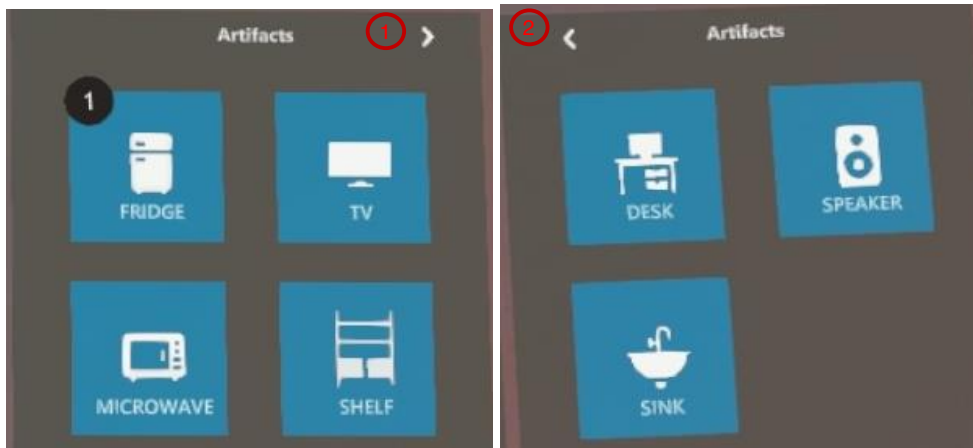
Figure 21: Artifact list UI on a new project

Figure 22: Augmented environment, experiment settings & artifact list during the creation of a new project.

4.3.2 During Study

When a study session is taking place and the participant of the study launches the application through their 'role', they are presented with a loading scene as shown in Figure 24 (left). In this stage, they are prompted with a 'Loading Experiment Scene' message along with

orbiting circles, while the evaluator is finishing setting up the current project.

After the evaluator presses the 'LAUNCH' button, the study moves to the 'During Study' mode. The settings (Figure 22 (1)) and artifact list (Figure 22 (2)) are hidden and the 'ARgus Workspace' (Figure 23) appears. When the evaluator presses the 'START' button (Figure 23 (1)) and the global experiment timer begins, the UI of the participant changes from the loading scene to the one in Figure 24.
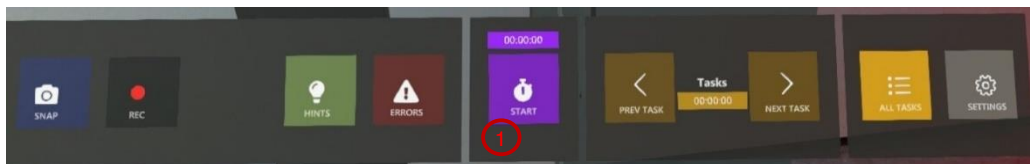
Figure 23: ARgus Workspace before officially starting experiment

The 'Start' button depicted in Figure 24 (right) will officially start the experiment for both users and initiate the experiment and task counter. These two start buttons launch different timers for the experiment with the first (Figure 23 (1)) considering possible introductions and explanations that might be needed for a study, considering their importance to experimenters. The second (Figure 24) initializes the official study timer, which accounts for the total time it takes a participant to complete all the scenario tasks of the evaluation; simultaneously launching the first task's timer as well.
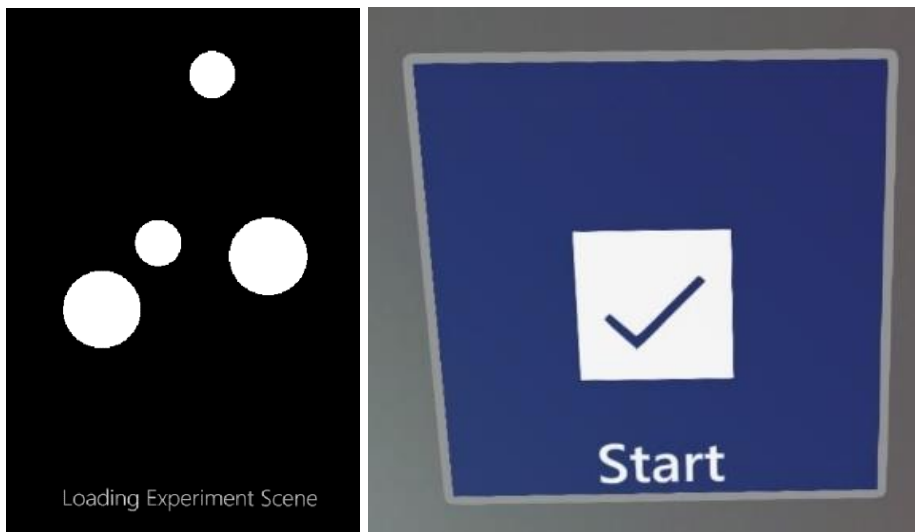
Figure 24: Participant waiting scene while evaluators are setting up (left) and button to start experiment (right).
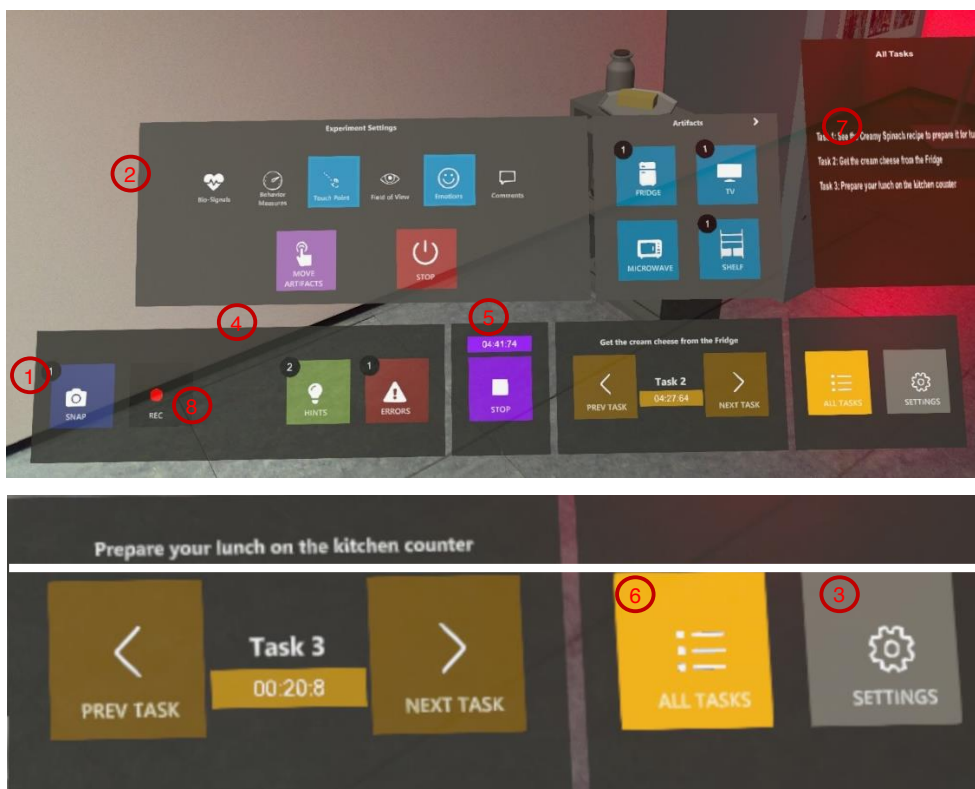
Figure 25: ARgus Workspace, study settings & the augmented scene during a study session with tasks open and artifacts movable in the scene (top), zoomed info & buttons of tasks and settings (bottom).

While a study session is running, the 'ARgus Workspace (1) in Figure 25 is always available to the evaluators in their peripheral view. This Figure shows indicative examples of the UIs of the 'ARgus Workspace' (1) and the study settings (2) as well as the augmented scene with the virtual artifacts. In this case, the study settings (2) are toggled from the 'SETTINGS' button in the 'ARgus Workspace (1) and showcase two new buttons, namely the 'MOVE ARTIFACTS' (4), which enables the virtual artifacts to be moved in the augmented environment, and the 'STOP' (5) which permits the evaluator to force-quit this mode and end the study session. Additionally, the 'ALL TASKS' button (6) in the 'ARgus Workspace' (1), shows a virtual textbox (7) of all the tasks the participant has to complete in the current scenario.

Figure 26: Functional buttons in the ARgus Workspace (from left to right): to take a screenshot, pause/stop an active recording, increase the hint counter, increase the error counter.

In further detail, the first two functional buttons of the 'ARgus Workspace' concern capturing the HoloLens glasses' view the and consist of the 'SNAP' (Figure 26 (1)), which allows the evaluator to take a screenshot of after a few seconds and the 'REC' (Figure 25 (8)), that begins a video recording and after its selection provides a pause and stop button along with a timer. Then, the third and fourth functional buttons are the 'HINTS' (3) and the 'ERRORS' (4) which act as counters of the amount of hints

61

the evaluator had to give to the participant and the mistakes the participant made accordingly.

Figure 27: Global timer & task timer and handler when experiment is running (left), Global timer when study has been paused by the evaluator (right)

Moreover, an evaluator can view and manipulate time related data in the 'ARgus Workspace' UI. Specifically, the system enables them to stop (Figure 27(1)) and re-start (Figure 27 (2)) the study's global timer in the event that the flow of the study session must be stopped due to an unpredictable cause. Moreover, information such as the current task and its duration are also accessible to evaluators in the virtual workstation along with functional 'PREV TASK' and 'NEXT TASK' buttons that can manually change the ongoing scenario task.
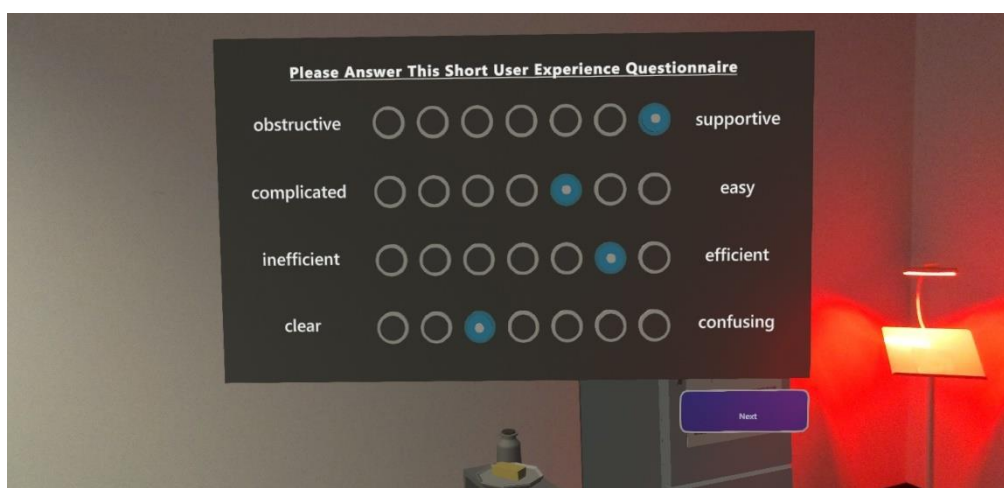
Figure 28: Example of the AR Questionnaire (Short UEQ).

Finally, the evaluator can observe participants answer a virtual questionnaire in the MR environment after the corresponding participant has concluded all the necessary scenario tasks. Figure 28 displays an example of an augmented Short – User Experience Questionnaire that is automatically issued after the completion of the scenario steps. Evaluators can navigate through the responses, during the session while the participants are filling them in, as well as post study.

4.3.3 Post-Study
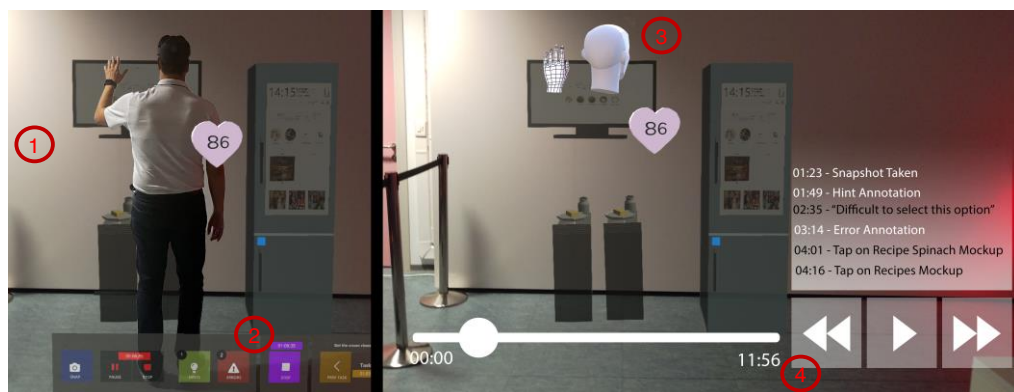
Figure 29: Mockup of the post-analysis phase offered by the ARgus Designer system. Evaluators can review snapshots that were captured during the experiment (left) as well as relieve a specific moment of the experiment (right).

Regarding the post-study phase of the evaluation process, the evaluator can retrieve all captured media (e.g. snapshots, videos) that were taken during the study sessions (Figure 29 (1)) as well as examine a specific user's performance metrics (e.g. errors made, hints given, times to complete each task etc.) (Figure 29 (2)). In addition, the system offers a replay option, where evaluators can relieve a study session. Evaluators by going to the site of the study can virtually replay the participant's interactions with the environment. The movement of the head and hands of the participant is virtually depicted in the MR world through the application along with other selected contextual information about the participant (Figure 29 (3)) such as their heart-rate. Finally, the ARgus

Designer system permits experts to relive a user-study from beginning to end but can also replay specific events of interest, e.g. after a snapshot was taken, when the user made an error or needed hint etc. (Figure 29 (4)).

Chapter 5

Cognitive Walkthrough Evaluation

A cognitive walkthrough evaluation study of the ARgus Designer system was conducted with the participation of six (6) User Experience (UX) experts with prior experience in designing systems for Intelligent Environments. The evaluation process was conducted at the Human-Computer Interaction Laboratory (HCI) of the Institute of Computer Science of the Foundation for Research and Technology - Hellas (ICS-FORTH). Cognitive walkthrough [47] is a technique for evaluating the overall user interface of a system and focuses on how a system is perceived by users the first time they use it without formal training. Its purpose is to identify whether or not a user can easily carry out specific tasks within a given system. The nature of the cognitive walkthrough defines it as one of the fastest forms of usability testing, since the user is required to only carry out small tasks. The study can be implemented prior to development and during the design phase of a system. The goal of the evaluation was to identify any potential issues regarding the concept, ascertain any unsupported features, and uncover possible

usability errors by noting the comments of experts before planning a large-scale user-based evaluation.

5.1 Process

During the evaluation, a conductor helped the UX experts to work through a series of tasks and asked them a set of questions. The process included a person who impersonated the participant of the simulated user study. The questions that the experts were inquired aimed to indicate how simple or challenging it was for them to interact with the system and observe the participant's actions. The UX experts were given fourteen (14) tasks, one by one, and after each task, the conductor asked various questions that revolved around the four key questions based on the work of [8]: (i) Will the user try and achieve the right outcome; (ii) will the user notice that the correct action is available to them; (iii) will the user associate the correct action with the outcome they expect to achieve, and (iv) if the correct action is performed, will the user see that progress is being made towards their intended outcome. The conductor was keeping notes for each task containing any comments or suggestions from the experts and recorded the answers for the questions.

5.2 Results

The findings of the evaluation revealed not only some positive comments, but also some issues concerning the design and functionality of the system.

In general, the system received positive feedback from the UX experts, both in terms of concept and implementation. Characteristically, regarding the concept of the pre-study phase where designers are able to build an MR scene, experts mentioned that "it is very interesting to

set-up the scene virtually in the way we want", and "it is helpful to have a virtual overview of how it would be in reality". Other comments included that this approach is "less costly" than having to physically develop and test prototypes. Regarding the ARgus Workstation and Settings menus, evaluators noted that they liked that the offered functionality "was always available", and affirmed that the contextual information that can be displayed at runtime is actually what an expert would like to know during an evaluation. The navigation to the experiment tasks received positive feedback as well, as it was "very easy, without unnecessary complexity". Additionally, with respect to the concept of observing the participants interact with the virtual artifacts and mockups, evaluators reported that "this is very useful as we want to know their mistakes, their movements and if something is confusing to them". Then, concerning the ability of evaluators to help participants by interacting with the virtual environment, it was mentioned that it is "a particularly useful feature, especially if the participant is stuck", and undeniably necessary for the "smooth flow of the experiment". Finally, the possibility of administering AR questionnaires to participants received mainly positive feedback with experts stating that this feature is "very helpful as you can observe the provided answers in real time, without overwhelming the user", and that it offers the possibility to "easily debrief the users regarding their answers without requiring any additional processing".

The most notable issues that were identified from the evaluation process, along with the suggested solution are summarized in the following list:

Issue 1: The placement of the virtual menus' (Settings and ARgus Workstation) in the augmented environment was not comfortable for all users and needs further investigation. In particular, some users stated

that the menus were 'too close', and others said that they were 'too far' to interact with them correctly.

Solution: The algorithm of the menu placement will be revised so as the menu instantiates at an appropriate position according to the user's height. Moreover, a pin functionality will be included so as to permit users to lock the menu's position in the virtual environment. Following these modifications, a user-based usability evaluation will be required for assessing the new functionality.

Issue 2: The location of the settings menu sometimes obscured the user's field of view (FOV) when placing artifacts in the virtual environment during the pre-study phase.

Solution: A pin functionality will be included so as to permit users to lock the menu's position in the virtual environment. In the case whereupon users have not pinned the menu when they are placing artifacts in the virtual environment, the menu will be automatically minimized.

Issue 3: When launched at runtime, the settings menu would sometimes overlap with the ARgus Workstation, if the user had manually changed its position.

Solution: The algorithm of positioning the two menus in the augmented environment will be improved in order to recognize when there is an overlap between them and reposition them accordingly.

Issue 4: The snapshot feature of the ARgus Workstation was missing sufficient feedback. In particular, it was unclear which part of the environment was going to be captured, while there was no indication regarding the time when the picture was going to be taken.

Solution: A distinctive frame –inspired by physical cameras' viewfinder– will be visible when the user selects the Snap option, so as to designate the part of the environment that is going to be photographed. Additionally, a countdown counter will appear in the center of that frame, properly informing the user when the snapshot is going to be captured.

Issue 5: In the pre-study phase, when the users inserted new virtual objects and artifacts into the environment, they could not easily detect where they were instantiated.

Solution: All new virtual objects and artifacts will be instantiated in front of the user (inside their FoV), at a convenient position in order to facilitate their manipulation.

Issue 6: During the process of adding new virtual artifacts into the environment, users sometimes struggled to fine tune their position and orientation.

Solution: Additional controls will be integrated to the system permitting users to set the position and orientation of an artifact more precisely.

Issue 7: A 'close' button was expected at the majority of the pop-up dialogs and menus. Specifically, users did not initially understand that the opened menus (e.g. settings at runtime) or dialogs (e.g. all tasks) could be toggled by the same button that they were opened.

Solution: A 'close' button will be incorporated in conjunction with a toggle functionality of any pop-up dialog so the user will have more flexibility and close them effortlessly.

Issue 8: Lack of appropriate feedback when participants completed an interaction (e.g. touching a screen). Particularly, in the displayed mockup it was not clear which interactive element the participant had selected.

Solution: Relevant feedback will be exhibited to the users after each interaction, with the information (i.e., selected UI element) persisting for a few seconds before disappearing. This way, users will have the time to observe the interactions, diminishing the possibility of missing important data.

Issue 9: Even though the system allowed users to relocate the menu, they sometimes struggled to manipulate it, since there was no indication of where they could handle it.

Solution: A distinct handle control will be included in the menu in order to facilitate its relocation in the virtual space.

In addition to the above issues, this process revealed some missing features that the system can incorporate before proceeding with the user-based evaluation:

- **Note-taking:** The ability to take a note when upon the user marked that an error occurred, or a hint was provided.
- **Gesture-based interaction:** Permit users to perform various functions such as closing the menus, take a snapshot, etc. via gestures.
- **Predefined hints:** In addition to unscripted hints, enable users to populate a list of hints before the study, and select the one to be delivered to the user during the study.

Finally, a short list of technological limitations was revealed, stemming from the immaturity of the current state-of-the-art technology, which are foreseen to be addressed in the near future:

- **Limited FoV:** The limited field of view of the HoloLens 2 device resulted in making virtual items appear out of sight, even if in the real world they would remain in the user's peripheral field of view.
- **Imprecise hand tracking:** The imprecise hand tracking of the headset made the interaction with the virtual components of the MR scene slightly cumbersome.

Chapter 6

Conclusion and Future Work

This work presented ARgus Designer, a tool whose purpose is to support the prototyping phase of developing AR/MR applications by assisting their user-based evaluation through AR. In particular, it will augment the evaluators' view during testing useful real-time features / measures (performance, behavior) as well as functionalities that aim to facilitate the evaluation process in an unobtrusive and seamless manner. In addition to empowering evaluators while conducting user studies with AR/MR applications, this system also facilitates the evaluation of applications aiming to be deployed in Intelligent Environments by superimposing virtual interactive artifacts with digital UIs and digitally augmenting the physical environment with interactive 3D objects using the HoloLens MR glasses. Furthermore, it will permit evaluators to edit the virtual scene of the experiment beforehand, enabling real-time configuration of the intelligent space without having to re-build and deploy the application if they wish to make alterations. The limitations of this work include the setting that the evaluation takes place; the user-based evaluation will

73

only be able to take place in a controlled lab setting, even though previous work has stated the importance of in-the-field studies.

From the cognitive walkthrough evaluation that was conducted to evaluate the concept and the functionality of the presented system, we received positive feedback from the UX experts that prove the usefulness of the tool. However, a number of issues were also identified concerning the design of the system, technological limitations and unsupported features, which will be imminently addressed to improve the overall usability of the ARgus Designer tool.

The following list describes the immediate and long-term plans for future work:

A. Address the issues revealed during the cognitive walkthrough evaluation and integrate the new functionality that the majority of participants identified as necessary.

B. Verify the tool's usability by conducting multiple formal user-based evaluation studies. In particular, one session will include 15 experts in designing applications for Intelligent Environments, in order to determine the usability of the "pre-study" facilities; a second session will include 20 HCI experts so as to assess the "during-" and "post-study" features of the ARgus Designer.

C. Conduct a user-based experiment with the participation of multiple users in the role of participants to assess how the overall user experience of the evaluator is affected.

D. In addition to AR, offer a dedicated desktop-based application to accommodate post-study analysis and replay.

E. Given the broadness of the present application domain of the tool, the scope can be easily expanded from prototyping in Intelligent

Environments to a more extensive and generalized spectrum of AR/MR applications with diverse case studies.

F. Incorporate a mechanism (that was initially in the requirements list but was not implemented, Requirement 5 of Section 3.2.1) for converting the participant's words into text and displaying it inside speech bubbles overplayed in the virtual world.

Bibliography

[1] Alexandrovsky, D., Putze, S., Bonfert, M., Höffner, S., Michelmann, P., Wenig, D., Malaka, R. and Smeddinck, J.D. 2020. Examining design choices of questionnaires in VR user studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), 1–21.

[2] Alexandrovsky, D., Putze, S., Schwind, V., Mekler, E.D., Smeddinck, J.D., Kahl, D., Krüger, A. and Malaka, R. 2021. Evaluating user experiences in mixed reality. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), 1–5.

[3] Arabatzis, D. 2020. *"Wizard of AmI: A system for building and enacting interactive prototypes in Intelligent Environments."* M.Sc. Thesis, Computer Science Department, University of Crete.

[4] Ashtari, N., Bunt, A., McGrenere, J., Nebeling, M. and Chilana, P.K. 2020. Creating augmented and virtual reality applications: Current practices, challenges, and opportunities. *Proceedings of the 2020 CHI conference on human factors in computing systems* (2020), 1–13.

[5] Augusto, J.C., Nakashima, H. and Aghajan, H. 2010. Ambient intelligence and smart environments: A state of the art. *Handbook of ambient intelligence and smart environments*. (2010), 3–31.

[6] Bach, C. and Scapin, D.L. 2004. Obstacles and perspectives for evaluating mixed reality systems usability. *Acte du Workshop MIXER, IUI-CADUI* (2004).

[7] Bai, Z. and Blackwell, A.F. 2012. Analytic review of usability evaluation in ISMAR. *Interacting with Computers*. 24, 6 (2012), 450–460.

[8] Blackmon, M.H., Polson, P.G., Kitajima, M. and Lewis, C. 2002. *Cognitive walkthrough for the web. CHI 2002 Conference on Human Factors in Computing Systems*. ACM Press.

[9] Büschel, W., Lehmann, A. and Dachselt, R. 2021. Miria: A mixed reality toolkit for the in-situ visualization and analysis of spatio-temporal interaction data. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), 1–15.

[10] Butcher, P.W., John, N.W. and Ritsos, P.D. 2020. VRIA: A web-based framework for creating immersive analytics experiences. *IEEE Transactions on visualization and computer graphics*. 27, 7 (2020), 3213–3225.

[11] Carrol, J.M. 1999. Five reasons for scenario-based design. *Proceedings of the 32nd annual hawaii international conference on systems sciences. 1999. hicss-32. abstracts and cd-rom of full papers* (1999), 11 pp.

[12] Cordeil, M., Cunningham, A., Bach, B., Hurter, C., Thomas, B.H., Marriott, K. and Dwyer, T. 2019. IATK: An immersive analytics toolkit. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), 200–209.

[13] Derby, J.L. and Chaparro, B.S. 2021. The Challenges of Evaluating the Usability of Augmented Reality (AR). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2021), 994–998.

[14] Dey, A., Billinghurst, M., Lindeman, R.W. and Swan, J.E. 2018. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI.* 5, (2018), 37.

[15] Dünser, A. and Billinghurst, M. 2011. Evaluating augmented reality systems. *Handbook of augmented reality.* Springer. 289–307.

[16] Edward, J., Ii, S. and Gabbard, J.L. 2005. Survey of user-based experimentation in augmented reality. *In 1st International Conference on Virtual Reality, Las Vegas* (2005).

[17] Fischer, G. and Scharff, E. 2000. Meta-design: design for designers. *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (2000), 396–405.

[18] Fjeld, M., Schar, S.G., Signorello, D. and Krueger, H. 2002. Alternative tools for tangible interaction: A usability evaluation. *Proceedings. International Symposium on Mixed and Augmented Reality* (2002), 157–318.

[19] Friston, S.J., Congdon, B.J., Swapp, D., Izzouzi, L., Brandstätter, K., Archer, D., Olkkonen, O., Thiel, F.J. and Steed, A. 2021. Ubiq: A system to build flexible social virtual reality experiences. *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (2021), 1–11.

[20] Frommel, J., Rogers, K., Brich, J., Besserer, D., Bradatsch, L., Ortinau, I., Schabenberger, R., Riemer, V., Schrader, C. and Weber, M. 2015. Integrated questionnaires: Maintaining presence in game environments for self-reported data acquisition. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (2015), 359–368.

[21] Haniff, D.J. and Baber, C. 2003. User evaluation of augmented reality systems. *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.* (2003), 505–511.

[22] Hart, S.G. 1986. NASA task load index (TLX). (1986).

[23] Henderson, S.J. and Feiner, S. 2009. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. *2009 8th IEEE International Symposium on Mixed and Augmented Reality* (2009), 135–144.

[24] Hewett, T.T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G. and Verplank, W. 1992. *ACM SIGCHI curricula for human-computer interaction*. ACM.

[25] Hubenschmid, S., Wieland, J., Fink, D.I., Batch, A., Zagermann, J., Elmqvist, N. and Reiterer, H. 2022. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. *CHI Conference on Human Factors in Computing Systems* (2022), 1–20.

[26] Irlitti, A., Hoang, T. and Vetere, F. 2021. Surrogate-Aloud: A Human Surrogate Method for Remote Usability Evaluation and Ideation in Virtual Reality. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), 1–7.

[27] Janssen, C.P., Kun, A. and van der Meulen, H. 2017. Calling while driving: An initial experiment with HoloLens. *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (2017).

[28] Jerald, J. 2015. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

[29] Kiourexidou, M., Natsis, K., Bamidis, P., Antonopoulos, N., Papathanasiou, E., Sgantzos, M. and Veglis, A. 2015. Augmented reality for the study of human heart anatomy. *International Journal of Electronics Communication and Computer Engineering*. 6, 6 (2015), 658.

[30] Knierim, P., Schwind, V., Feit, A.M., Nieuwenhuizen, F. and Henze, N. 2018. Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), 1–9.

[31] Krüger, J., Niessner, M. and Stückler, J. A Design and Application Space for Visualizing User Sessions of Virtual and Mixed Reality Environments.

[32] Leonidis, A., Arampatzis, D., Louloudakis, N. and Stephanidis, C. 2017. The AmI-Solertis system: creating user experiences in smart environments. *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (2017), 151–158.

[33] Lobo, M.J., Hurter, C. and Irani, P. 2020. Flex-ER: a platform to evaluate interaction techniques for immersive visualizations.

Proceedings of the ACM on Human-Computer Interaction. 4, ISS (2020), 1–20.

[34] MacDonald, C.M. and Atwood, M.E. 2013. Changing perspectives on evaluation in HCI: past, present, and future. *CHI'13 extended abstracts on human factors in computing systems*. 1969–1978.

[35] Marques, B., Silva, S., Teixeira, A., Dias, P. and Santos, B.S. 2022. A vision for contextualized evaluation of remote collaboration supported by AR. *Computers & Graphics*. 102, (2022), 413–425.

[36] Marques, B., Teixeira, A., Silva, S., Alves, J., Dias, P. and Santos, B.S. 2020. A conceptual model for data collection and analysis for ar-based remote collaboration evaluation. *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (2020), 1–2.

[37] Murray, J.T. 2022. RealityFlow: Open-Source Multi-User Immersive Authoring. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2022), 65–68.

[38] Nebeling, M., Speicher, M., Wang, X., Rajaram, S., Hall, B.D., Xie, Z., Raistrick, A.R., Aebersold, M., Happ, E.G. and Wang, J. 2020. MRAT: The mixed reality analytics toolkit. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), 1–12.

[39] Oh, J.-Y. and Hua, H. 2006. User evaluations on form factors of tangible magic lenses. *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality* (2006), 23–32.

[40] Olsson, T., Lagerstam, E., Kärkkäinen, T. and Väänänen-Vainio-Mattila, K. 2013. Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. *Personal and ubiquitous computing*. 17, 2 (2013), 287–304.

[41] Olsson, T. and Salo, M. 2011. Online user survey on current mobile augmented reality applications. *2011 10th IEEE International Symposium on Mixed and Augmented Reality* (2011), 75–84.

[42] Plattner, H., Meinel, C. and Weinberg, U. 2009. *Design-thinking*. Springer.

[43] Porter, S.R., Marner, M.R., Smith, R.T., Zucco, J.E. and Thomas, B.H. 2010. Validating spatial augmented reality for interactive rapid prototyping. *2010 IEEE International Symposium on Mixed and Augmented Reality* (2010), 265–266.

[44] Pranoto, H., Tho, C., Warnars, H.L.H.S., Abdurachman, E., Gaol, F.L. and Soewito, B. 2017. Usability testing method in augmented reality application. *2017 International Conference on Information Management and Technology (ICIMTech)* (2017), 181–186.

[45] Putze, S., Alexandrovsky, D., Putze, F., Höffner, S., Smeddinck, J.D. and Malaka, R. 2020. Breaking the experience: Effects of

questionnaires in vr user studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), 1–15.

[46] Reipschläger, P., Brudy, F., Dachselt, R., Matejka, J., Fitzmaurice, G. and Anderson, F. 2022. AvatAR: An Immersive Analysis Environment for Human Motion Data Combining Interactive 3D Avatars and Trajectories. *CHI Conference on Human Factors in Computing Systems* (2022), 1–15.

[47] Rieman, J., Franzke, M. and Redmiles, D. 1995. Usability evaluation with the cognitive walkthrough. *Conference companion on Human factors in computing systems* (1995), 387–388.

[48] Rokhsaritalemi, S., Sadeghi-Niaraki, A. and Choi, S.-M. 2020. A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences*. 10, 2 (2020), 636.

[49] Santi, G.M., Ceruti, A., Liverani, A. and Osti, F. 2021. Augmented reality in industry 4.0 and future innovation programs. *Technologies*. 9, 2 (2021), 33.

[50] Santos, M.E.C., Taketomi, T., Sandor, C., Polvi, J., Yamamoto, G. and Kato, H. 2014. A usability scale for handheld augmented reality. *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (2014), 167–176.

[51] Schön, D.A. 2017. *The reflective practitioner: How professionals think in action*. Routledge.

[52] Schwerdtfeger, B., Reif, R., Gunthner, W.A., Klinker, G., Hamacher, D., Schega, L., Bockelmann, I., Doil, F. and Tumler, J. 2009. Pick-by-vision: A first stress test. *2009 8th IEEE international symposium on mixed and augmented reality* (2009), 115–124.

[53] Schwind, V., Knierim, P., Haas, N. and Henze, N. 2019. Using presence questionnaires in virtual reality. *Proceedings of the 2019 CHI conference on human factors in computing systems* (2019), 1–12.

[54] Sheikh, S., Heyat, M.B.B., AlShorman, O., Masadeh, M. and Alkahatni, F. 2021. A Review of Usability Evaluation Techniques for Augmented Reality Systems in Education. *2021 Innovation and New Trends in Engineering, Science and Technology Education Conference (IETSEC)* (2021), 1–6.

[55] Sicat, R., Li, J., Choi, J., Cordeil, M., Jeong, W.-K., Bach, B. and Pfister, H. 2018. DXR: A toolkit for building immersive data visualizations. *IEEE transactions on visualization and computer graphics*. 25, 1 (2018), 715–725.

[56] Speicher, M., Hall, B.D. and Nebeling, M. 2019. What is mixed reality? *Proceedings of the 2019 CHI conference on human factors in computing systems* (2019), 1–15.

[57] Speicher, M., Hall, B.D., Yu, A., Zhang, B., Zhang, H., Nebeling, J. and Nebeling, M. 2018. XD-AR: challenges and opportunities in cross-device augmented reality application development. *Proceedings of the ACM on Human-Computer Interaction*. 2, EICS (2018), 1–24.

[58] Stefanidi, E., Arampatzis, D., Leonidis, A., Korozi, M., Antona, M. and Papagiannakis, G. 2020. Magiplay: An augmented reality serious game allowing children to program intelligent environments. *Transactions on Computational Science XXXVII*. Springer. 144–169.

[59] The 5 Stages in the Design Thinking Process: *https://www.interaction-design.org/literature/article/5-stages-in-the-design-thinking-process*. Accessed: 2022-08-18.

[60] Tonnis, M., Lange, C. and Klinker, G. 2007. Visual longitudinal and lateral driving assistance in the head-up display of cars. *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (2007), 91–94.

[61] Tonnis, M., Sandor, C., Klinker, G., Lange, C. and Bubb, H. 2005. Experimental evaluation of an augmented reality visualization for directing a car driver's attention. *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)* (2005), 56–59.

[62] Voit, A., Mayer, S., Schwind, V. and Henze, N. 2019. Online, VR, AR, Lab, and In-Situ: comparison of research methods to evaluate smart artifacts. *Proceedings of the 2019 chi conference on human factors in computing systems* (2019), 1–12.

[63] Wagener, N., Stamer, M., Schöning, J. and Tümler, J. 2020. Investigating Effects and User Preferences of Extra-and Intradiegetic Virtual Reality Questionnaires. *26th ACM Symposium on Virtual Reality Software and Technology* (2020), 1–11.

[64] Wang, X. and Dunston, P.S. 2011. Comparative effectiveness of mixed reality-based virtual environments in collaborative design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 41, 3 (2011), 284–296.

[65] Young, R.M. and Barnard, P. 1986. The use of scenarios in human-computer interaction research: Turbocharging the tortoise of cumulative science. *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface* (1986), 291–296.

[66] Zainuddin, N.M.M., Zaman, H.B. and Ahmad, A. 2011. Heuristic evaluation on augmented reality courseware for the deaf. *2011 International Conference on User Science and Engineering (i-USEr)* (2011), 183–188.

[67] Zhou, T. 2018. Examining user adoption of mobile augmented reality applications. *International Journal of E-Adoption (IJEA).* 10, 2 (2018), 37–49.

Appendix I.

The research activity related to this thesis has so far produced the following publication:

Helen Stefanidi, Asterios Leonidis, Maria Korozi, George Papagiannakis, "The ARgus Designer: Supporting experts while conducting user studies of AR/MR applications", 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)

Appendix II.

Tasks that UX Experts had to complete during the Cognitive Walkthrough evaluation

The UX experts were given the following introduction: "You are an evaluator who wishes to run a user-study for an MR application for the intelligent kitchen and you are using the ARgus Designer to configure this intelligent environment with virtual artifacts. After setup, you will invite a participant (available for this purpose) to complete a set of predefined tasks, available through the tool, while you will also be asked to follow a separate set of tasks that will be given to you, one by one."

The tasks that the evaluator was asked to follow were:

1. Set up an MR scene for the Smart Home project
2. Select the Emotions feature from the Settings menu
3. Put the corrects mockup group in the displays
4. See how many tasks the participant has to complete
5. Start the experiment
6. Read the first task to the participant and observe their interactions
7. Go to the next task in the experiment
8. Offer voice hint to participant and note that they needed help
9. See participant's heart-rate
10. Take a snapshot of the participant because they are stressed
11. Help the participant complete the task by opening the fridge
12. Note that an error was made
13. Proceed to the last task and read it to the participant
14. Observe the participant fill in the AR Questionnaire