Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

# A vision for contextualized evaluation of remote collaboration supported by AR

Bernardo Marques[a,], Samuel Silva[a], António Teixeira[a], Paulo Dias[a], Beatriz Sousa Santos[a]

[a] *IEETA, DETI, University of Aveiro, Aveiro, 3810-193, Portugal*

## ARTICLE INFO

*Article history*:
Received October 9, 2021

*Keywords:* Remote Collaboration, Augmented Reality, Characterization Collaborative Process, Guidelines, Evaluation Toolkit, User Study.

## ABSTRACT

Remote collaboration using Augmented Reality (AR) has potential to support physically distributed team-members that need to achieve a common goal by increasing knowledge retention, improving understanding and awareness of the problem and its context. In this vein, the path to achieve usable, realistic and impactful solutions must entail an explicit understanding regarding how remote collaboration occurs through AR and how it may help contribute to a more effective work effort. Thus, characterization and evaluation of the collaborative process is paramount, but a particularly challenging endeavor, due to the multitude of aspects that define the collaboration effort. In this context, the work presented here contributes with a critical analysis, discussing current evaluation efforts, identifying limitations and opportunities. Then, we outline a conceptual framework to support researchers in conducting evaluations in a more structured manner. To instrument this vision, an evaluation toolkit is proposed to support contextual data collection and analysis in such scenarios and obtain an additional perspective on selected dimensions of collaboration. We illustrate the usefulness and versatility of the toolkit through a case study on remote maintenance, comparing two distinct methods: sharing of video and AR-based annotations. Last, we discuss the results obtained, showing the proposed vision allows to have an additional level of insights to better understand what happened, eliciting a more complete characterization of the work effort.

## 1. Introduction

Collaboration has the potential to achieve more effective solutions for challenging problems [1]. It has evolved from simple co-located situations to more complex remote scenarios, encompassing several team members with different experiences, expertise's and multidisciplinary backgrounds. Remote collaboration can be described as the process of joint and interdependent activities between physically distributed collaborators performed to achieve a common goal [2, 3, 4]. This activity has become essential in many situations, as is the case of industrial, medical, and educational domains, among others [5, 6]. To address such activities, remote solutions have been growing in terms of scale, complexity, and interdisciplinarity, entailing not only the mastery of multiple domains of knowledge, but also a strong level of proficiency in each [5, 6].

Scenarios of remote collaboration imply that collaborators establish a joint effort to align and integrate their activities in a seamless manner. To address this, and overcome the fact team-members do not share a common space/world, there is an increasing interest in using Augmented Reality (AR) in this context [7, 8, 9, 10]. Remote collaboration mediated by AR combines the advantages of virtual environments and the seamless integration with the real-world objects and other collaborators by overlying responsive computer-generated information on top of the real-world environment [2, 11, 12], allowing to establish a common ground, analogous to their understanding

---

*e-mail:* `bernardo.marques@ua.pt` (Author 1)

of the physical space, i.e., serve as a basis for situation mapping, allowing identification of issues, and making assumptions and beliefs visible [13, 14, 15, 16]. These solutions can be used to empower workers that require knowledge from professionals unavailable on-site [17]. Remote experts can provide guidance, highlight specific areas of interest or share real-time spatial information [14, 9, 18, 10] in the form of visual communication cues, e.g., pointers, annotations, hand gestures, among others [17, 19, 3, 20, 21, 9, 22]. These solutions can better support analysis, discussion and resolution of complex problems and situations, given their ability to enhance alertness, awareness, and understanding of the situation [23].

In the past decade, the community has been particularly active in this domain, concentrating efforts on creating the enabling technology to support the design and creation of an AR-based shared understanding. As the field matures and with the growing number of prototypes, the path to achieve usable, realistic and impactful solutions must entail an explicit understanding regarding how remote collaboration occurs through AR and how it may help contribute to a more effective work effort.

Therefore, evaluating such scenarios becomes an essential, but difficult endeavor [23, 24, 25, 26], given the lack of methods and frameworks to guide the characterization of the collaborative process [27, 9, 28, 29]. This is substantiated by Bai et al. reporting that *"it can be hard to isolate the factors that are specifically relevant to collaboration"* [30]. In fact, this is further evident in remote scenarios, since the logistics associated with carrying out evaluations in these multifaceted contexts is even more demanding due to a significant number of variables that may affect the way teams collaborate [27, 9]. Ratcliffe et al. report that *"remote settings introduce additional uncontrolled variables that need to be considered by researchers, such as potential unknown distractions, (...) participants and their motivation, and issues with remote environmental spaces"* [31].

Also, Dey et al. suggest the existence of *"opportunities for increased user studies in collaboration"* and the need for *"a wider range of evaluation methods"* [32]. In this vein, Ens et al. emphasize that *"frameworks for describing groupware and MR systems are not sufficient to characterize how collaboration occurs through this new medium"* [9]. Additionally, Ratcliffe et al. suggest that *"the infrastructure for collecting and storing this (mass) of XR data remotely is currently not fully implemented, and we are not aware of any end-to-end standardised framework"* [31]. As such, conducting thorough evaluations is paramount to retrieve the necessary data for more comprehensive analysis that help provide a better perspective on the different factors of collaboration supported by AR. Hence, integration of proper characterization and evaluation methods, covering different contexts of use and tasks are of utmost importance.

In this paper, we analyse existing evaluation efforts on remote collaboration using AR to provide a high-level overview. Motivated by the challenges reported, we present a conceptual framework for supporting researchers in obtaining an additional perspective on several dimensions of collaboration. Then, we propose the *CAPTURE* toolkit, a first instantiation towards the vision proposed, aiming to provide a strategy that monitors data concerning the level of collaboration, behaviour and performance of each intervening party, individual and as a team, as well as contextual data. To illustrate the advantages of the framework, the toolkit usefulness and versatility are demonstrated through a case study in a remote maintenance scenario, comparing two distinct methods: sharing of video and AR-based annotations. Then, the results obtained are discussed, showing that the proposed vision allows having an additional level of insights to better understand what happened, eliciting a more complete characterization of the collaborative work effort.

The remainder of this paper is organized as follows. Section 2 overviews existing evaluation efforts on remote collaboration mediated by AR. Section 3 proposes our conceptual framework for essential aspects that must be addressed. Section 4 describes the *CAPTURE* toolkit and Section 5 applies it through a user study on a remote maintenance scenario. Section 6 presents and discusses the main results. Finally, concluding remarks and future research opportunities are drawn in Section 7.

## 2. Background and Challenges for Evaluation of AR-based Remote Collaboration

This section reports existing evaluation efforts addressing collaborative AR user studies. The goal is to understand how evaluation has been conducted in such scenarios, provide a high-level overview, and identify existing challenges and gaps.

According to Merino et al. *"as MR/AR technologies become more mature, questions that involve human aspects will gain focus in MR/AR research. Consequently, we expect that future MR/AR papers will elaborate on human-centered evaluations that involve not only the analysis of user performance and user experience, but also the analysis of other scenarios, like understanding the role of MR/AR in working places and in communication and collaboration"* [26]. However, there is no standard methodology for characterization and evaluation, specifically tailored to assess how remote collaboration occurs through AR.

The literature shows that studies that evaluate their solutions rely on single-user methods, mainly focused on the comparison of technological aspects or interaction mechanisms, which are not the most adequate for multifaceted solutions that aim to support distributed team collaboration [25, 2, 32, 9, 27, 29]. Also, most studies focus exclusively on the performance of one collaborator, i.e., on-site, or remote. This means evaluation usually does not consider interaction, and communication among team-members, and is not conducted in distributed scenarios, as should be the case to establish experimental conditions closer to real scenarios, Likewise, focus is given to the technological aspects of the solution being used, as well as to quantifying the effectiveness in completing the tasks, which mostly lack difficulty, diversity and ecological validity [30, 7, 32, 2, 3, 29].

Moreover, the majority of studies are formal, conducted in laboratories, collecting objective and subjective data at the end of the tasks through standard practices with fixed answers like scale-based questionnaires (e.g., System Usability Scale (SUS), NASA Task Load Index (TLX), among others) or direct observation [7, 19, 33, 34, 32, 9, 27]. Adding to these data, only a reduced set of studies include measurements collected during the collaborative process (e.g., task duration and error/accuracy), as well as explicit communication (e.g., spoken messages or gestural cues), ease of collaboration and others [30, 26]. While this is the case, the collection of more contextual and behavioral data is often not considered or hindered due to the complexity it entails regarding acquisition, processing and analysis, and more important, the lack of guidelines to inform researchers on what dimensions of collaboration should be collected and how.

Therefore, current frameworks are not tailored to characterize how collaboration mediated by AR occurs [25, 9, 28, 31], falling short to retrieve the necessary amount of data for more comprehensive analysis. As as consequence, without the appropriate methods, the research community does not accumulate enough experience to improve the work effort [35, 30, 25, 32, 3, 9, 26, 28, 29]. Thus, as the field of remote collaboration using AR matures, evaluation needs to move beyond a simple assessment of how the technology works, as it becomes essential to understand different aspects of collaboration itself, including how teams work together, how communication happens, how AR is used to create a common ground, among others. This should provide a richer output of the evaluation stage, balancing the design against requirements and leading to a more informed refinement of the context of use and system features, e.g., in line with the life-cycle for Human-Centered Design (HCD) described in the principles and activities associated with the [ISO 9241-210][1].

Given the challenges and constraints involved in evaluating the way collaboration occurs through AR, we argue it is paramount to address a set of important topics, namely: 1- **conduct more collaborative-centric evaluations**, i.e., move beyond usability testing, which fails to obtain a more comprehensive understanding of the work effort. Equally important, 2- **develop evaluation strategies including contextual data collection and visualization**, i.e., collect a richer data set to better understand how AR contributes to the collaborative process, in order to shape more effective collaboration.

## 3. A vision for contextualized Collaborative AR evaluation

The area being addressed in this work is part of a complex phenomenon. To allow answering existing problems, it is necessary to systematize knowledge and perspectives, so that it can be applied transversely. For this, it is necessary the creation of evaluation frameworks, i.e., capitalize on the hierarchies and dimensions of collaboration from ontologies and taxonomies, as well as the development of tools that allow contextualizing the use of collaborative solutions.

Taking into account the challenges and needs identified in the previous section, Figure 1 structurally presents an evaluation framework of the collaborative process when using a given tool, with a proposal of several levels of information that must be considered for contextualization, derived through a HCD methodology. In this effort, we argue that the evaluation process must be addressed by the research community, namely the definition of the evaluation purpose, as well as the team characteristics and the details of the collaborative tasks. Also, carefully establish the experimental setup and design. Equally important, explore contextualized data gathering and analysis, which requires the creation of novel tools. This last, being the aspect this work further contributes. Next, we elaborate on these with more detail.

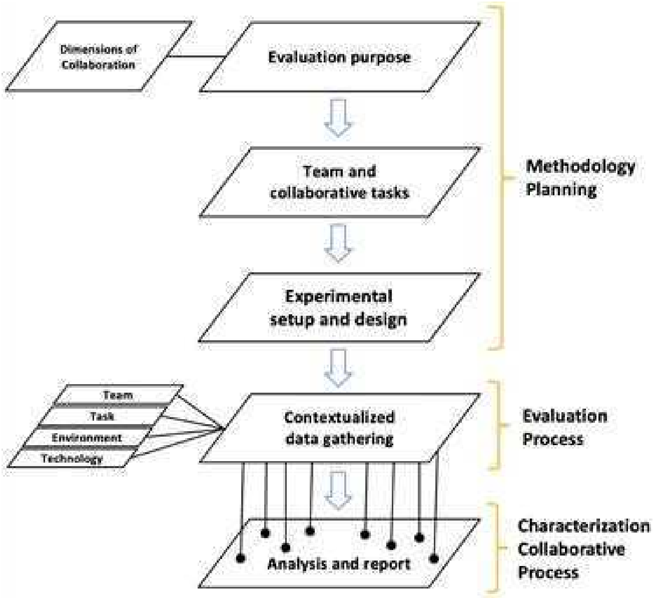


**Fig. 1. Conceptual framework for helping researchers evaluate of AR-remote collaboration in a more structured manner.**

### *3.1. Evaluation purpose*

To begin, the scope must be defined, taking into account existing dimensions of collaboration to clarify what will be evaluated, so that relevant research questions are formulated in the design phase and answered in the evaluation analysis [24].

### *3.2. Team and collaborative tasks*

Also important, determine the team-members' characteristics, i.e., role structure, coupling level, life-span, technology literacy and multidisciplinarity. In this context, participants with different ages, perspectives, motivations, and multidisciplinary background should be considered, which might lead to more relevant insights. Moreover, understanding of VR/AR, as well as remote tools is a benefit for the adaptation, thus removing the 'wow factor' that makes participants feel excitement or admiration towards such technologies. Besides, participants should only perform one role, i.e., on-site or remote, so that they are only exposed to a set of tasks, concerns and responsibilities.

Furthermore, the collaborative tasks goals must be clearly established including which team-members will be accountable for achieving each completion stage. It is also important to consider if the tasks are performed indoor, outdoor, or mixed between the two; A balance must be kept between task complexity and duration. Tasks must be complex and long enough to encourage interaction through AR. However, longer tasks may cause fatigue or boredom, affecting the evaluation outcomes. Equally relevant, tasks can introduce deliberated drawbacks, i.e., incorrect, contradictory, vague or missing information, to force more complex situations and elicit collaboration.

---

[1] iso.org/standard/77520.html

### 3.3. Experimental setup and design

Establish the experimental setup and design are equally key. When considering prototypes, evaluation under laboratory settings should be used. Afterwards, when considering more mature solutions, evaluation should be made in the field, with real stakeholders and domain experts, moving beyond typical laboratory settings to increase the ecological validity of the evaluations. Regarding the environment, two separated rooms in the same/different building(s) should be used. Otherwise, participants must be separated by some kind of physical barrier when in the same room. Furthermore, an adaptation period must be provided so that participants can explore the technology possibilities before the tasks, individually and as a team. Besides, a proper amount of time must be defined for other aspects, e.g., presentation of the study, pre- and post-task questionnaires, team interview, and others.

### 3.4. Contextualized data gathering

As well observed by Merino et al. [26], future works on Mixed and Augmented Reality (MR/AR) will elaborate on human-centered evaluations involving not only the analysis of user experience and performance, but also understanding the role of such technologies in working places, in communication and in collaboration. In this scope, contextual information helps inform the conditions in which the collaboration took place. It can also be used for understanding interaction and communication changes, namely if the surroundings affected the way teams collaborate, in such a way that they needed to adapt it. Also, it helps portrait the conditions in which team-members performed a given action, received information or requested assistance, which can be used to assess uncommon situations or identify patterns that can lead to new understanding of a given artifact, as well as identify new research opportunities.

Without comprehending contextual information, it becomes difficult to assess important variables related to the collaborative process, which means the findings reported may be misleading or of limited value. Hence, these aspects have an important impact on how the studies must be prepared and how they were conducted, influencing situation understanding, team-members communication, performance, and usage of AR.

Literature shows that a better evaluation process can be supported by improved data collection and data visualization tools [35, 36]. In particular, the following factors are crucial and must be taken in account to better understand the real impact of each aspect in the collaborative effort: team, tasks, context and AR-based tool [28]. Through these, a wide range of information is provided when performing judgment over the results and establishing conclusions.

Therefore, data collection while team-members collaborate, considering different forms of measurement according to the evaluation goals is paramount and should include:

- **pre-task measures** like demographic questionnaires (e.g., age, gender, occupation, years of experience, etc.), information on participants background: if they knew each other, previous experience with VR/AR technologies and remote tools, among other aspects;
- **runtime measures** may comprise:
  - *performance metrics* including overall duration of specific events, number and type of errors; number and type of interactions; frequency of using each feature of the tool; screenshots of the enhanced content;
  - *behaviour metrics* including conversational analysis (e.g., frequency of conversational turns, number of questions or interruptions, and dialog length, duration of overlapping speech); physical movement around the environment; number of hand gestures; physiological variables and emotions; eye gaze;
  - *collaboration metrics* including the level of effectiveness; perception; interest; engagement; awareness; togetherness; mental stress;
  - researchers may collect audio (or video) and register *interesting events* including the type (e.g., guide, request, express, propose) and frequency of communication (e.g., never, sometimes, often, continuously), if the goals were accomplished, difficulties detected, if the participants requested assistance and how many types, among other relevant aspects.
- **post-task measures** can encompass:
  - register usability towards the tools(s) used;
  - record *collaboration metrics* including the level of effectiveness; perception; interest; engagement; awareness; togetherness; mental stress, etc;
  - collect participants reactions, opinions and preferences through semi-structured interviews.

### 3.5. Analysis and report

The use of more contextualized approaches will provide ground to improve how research is analyzed and reported. Hence, increasing the awareness of researchers about the different dimensions of collaboration and the need to improve how the nuances associated to the collaborative effort are described. In turn, a more systematic characterization can lead to a community setting that enables easier communication, understanding, reflection, comparison and refining, building on existing research while fostering harmonization of perspectives for the field. In this context, some noticeable recommendations are:

- researchers can profit from the outcomes generated to improve the level of detail provided in their reports;
- the collaborative context needs to be widely described, allowing the creation of a better understanding of the surrounding conditions, including relations between individuals, their interconnection as a team, how AR was used, the characteristics of the environment, and others;
- the outcomes can help identify limitations and promising functionalities regarding AR, providing opportunities for future work in a technical level;
- the insights obtained may also lead to improvements in individual behaviour and team collaboration in specific procedures and tasks over longer periods of time.

## 4. Toolkit for Distributed Evaluations using AR

Given the challenges in evaluating the way remote collaboration occurs, the absence of frameworks and tools, this section describes *CAPTURE - Contextual dAta Platform for remoTe aUgmented Reality Evaluation*, a first instantiation towards addressing the vision previously described, in particular the need to include more contextual data in the evaluation of the collaborative process (Figure 2), following the conceptual model [28].

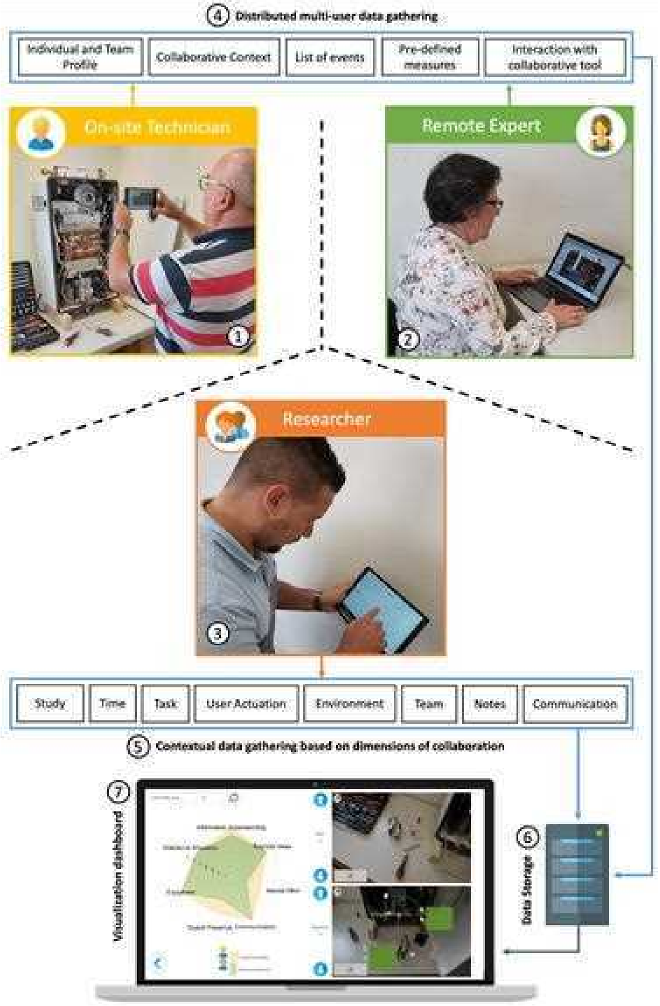


**Fig. 2. Scenario of remote collaboration using an AR-based tool instrumented with the *CAPTURE* toolkit: 1- On-site technician requiring assistance; 2- Expert using AR to provide remote guidance; 3 - Researcher(s) following the evaluation process; 4- Distributed multi-user data gathering; 5- Contextual data collection based on existing dimensions of collaboration [37]; 6- Evaluation data storage; 7- Visualization dashboard for analysis of the collaborative process.**

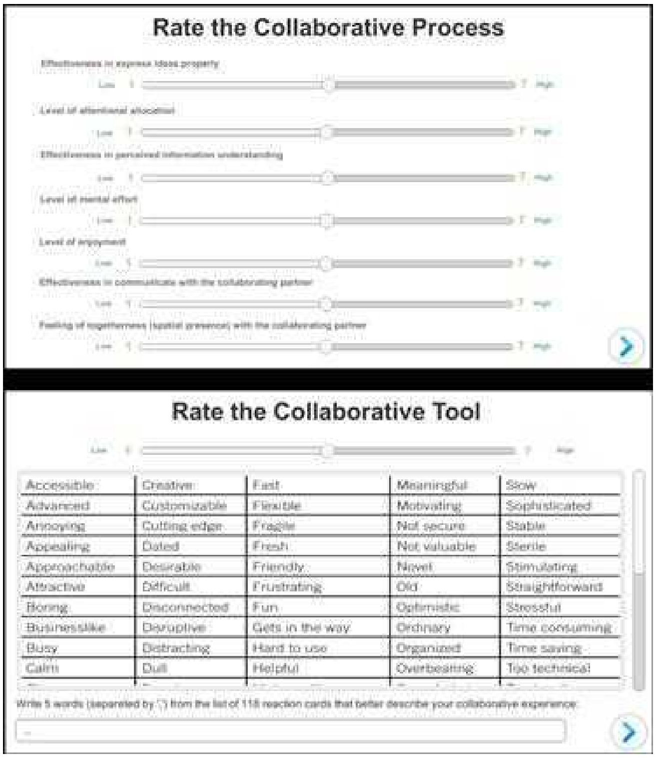
To inform the conceptualisation/development, we conducted brainstorm sessions with domain experts (academics, including faculty members and researchers) sharing several years of expertise in HCI, VR/AR, Visualization and remote collaboration, who co-authored multiple publications, and projects on these subjects. Hence, the toolkit must support:

- data gathering at distributed locations in synchronous and asynchronous manner;
- explicit input on different dimensions of collaboration, following a taxonomy for Collaborative AR [38, 39] and an evaluation ontology for remote scenarios [37];
- data collection regarding team interaction, custom logging and registration of interesting events according to the selected scenarios of remote collaboration;
- easy instrumentation into remote tools by providing ready to use scripts and prefabs for non-experts in programming, i.e., each process can be configured via visual editors;
- modularity to ensure adaptation to different goals;
- data storage and aggregation via a centralized server;
- post-task analysis through a visualization dashboard.

To elaborate, for team-members, the *CAPTURE* toolkit provides native off-the-shelf modules to support explicit input and data gathering regarding (Figure 2 - 4):

- *individual and team profile*: demographic data, knowledge of other collaborators, participants background, emotional state [40], experience with AR and remote tools;
- *collaborative context*: details on the task and the environment, like the number of completion stages, resources available or the amount of persons, movement and noise in the surrounding space;
- *list of events*: task duration, augmented content shared and received, and other relevant occurrences;
- *pre-defined measures*: characteristics associated to the collaborative process, including, but not limited to, easy to communicate or express ideas and the level of spatial presence, enjoyment, mental effort, information understanding, attention allocation or others (Figure 3 - top). Also, the Microsoft reaction card methodology [41] to have a grasp on team-members reaction towards the tool used for shared understanding (Figure 3 - bottom);
- *interaction with the collaborative tool*: duration of the collaborative process and specific events, e.g., when creation of content is started or completed, number and type of interactions, frequency of using each feature, as well as captures of the augmented instructions being shared.

Regarding pre-defined measures, the aspects of collaboration proposed are the result of carefully survey existing literature to create a list of important topics facing the lack of methodologies and frameworks. This list was presented to the experts, who had an important role in selecting, analysing and filtering said topics of collaboration by voting about the ones they considered being more relevant. To elaborate, we took inspiration from the questionnaires used by [42, 43, 44, 45, 3, 46], as well as the works by [47, 48, 19, 11, 49, 50, 21, 22]. Nevertheless, other aspects of collaboration can be considered according to the evaluation scope due to the inherent flexibility provided by the *CAPTURE* toolkit implementation, as described below.



Rate the Collaborative Process

Effectiveness in express ideas properly: Low 1 — 7 High

Level of attentional allocation: Low 1 — 7 High

Effectiveness in perceived information understanding: Low 1 — 7 High

Level of mental effort: Low 1 — 7 High

Level of enjoyment: Low 1 — 7 High

Effectiveness in communicate with the collaborating partner: Low 1 — 7 High

Feeling of togetherness (spatial presence) with the collaborating partner: Low 1 — 7 High

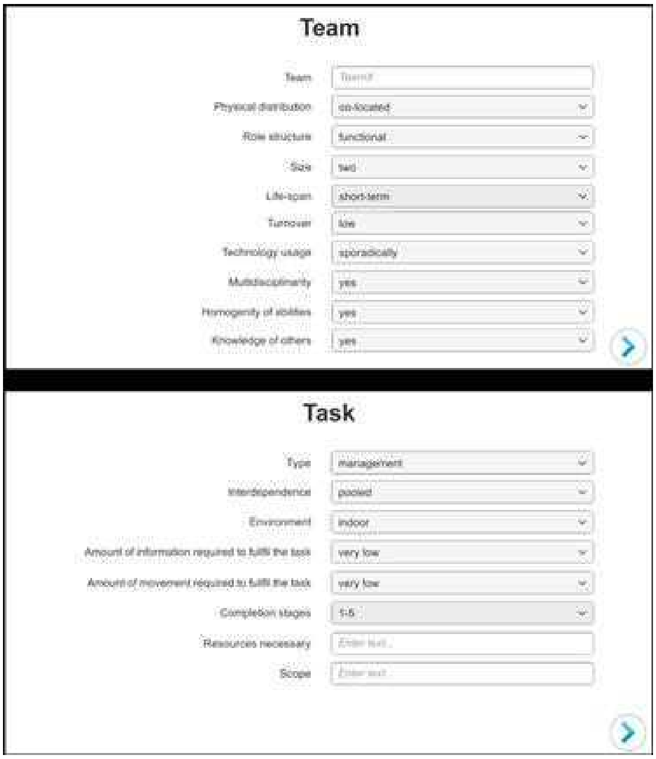
Rate the Collaborative Tool

Low 1 — 7 High

| | | | | |
|---|---|---|---|---|
| Accessible | Creative | Fast | Meaningful | Slow |
| Advanced | Customizable | Flexible | Motivating | Sophisticated |
| Annoying | Cutting edge | Fragile | Not secure | Stable |
| Appealing | Dated | Fresh | Not valuable | Sterile |
| Approachable | Desirable | Friendly | Novel | Stimulating |
| Attractive | Difficult | Frustrating | Old | Straightforward |
| Boring | Disconnected | Fun | Optimistic | Stressful |
| Businesslike | Disruptive | Gets in the way | Ordinary | Time consuming |
| Busy | Distracting | Hard to use | Organized | Time saving |
| Calm | Dull | Helpful | Overbearing | Too technical |

Write 5 words (separated by ';') from the list of 118 reaction cards that better describe your collaborative experience:

**Fig. 3.** ***CAPTURE*** **toolkit - example of pre-defined scenes associated with post-task measurements. Top - questionnaire regarding the collaboration process; Bottom - questionnaire regarding the collaborative tool.**



Team

Team: Team#
Physical distribution: co-located
Role structure: functional
Size: two
Life-span: short-term
Turnover: low
Technology usage: sporadically
Multidisciplinarity: yes
Homogenity of abilities: yes
Knowledge of others: yes

Task

Type: management
Interdependence: pooled
Environment: indoor
Amount of information required to fulfil the task: very low
Amount of movement required to fulfil the task: very low
Completion stages: 1-5
Resources necessary: Enter text...
Scope: Enter text...

**Fig. 4.** ***CAPTURE*** **toolkit - example of pre-defined scenes associated with selected dimensions of collaboration. Top - characteristics of the Team; Bottom - characteristics of the Task.**

As for the researcher(s), the toolkit provides native off-the-shelf modules to support explicit input regarding (Figure 2 - 5):

- *Study*: area of application, research context and study type;
- *Time*: synchronicity, duration and predictability;
- *Team*: distribution, role structure, size, life-span, turnover, multidisciplinarity, technology usage, homogeneity of abilities, and knowledge of others (Figure 4 - top);
- *Task*: scope and type of task, interdependence, amount of information and movement required to fulfil the task, number of completion stages, resources necessary to achieve the goal (Figure 4 - bottom);
- *User Actuation*: capacity to passive-view, interact/explore, share/create, as well as level of symmetry;
- *Communication*: structure, mode, intent, frequency and duration;
- *Environment*: amount of noise, level of brightness, number of persons in the environment, weather conditions and resources available;
- *Notes*: interesting events, notes, comments or difficulties, as well as if the goals were achieved and the amount of physical movement conducted by the team-members.

At the system level, *CAPTURE* consists of a Unity Package that can easily be added to existing collaborative solutions in Unity. All data gathered from the different team-members and researcher(s) during collaboration sessions is stored in a central server for post-evaluation analysis through a visualization dashboard (Figure 5), which allows reviewing the work effort of a particular team or set of teams, as well as compare different tools, if that is the evaluation scope [51].

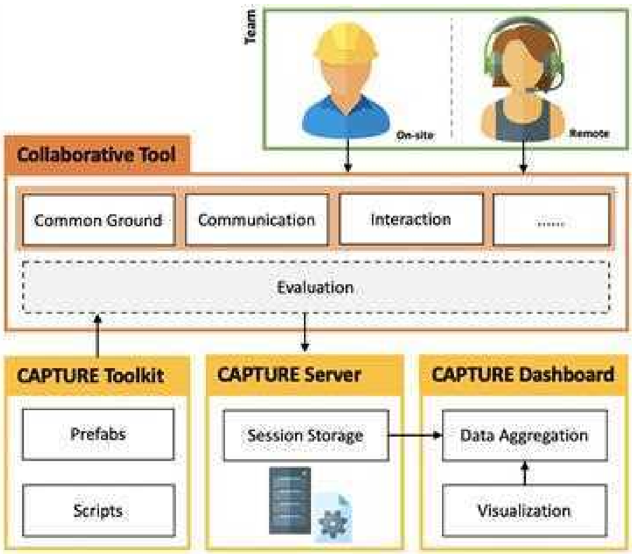
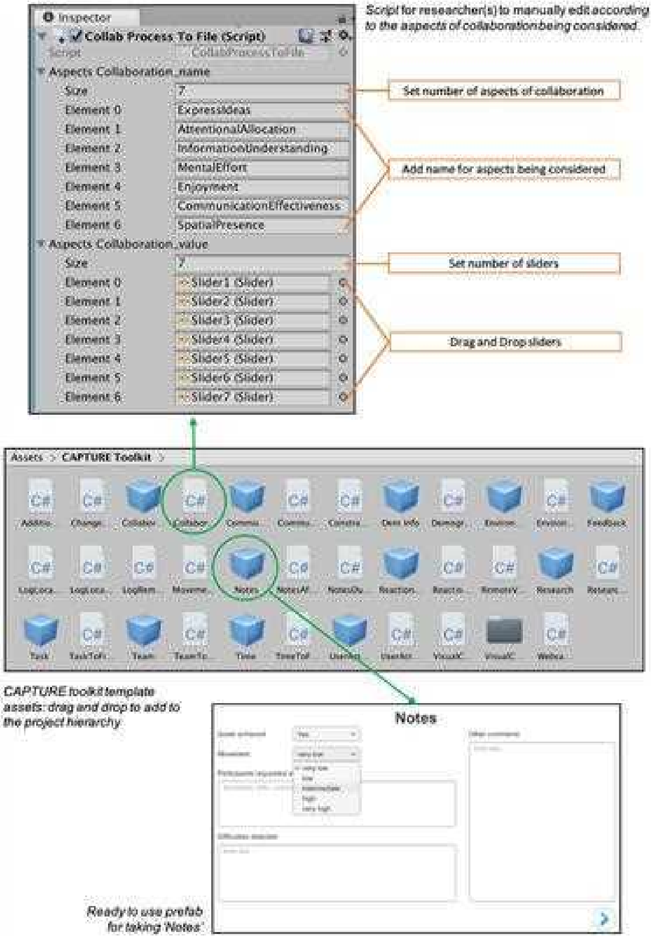


**Fig. 5.** ***CAPTURE*** **architecture. The toolkit can be integrated into a collaborative tool via visual editor. All data collected during collaboration is stored in a central server, which can be analyzed during post-task analysis through the visualization dashboard.**

The modules of the proposed toolkit can be integrated into existing remote tools via visual editors, i.e., with minimal need for programming skills (Figure 6). It is possible to drag and drop ready to use prefabs and editable scripts into Unity 3D projects, which can be modified according to the evaluation scope in the inspector module. Figure 6 illustrates the example of the collaborative process script, which researchers can manually edit (set the number of elements, add relevant aspects of collaboration to be assessed, etc.) according to the evaluation scope. This dynamic approach allows researchers to re-use scripts over different evaluation sessions according to the collaborative effort being considered. For development, Unity 3D was used based on C# scripts. Communication between each instance is performed over Wi-Fi through calls to a PHP server.



Inspector
Collab Process To File (Script)
Script: CollabProcessToFile
Aspects Collaboration_name
Size: 7
Element 0: ExpressIdeas
Element 1: AttentionalAllocation
Element 2: InformationUnderstanding
Element 3: MentalEffort
Element 4: Enjoyment
Element 5: CommunicationEffectiveness
Element 6: SpatialPresence
Aspects Collaboration_value
Size: 7
Element 0: Slider1 (Slider)
Element 1: Slider2 (Slider)
Element 2: Slider3 (Slider)
Element 3: Slider4 (Slider)
Element 4: Slider5 (Slider)
Element 5: Slider6 (Slider)
Element 6: Slider7 (Slider)

Script for researcher(s) to manually edit according to the aspects of collaborationbeing considered.

Set number of aspects of collaboration

Add name for aspects being considered

Set number of sliders

Drag and Drop sliders

CAPTURE toolkit template assets: drag and drop to add to the project hierarchy

Ready to use prefab for taking 'Notes'

**Fig. 6. Overview of the *CAPTURE* toolkit assets: ready to use scene prefabs and editable scripts, which researchers may modify according to the aspects of collaboration being considered for the evaluation.**

In short, the field needs to have more contextualized evaluation strategies, allowing to learn more regarding how technology address the collaborative process. All of this can support an effort towards systematized data, which may support the proposal of guidelines in the future, resulting from the experience and knowledge accumulated through the analysis from multiple research teams and different technology approaches with contextualized information. This effort will allow to use these recommendations to jump-start the quality of current and novel solutions right from the very beginning of its conceptualization, which have already been proven useful in remote scenarios.

## 5. User Study on a Remote Maintenance Scenario

A user study was conducted to compare the collaborative process of distributed teams using two distinct tools when instrumented with *CAPTURE*: Video Chat and AR-based Annotations. These were proposed following a user-centered approach with partners from the industry sector to probe how AR could provide solutions to support their collaborative needs.

### 5.1. Experimental Setup

To create a common ground between distributed team-members, two distinct methods were provided: a video chat tool and an AR-based annotation tool. Next, a brief description of the main features of each tool is provided. To clarify, the hardware used was the same for both methods, only the characteristics of the tool changed. Also, both tools were developed using the Unity 3D game engine, based on C# scripts. Communication was provided over Wi-Fi through WebRTC calls to a dedicated server. To place the augmented content in the real-world environment, we used the Vuforia library.

### 5.1.1. Video Chat Tool

The first method uses video chat features to provide support (Figure 7). On-site participants can point a handheld device to the situation context, which is shared though live video stream with the remote expert. In this context, the face of the expert is visible at all times, while the on-site participant may change between showing the task context or his face using the back and front cameras of the device. Besides, team-members can share text messages using the chat to ensure important messages are kept visible. Using these features, team-members may communicate and discuss the content being captured to express the main difficulties, identify areas of interest or the remote expert to inform where to act and what to do.

**Fig. 7. Video Chat tool for remote collaboration.**

#### *5.1.2. AR-based Annotation Tool*

The second method uses AR-based annotations as additional layers of information (Figure 8). On-site participants can point a handheld device to capture the situation context. Using audio communication and annotation features like drawing, placing pre-defined shapes or notes, as well as sorting annotations, the participant can edit the capture to illustrate difficulties, identify specific areas of interest or indicate questions. Then, the capture is sent to the remote expert to suggest instructions accordingly i.e., inform where to act, and what to do, using similar annotation features. Afterwards, the on-site participant receives the annotations. The handheld device can be placed on top of a surface to follow the instructions in a hands-free setting. At any time, it can be picked up to perform an augmentation of the annotations, by re-aligning with the real world.

**Fig. 8. AR-based Annotation tool for remote collaboration.**

### *5.2. Experimental Design*

A within-group experimental design was used. The null hypothesis (H0) considered was that the two experimental conditions are equally usable and acceptable to conduct the selected maintenance tasks. The independent variable was the information display method provided during the collaborative process, with two levels corresponding to the experimental conditions: *C1- Video Chat* and *C2- AR-based Annotations*. For both experimental conditions, the tools used provided a similar level of user actuation for both team-members, having identical features to view (C1 and C2), create, share and interact with augmented content (C2). Performance measures and participants' opinion were the dependent variables. Participants' demographic data, as well as previous experience with AR and collaborative tools were registered as secondary variables.

### *5.3. Tasks*

We focused on a case study where an on-site participant using a handheld device had to perform a maintenance procedure while being assisted from a remote expert using a computer. The tasks require accomplishing the following steps (Figure 9): 1- replace interconnected components, 2- plug and unplug some energy modules, 3- remove a specific sensor, as well as 4- integrate new components into the equipment. For each condition, different tasks were used to minimize bias, i.e., learning effect. Nevertheless, we defined these tasks based on feedback from our industry partners regarding their usual work activities and needs, while ensuring a similar level of difficulty and resources.

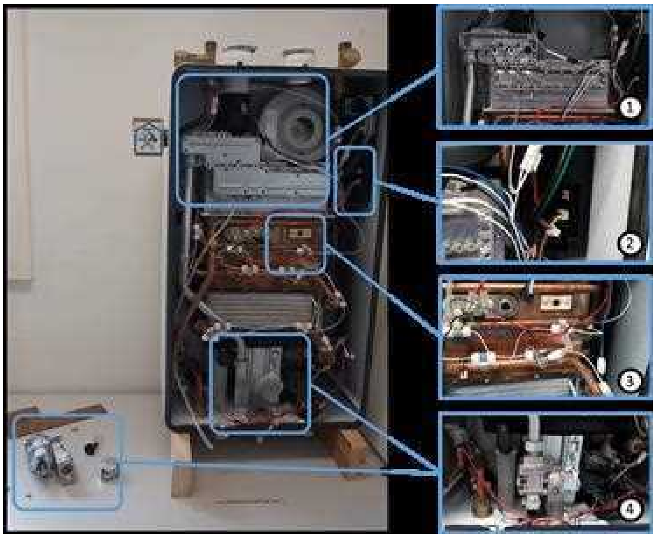


**Fig. 9. Illustration of some of the completion stages associated with the maintenance tasks used in the study: 1- replace interconnected components; 2- plug and unplug some energy modules; 3- remove a specific sensor; 4- integrate new components into the equipment.**

Each task was a defined-problem with 4 completion stages, forcing team-members to communicate in a continuous way while acting alternately (reciprocal interdependence) in an indoor environment with controlled illumination conditions and reduced noise. Besides the participants and researchers, no other individuals were present. The on-site participant needed to use different hand tools to perform the procedures, although low physical movement was required.

### *5.4. Measurements*

All data was collected through the *CAPTURE* toolkit for all conditions, including standard measures found in literature like task performance based on the overall total time, i.e., time needed to complete the tasks, answer to questionnaires and participation in a brief interview, as well as task time, i.e., time required for successfully fulfill the task in a collaborative manner. Besides, novel measures, taking advantage of the toolkit off-the-shelf modules, i.e., information on selected dimensions of collaboration (e.g., time, team; task; user actuation, communication, environment); the overview of the collaborative process (e.g., easy to communicate or express ideas, level of spatial presence, enjoyment, mental effort, information understanding and attention allocation) at the end of the tasks; participants emotional state, before and after the task fulfilment; participants preferences and opinion, also at the end. Hence, the toolkit was integrated into an existing video chat tool, as well as an AR-based tool [Omitted for review] using stabilized annotations, following prior work with partners from the Industry sector.

### 5.5. Procedure

Participants were instructed on the experimental setup, the tasks and gave their informed consent. Then, they were introduced to both tools and a time for adaptation was provided. Participants would act as on-site technicians with condition C1 and then C2, always in this order, while a researcher was the remote counterpart to ensure the instructions were correctly transmitted. We used this approach to facilitate collaboration, as having participants also act as the remote counterpart would add an additional level of complexity, which we believe was not necessary. Since this role was ensured by one of the researchers, we recognize that it is not the same as having a participant, but still allows to have a granular view of the work effort, since not all collaborative processes are created equal. Hence, the researcher also followed the same procedure during the evaluation. We argue that the data collected from this role convey a variability in the way collaboration occurred and in what works or not, depending on the team-members, which demonstrates the ability of the measures used to have some granularity in the evaluation of how the collaborative process took place.

Participants started with a demographic questionnaire. In the next stage, they completed the maintenance tasks while observed by a researcher who assisted them if necessary, and registered any relevant event. Immediately after completing the tasks using the conditions, participants answered a post-study questionnaire regarding the collaborative process, as well as their preferences towards the tool used. Then, a small interview was conducted to understand participants' opinion regarding their collaboration with each condition. The data collection was conducted under the guidelines of the Declaration of Helsinki. Also, all measures were followed to ensure a COVID-19 safe environment during each session of the user study.

### 5.6. Participants

We recruited 26 participants (9 female - 34.7%), whose ages ranged from 20 to 63 years old (M = 33.1, SD = 11.7). Participants had various professions, e.g., Master and PhD students, Researchers and Faculty members from different fields, as well as Software Engineers, Front-End Developers and an Assembly Line Operator. With respect to *individual and team profile*, 14 participants had prior experience with AR and 24 with collaborative tools. With the exception of 1 team, all collaborators had knowledge of each other prior to the study.

## 6. Results and Discussion

This section presents and discusses the main results obtained from the analysis of the data collected through *CAPTURE*.

### 6.1. Overall total time and task time

As for the total duration, sessions lasted 32 minutes on average (SD = 3.10) using condition C1 and 28 minutes on average (SD = 3.03) using condition C2 (Figure 10). Regarding task duration, it lasted 16 minutes on average (SD = 2.68) using condition C1 and 12 minutes on average (SD = 2.66) using condition C2. Therefore, participants were quicker on average to perform the tasks when using condition C2, despite having a higher data variability when compared to condition C1.

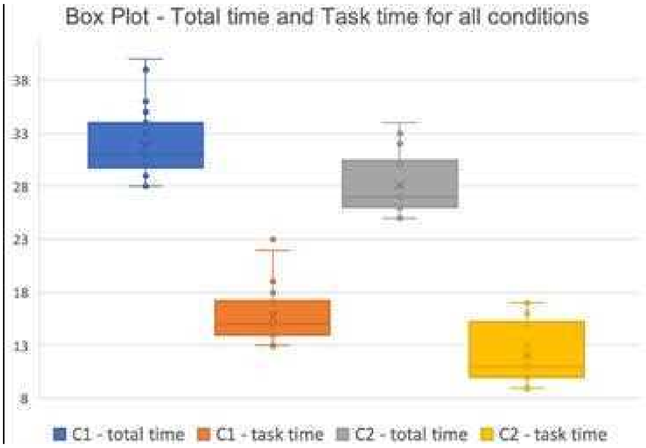


**Fig. 10. Total time and task time with the two conditions (in minutes).** ***C1: video chat tool; C2: AR-based annotation tool.***

### 6.2. Overview of the collaborative process

Regarding condition C1, participants rated the collaborative process (Likert-type scale: 1- Low; 7- High) as following (Figure 11 - top): express ideas (median= 4.5), attentional allocation (median= 4), information understanding (median= 5), mental effort (median= 5), enjoyment (median= 4), communication (median= 5), spatial presence (median= 5.5). As for condition C2, participants rated the collaborative process as following (Figure 11 - bottom): express ideas (median= 6), attentional allocation (median= 7), information understanding (median= 7), mental effort (median= 2), enjoyment (median= 6), communication (median= 6), spatial presence (median= 5).

Hence, it is possible to understand that for the majority of aspects of collaboration, i.e., easy to share ideas properly, level of attention allocation, level of information understanding, level of enjoyment and easy to communicate, condition C2 was rated higher by the participants. Regarding the level of mental effort, participants rated higher condition C1, possibly due to the diminished level of attentional allocation this condition had, which lead to some communication arguing in order to understand where to perform some activities. Therefore, these results suggest that the AR-based annotation tool was better in such aspects of collaboration when compared to the video alternative.

In contrast, for condition C1 the level of spatial presence was higher. This might be associated to the fact that this condition supported live video sharing between team-members, which may have an impact on participants feeling of togetherness with their collaborative counterparts, since it was possible to see the remote expert at all times during the task duration. On the other side, condition C2 provided stabilized AR-based annotations on top of captures/images of the task context. This condition did not allow to see the remote expert during the task procedures, which may have affected participants reaction towards the level of spatial presence, although not with any major difference.

In this context, a smaller data variability can also be observed for easy to share ideas properly, level of information understanding, level of mental effort, easy to communicate and level of spatial presence, when analysing the box plots of condition C1 and C2, as illustrated by Figure 11.

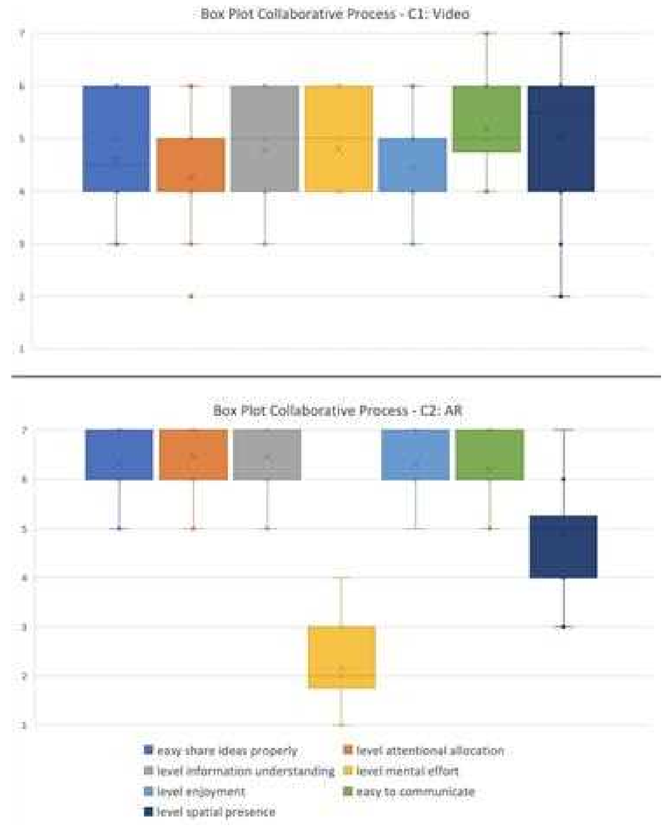


**Fig. 11. Overview of the collaborative process outcomes for all teams during a scenario of remote maintenance, including all the selected measures collected: easy to share ideas properly, as well as communicate, level of attentional allocation, information understanding, mental effort, enjoyment, spatial presence.** ***Top - C1: video chat tool; Bottom - C2: AR-based annotation tool*****. Data displayed using a Likert-type scale: 1- Low; 7- High.**

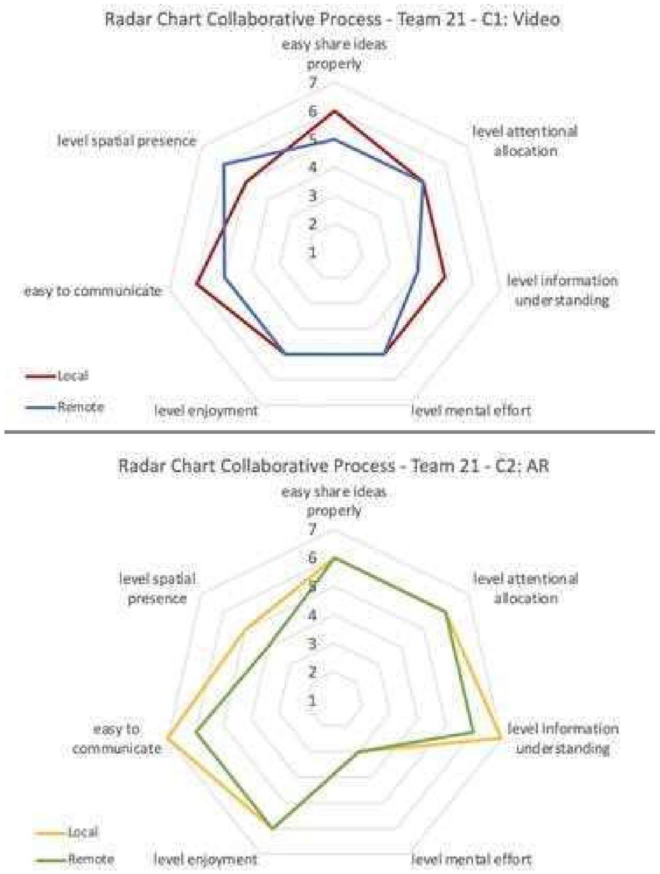


**Fig. 12. Collaborative process for the same team during remote maintenance using the two tools:** ***Top - C1: video chat tool; Bottom - C2: AR-based annotation tool*****. Data displayed using a Likert-type scale: 1- Low; 7- High.**

Through the visualization dashboard of the *CAPTURE* toolkit, it is possible to analyse the collaborative process at the end of an evaluation session for a specific team, or set of different teams. In particular, it is possible to analyse the aspects of collaboration obtained from the use of different tools for the elements of the same team, as explored in this study, which is illustrated in Figure 12, through a random selection.

Naturally, following the results presented above, when using condition C2, the team had a better collaborative performance when compared to the results of condition C1. Nevertheless, by analysing the elements of each team individually, such type of visualization allows to identify aspects of collaboration that could be useful to improve over time, or that may be relevant to update in the collaborative tool being used. For example, when using condition C2, the on-site participant rated the level of spatial presence lower. This fact may suggest that in order to improve the feeling of togetherness, the AR-based annotation tool might benefit from including video sharing in its features.

### 6.3. Participants Preferences and Opinion

With respect to participants experience with the tools, 44 reaction cards were selected to characterize condition C1, including 5 neutral, 9 negative and 30 with positive meaning. Likewise, 46 were selected to characterize condition C2, including 3 neutral, 1 negative and 40 with positive meaning (Figure 13). The following top 10 reaction cards represent participants most selected expressions to characterize each condition: C1 - accessible, collaborative, helpful, flexible, simplistic, familiar, usable, unrefined, expected and time-consuming; C2 - helpful, empowering, collaborative, appealing, easy-to-use, engaging, flexible, novel, innovative and advanced.

However, when analysing participants emotional state, collected before and after the tasks, a clearer perspective is attained. To elaborate, regarding condition C1, participants emotional state before the study varied among joy (11 out of 26), surprise (3 out of 26), excitement (8 out of 26) and contempt (4 out of 26) (Figure 14 - top). Then, after the study, it varied among joy (7 out of 26), surprise (1 out of 26), excitement (1 out of 26) and contempt (17 out of 26) (Figure 14 - top). As for condition C2, participants emotional state before the study varied among joy (12 out of 26), surprise (3 out of 26), excitement (7 out of 26) and contempt (4 out of 26) (Figure 14 - bottom). Then, after the study, it varied among joy (6 out of 26), surprise (4 out of 26) and excitement (6 out of 26) (Figure 14 - bottom).

Hence, it is possible to verify that for condition C1, there was a decrease in the number of participants feeling joy, surprise and excitement at the end of the study, which lead to a significant rise associated to the emotional state of contempt. Contrarily, regarding condition C2, there were no occurrences

**Fig. 13. Participants total reaction cards regarding the collaborative tools.** ***C1: video chat tool; C2: AR-based annotation tool.*** **A larger font size means that the word was selected by more participants (higher frequency). Red - negative meaning; gray - neutral meaning; green - positive meaning [41].**

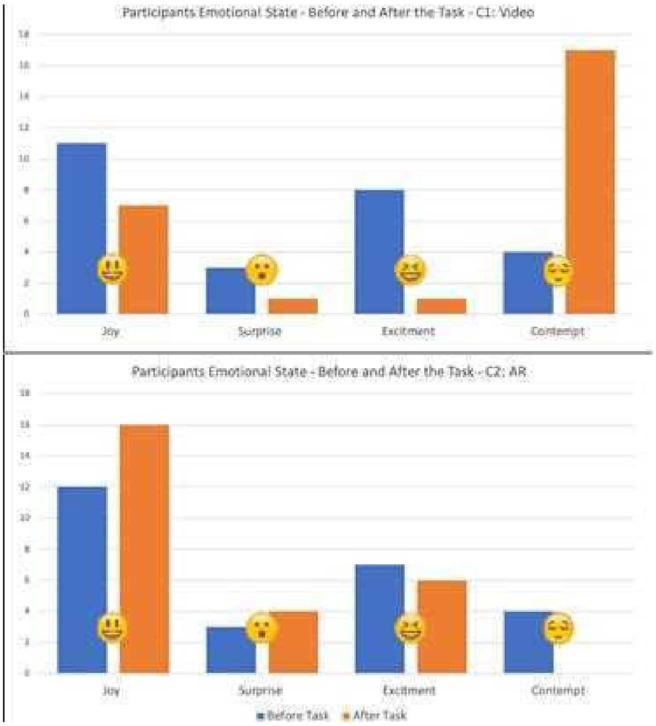


**Fig. 14. Participants emotional state before (top) and after (bottom) the tasks for each condition.** ***C1: video chat tool; C2: AR-based annotation tool.***

of contempt, while joy and surprise had higher number of participants expressing those feelings. As for excitement, although the number of participants that reported such feeling is lower, it is very close to the values reported at the beginning of the study. As such, condition C2 presents significant higher values for emotions correlated with positive connotation, e.g., joy, surprise and excitement when compared to condition C1, which only presents a higher value for contempt (neutral connotation).

In addition, Figure 15 presents participants satisfaction regarding the collaborative tools used through a box plot representation, which illustrates clearly that condition C2 was preferred when compared to condition C1, following the analysis statement of participants emotional state.

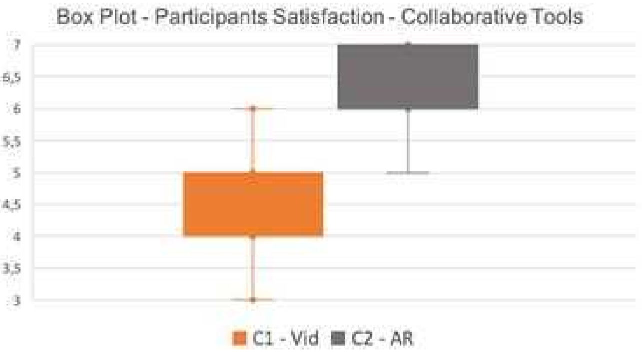


**Fig. 15. Participants satisfaction towards the tools.** ***C1: video chat tool; C2: AR-based annotation tool.*** **Data displayed using a Likert-type scale: 1-Low; 7- High.**

The interviews conducted at the end of the study also emphasize that the majority of participants preferred condition C2, since it enabled seeing non-verbal cues aligned with the task context, which they mentioned contributed to express themselves better through the augmented features, while also having a greater perception of where to perform a given action.

Next, some comments by the participants are presented to provide additional context to the statement previously made:

- regarding the **level of attentional allocation and information understanding** with condition C1, one participant emphasized the following: *"although the video tool is more familiar and quicker to start collaborate, when I needed to express myself about the equipment components or the tools I should use, that's when I started noticing the lack of support. This lead me to repeat the same ideas in different ways to properly explain the desired goal, and the same also happened to my colleague"*;

- as for the **level of mental effort** with condition C1, another participant outlined that *"besides the use of voice, the absence of support to highlight an area of interest or express myself when using the video tool makes me prefer the use of AR-based annotations, in particular for more complex procedures, even though it was a novelty to me and I needed to learn and adapt to it."*;

- concerning **easy to share ideas properly** with condition C2, a different participant reported that *"the use of AR-based annotations allowed me to interact more naturally, while also better comprehend where to perform a given* *action"*. In regards to **attentional allocation** with condition C2, the same participant commented that *"having the handheld device displaying the annotations near the equipment, allowed me to perform the maintenance tasks easily when compared to the video, since in this last there was no content besides the text chat I could use to remember what to do, or to confirm my actions"*;

- with respect to the **level of mental effort and spatial presence**, an additional participant mentioned the following regarding condition C2: *"since I'm familiarized with remote video tools in my daily activities, I was expecting that the absence of video would affect my collaboration with the remote expert. Nevertheless, since the AR-based tool focused more on the task itself, I was engaged in such a way, that not viewing the expert did not affect me at all"*.

Regarding additional comments/suggestions, some participants (7 out of 26) emphasize condition C2 could help create documentation for scenarios where identical tasks may occur. Actually, the AR-based tool already supports revisiting existing annotations, a feature identified as useful by industry partners during the design of a remote maintenance support platform [52]. Nevertheless, for the case study reported, such feature was not made available, since the tasks used did not imply repeating particular activities.

Another topic raised by some participants (18 out of 26) was the possible inclusion of Head Mounted Displays (HMDs), which they consider may further enhance their performance, since it supports a hands-free setting. Likewise, the AR-based tool used already supports HMD, as described in prior work [53]. Since our goal was not to compare different set-ups, we decided not to include such type of device at this moment.

In addition, 4 out of 26 participants referred to possible limitations regarding the use of mobile devices as means to answer a questionnaire, since they were used to doing so on computers. They reported that for questions using drop-down menus and multiple-choice options it was easy to select the desired answer. As for the ones requiring text entry, the process could be slower and tiring. Yet, they understood the usefulness/relevance due to the fact of monitoring real-life scenarios. For example, CAPTURE is ready to be used in industry contexts, in which most technicians may find themselves without a computer. Furthermore, having these target users answering relevant questions after the tasks provides more useful insights than having them filling the questionnaires at the end of a workday on a more suitable device for writing. In this vein, we argue a compromise was required and that the solution provided takes these constraints into consideration. Nevertheless, this also opens new opportunities to propose novel forms of providing input in such scenarios. Furthermore, following the possible inclusion of HMDs in such scenarios and their similar (or even worst) capacity to answer questionnaires, this is also an open topic. Although it is possible to create text with such devices, e.g., hand interaction, literature shows it may not be the best approach. An alternative may be to use a keyboard linked to the HMD device just for answering the existing questionnaires, or perhaps, support voice/sound data collection, and later convert that into text, either via automatic or semi-automatic means. Nevertheless, more than ease of filling out questionnaires, what really matters is evaluating and monitoring collaboration in the best way possible. Therefore, as mentioned, on-the-fly feedback is essential. The choice of the most adequate input form for collecting information from the questionnaires may depend on the hardware available/being used, or on the person designing the study. Overall, the idea is that the toolkit is flexible enough to support all these options.

Last, a reduced number of participants (5 out of 26) suggested viewing the remote expert, not as a basic feature, but as an option for specific cases which may help increase empathy and trust during the collaboration process.

### 6.4. Final remarks

To summarize the added value of our proposal, and how it compares to existing approaches, the conceptual framework instantiated through the *CAPTURE* toolkit allows to retrieve additional amounts of contextual data, as well as selected aspects of collaboration according to the evaluation scope, (usually ignored in existing evaluations found in literature), for more comprehensive analysis using the visualization dashboard.

Another aspect that must be emphasized, is the capacity to adapt to the available data collection instruments. Although self-report was used to gather the emotional response, *CAPTURE* can adapt to support the inclusion of external sensors (e.g., biomedical devices), if necessary for different scenarios.

With all things considered, it is possible to better understand the phenomenon, i.e., recognize when selected aspects of collaboration affect the work effort. By having these insights, it is possible to more easily identify key issues that need to be tackled to ensure a proper shared understanding is attained by distributed team-members in future sections of remote collaboration. By doing so, the research community can evolve from simple evaluations on how technology works, to more complex evaluations aimed to capture a better perspective on the different factors of collaboration supported by AR, which may lead to a more effective collaborative process over time. Hence, we have shown that a better characterization of the collaborative process can be successfully used to provide an additional perspective on the nuances of remote collaboration mediated by AR, which without contextual data would not be possible.

Altogether, due to the flexibility and range of the proposed conceptual model, the instrumentation through the *CAPTURE* toolkit establishes itself as a general-purpose evaluation approach, providing data that otherwise would be difficult to obtain and analyze. While we must be prudent with generalizing our findings, we expect our insights to be valuable for future reproduction in other domains beside maintenance context.

To finish, the continuous observation of contextual data in other tools and with other users may allow, in the future, to create guidelines, supported by experimental data, which can guide the initial development of novel collaborative solutions.

## 7. Conclusions and Future Work

As a contribution, a critical analysis on collaborative user studies mediated by AR is presented, showing that most studies rely on single-user methods, not adapted to collaborative scenarios and that existing frameworks are not well suited to characterize how collaboration occurs. Motivated by these, we presented a conceptual framework to support researchers in designing better evaluations based on retrieving contextualized data for more comprehensive analysis of the collaborative process.

To instantiate this framework, the *CAPTURE* toolkit was proposed to assist with more user centric evaluations, allowing to easily analyze the collaborative process of a particular team or comparison between a set of teams or different tools. During the analysis of the results obtained, it was possible to realize that the contextual data allowed us to understand participants ease to communicate and to share ideas, and the level of attention allocation, spatial presence or others. Also, measure emotional state, and reaction towards the tools used. In this vein, participants felt AR supports more natural interaction, which contributes to increase empathy, interest and collaboration.

By having a grasp on these aspects, typically not reported in the literature, but which are very informative/valuable to understand where the focus of the work, it is possible to better define how research should progress and how the tools can evolve. Hence, conduct comparative analysis of distributed teams may benefit researchers in better understanding the collaborative phenomenon, when compared to how its being currently reported, designing novel methods and improve the collaborative effort. This reinforces, once again, the need to evolve and make these experiences more contextualized and better reported, so that the research community can move into a phase of producing guidelines for remote scenarios supported by AR.

Later, we intend to support data/voice collection, both during the collaborative process among the remote team members, and as an additional data input during the post-task assessment. We envision it may be relevant for researchers having metrics that can be automatically calculated and brought for analysis through an updated version of the visualization dashboard, e.g., characteristics of the dialog, during synchronous collaboration (e.g., number of questions, interruptions, occurrences of specific words). One possible way being considered is supporting some form of synchronization so that all user-related events are synchronized with video/voice streams captured in the study.

Furthermore, we plan to share the toolkit with the research community, which may elicit newer data gathering/visualization requirements. Also, conduct field studies with experts from the industry sector to demonstrate the framework use in real scenarios. Last, pursue the creation of guidelines to elicit more complete evaluations in such scenarios.

## Acknowledgments

We would like to thank the reviewers for their thoughtful comments and suggestions towards improving on an earlier version of this manuscript. To everyone involved in case studies, and discussion groups, thanks for your time and expertise, in particular Pedro Cruz and António Rocha, as well as Andreia Santos, Francisco Marques and Júlia Santos. This research was developed in the scope of the PhD grant [SFRH/BD/143276/2019], funded by FCT - Foundation for Science and Technology. It was also supported by IEETA - Institute of Electronics and Informatics Engineering of Aveiro, funded by National Funds through FCT, in the context of the project [UID/CEC/00127/2019]. The user study of this research was possible due to an industrial collaboration under the Smart Green Homes Project [POCI-01-0247-FEDER-007678], a co-promotion between Bosch Termotecnologia S.A. and the University of Aveiro. It is financed by Portugal 2020 under COMPETE 2020, and by the European Regional Development Fund.

## References

[1] Kim, K, de Melo, CM, Norouzi, N, Bruder, G, Welch, GF. Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. Conference on Virtual Reality and 3D User Interfaces, IEEE VR 2020;:529–538.

[2] Kim, K, Billinghurst, M, Bruder, G, Duh, HB, Welch, GF. Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008-2017). IEEE Transactions on Visualization and Computer Graphics 2018;24:2947–2962.

[3] Kim, S, Billinghurst, M, Lee, G. The effect of collaboration styles and view independence on video-mediated remote collaboration. Computer Supported Cooperative Work: CSCW: An International Journal 2018;27(3-6):569–607.

[4] Gervasi, R, Mastrogiacomo, L, Franceschini, F. A conceptual framework to evaluate human-robot collaboration. The International Journal of Advanced Manufacturing Technology 2020;108(3):841–865.

[5] Lukosch, S, Billinghurst, M, Alem, L, Kiyokawa, K. Collaboration in Augmented Reality. In: Computer Supported Cooperative Work, CSCW 2015; vol. 24. 2015, p. 515–525.

[6] Schneider, M, Rambach, J, Stricker, D. Augmented reality based on edge computing using the example of remote live support. In: 2017 IEEE International Conference on Industrial Technology. 2017, p. 1277–1282.

[7] Billinghurst, M, Clark, A, Lee, G. A Survey of Augmented Reality. Foundations and Trends in Human–Computer Interaction 2015;8:73–272.

[8] Jalo, H, Pirkkalainen, H, Torro, O, Kärkkäinen, H, Puhto, J, Kankaanpää, T. How Can Collaborative Augmented Reality Support Operative Work in the Facility Management Industry? In: Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. 2018, p. 41–51.

[9] Ens, B, Lanir, J, Tang, A, Bateman, S, Lee, G, Piumsomboon, T, et al. Revisiting Collaboration through Mixed Reality: The Evolution of Groupware. International Journal of Human-Computer Studies 2019;131:81–98.

[10] Wang, P, Bai, X, Billinghurst, M, Zhang, S, Zhang, X, Wang, S, et al. Ar/mr remote collaboration on physical tasks: A review. Robotics and Computer-Integrated Manufacturing 2021;72:102071.

[11] Aschenbrenner, D, Rojkov, M, Leutert, F, Verlinden, J, Lukosch, S, Latoschik, ME, et al. Comparing different augmented reality support applications for cooperative repair of an industrial robot. IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct) 2018;:69–74.

[12] Bottani, E, Vignali, G. Augmented reality technology in the manufacturing industry: A review of the last decade. IISE Transactions 2019;51(3):284–310.

[13] Wang, X, Ong, SK, Nee, AYC. A comprehensive survey of augmented reality assembly research. Advances in Manufacturing 2016;4(1):1–22.

[14] Hall, M, McMahon, CA, Bermell-Garcia, P, Johansson, A, Ravindranath, R. Capturing synchronous collaborative design activities: A state-of-the-art technology review. In: Proceedings of International Design Conference, DESIGN 2018. 2018, p. 347–358.

[15] Lee, G, Kang, H, Lee, J, Han, J. A user study on view-sharing techniques for one-to-many mixed reality collaborations. IEEE Conference on Virtual Reality and 3D User Interfaces, IEEE VR 2020;:343–352.

[16] Ludwig, T, Stickel, O, Tolmie, P, Sellmer, M. shARe-IT: Ad hoc Remote Troubleshooting through Augmented Reality. Computer Supported Cooperative Work (CSCW) 2021;30(1):119–167.

[17] Gurevich, P, Lanir, J, Cohen, B. Design and implementation of teleadvisor: a projection-based augmented reality system for remote collabora-

tion. Computer Supported Cooperative Work: CSCW: An International Journal 2015;24(6):527–562.

[18] Zigart, T, Schlund, S. Evaluation of augmented reality technologies in manufacturing - a literature review. In: Proceedings of the AHFE 2020 International Conference on Human Factors and Ergonomics. 2020, p. 75–82.

[19] Fakourfar, O, Ta, K, Tang, R, Bateman, S, Tang, A. Stabilized annotations for mobile remote assistance. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016, p. 1548–1560.

[20] Kim, S, Billinghurst, M, Lee, C, Lee, G. Using freeze frame and visual notifications in an annotation drawing interface for remote collaboration. KSII Transactions on Internet and Information Systems 2018;12(12):6034–6056.

[21] Kim, S, Lee, G, Huang, W, Kim, H, Woo, W, Billinghurst, M. Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019, p. 1–13.

[22] Kim, S, Lee, G, Billinghurst, M, Huang, W. The combination of visual communication cues in mixed reality remote collaboration. Journal on Multimodal User Interfaces 2020;14(4):321–335.

[23] Neale, DC, Carroll, JM, Rosson, MB. Evaluating computer-supported cooperative work: Models and frameworks. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work. CSCW '04; 2004, p. 112–121.

[24] Hamadache, K, Lancieri, L. Strategies and taxonomy, tailoring your CSCW evaluation. In: Groupware: Design, Implementation, and Use. 2009, p. 206–221.

[25] Antunes, P, Herskovic, V, Ochoa, SF, Pino, JA. Reviewing the quality of awareness support in collaborative applications. Journal of Systems and Software 2014;89:146–169.

[26] Merino, L, Schwarzl, M, Kraus, M, Sedlmair, M, Schmalstieg, D, Weiskopf, D. Evaluating Mixed and Augmented Reality: A Systematic Literature Review (2009–2019). In: IEEE International Symposium on Mixed and Augmented Reality, ISMAR. 2020,.

[27] Belen, RAJ, Nguyen, H, Filonik, D, Favero, DD, Bednarz, T. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. AIMS Electronics and Electrical Engineering 2019;3:181.

[28] Marques, B, Teixeira, A, Silva, S, Alves, J, Dias, P, Santos, BS. A Conceptual Model for Data Collection and Analysis for AR-based Remote Collaboration Evaluation. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct). 2020, p. 1–2.

[29] Marques, B, Teixeira, A, Silva, S, Alves, J, Dias, P, Santos, BS. A critical analysis on remote collaboration mediated by augmented reality: Making a case for improved characterization and evaluation of the collaborative process. Computers & Graphics 2021;:1–17.

[30] Bai, Z, Blackwell, AF. Analytic review of usability evaluation in ISMAR. Interacting with Computers 2012;24(6):450 – 460.

[31] Ratcliffe, J, Soave, F, Bryan-Kinns, N, Tokarchuk, L, Farkhatdinov, I. Extended Reality (XR) Remote Research: a Survey of Drawbacks and Opportunities. In: CHI Conference on Human Factors in Computing Systems. 2021, p. 1–13.

[32] Dey, A, Billinghurst, M, Lindeman, RW, Swan, JE. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. Frontiers in Robotics and AI 2018;5:37.

[33] Anton, D, Kurillo, G, Bajcsy, R. User experience and interaction performance in 2d/3d telecollaboration. Future Generation Computer Systems 2018;82:77–88.

[34] Piumsomboon, T, Lee, G, Ens, B, Thomas, B, Billinghurst, M. Superman vs giant: A study on spatial perception for a multi-scale mixed reality flying telepresence interface. IEEE Transactions on Visualization and Computer Graphics 2018;24(11):2974–2982.

[35] Araujo, RM, Santoro, FM, Borges, MRS. A conceptual framework for designing and conducting groupware evaluations. International Journal of Computer Applications in Technology 2004;19(3):139–150.

[36] Araujo, RM, Santoro, FM, Borges, MRS. The CSCW lab ontology for groupware evaluation. In: 8th International Conference on Computer Supported Cooperative Work in Design; vol. 2. 2004, p. 148–153.

[37] Marques, B, Silva, S, Dias, P, Santos, BS. An Ontology for Evaluation of Remote Collaboration using Augmented Reality. In: European Conference on Computer-Supported Cooperative Work, ECSCW: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies - Posters & Demos, Reports of the European Society for Socially Embedded Technologies. 2021, p. 1–8.

[38] Marques, B, Silva, SS, Alves, J, Araujo, T, Dias, P, Santos, BS. A conceptual model and taxonomy for collaborative augmented reality. IEEE Transactions on Visualization & Computer Graphics 2021;:1–18.

[39] Marques, B, Ara, T, Silva, S, Dias, P. Visually exploring a Collaborative Augmented Reality Taxonomy. In: Information Visualization, IV. 2021, p. 1–6.

[40] Izard, CE. Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. Perspectives on Psychological Science 2007;2(3):260–280.

[41] Barnum, CM. Usability Testing Essentials: Ready, Set...Test! 1st ed.; San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2010.

[42] Gutwin, C, Greenberg, S. The effects of workspace awareness support on the usability of real-time distributed groupware. ACM Transactions on Computer-Human Interaction 1999;6(3):243–281.

[43] Nilsson, S, Johansson, B, Jonsson, A. Using ar to support cross-organisational collaboration in dynamic tasks. IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2009;:3–12.

[44] Kim, S, Lee, G, Sakata, N, Billinghurst, M. Improving co-presence with augmented visual communication cues for sharing experience through video conference. ISMAR 2014 - IEEE International Symposium on Mixed and Augmented Reality - Science and Technology 2014, Proceedings 2014;:83–92.

[45] Kim, S, Billinghurst, M, Lee, C, Lee, G. Using freeze frame and visual notifications in an annotation drawing interface for remote collaboration. KSII Transactions on Internet & Information Systems 2018;12:6034–6056.

[46] Huang, W, Kim, S, Billinghurst, M, Alem, L. Sharing hand gesture and sketch cues in remote collaboration. Journal of Visual Communication and Image Representation 2019;58:428–438.

[47] Patel, H, Pettitt, M, Wilson, JR. Factors of collaborative working: A framework for a collaboration model. Applied ergonomics 2012;43(1):1–26.

[48] Johnson, S, Gibson, M, Mutlu, B. Handheld or handsfree? remote collaboration via lightweight head-mounted displays and handheld devices. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2015, p. 1825–1836.

[49] Teo, T, Lee, G, Billinghurst, M, Adcock, M. Investigating the use of different visual cues to improve social presence within a 360 mixed reality remote collaboration. ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry 2019;.

[50] Piumsomboon, T, Dey, A, Ens, B, Lee, G, Billinghurst, M. The effects of sharing awareness cues in collaborative mixed reality. Frontiers Robotics AI 2019;6.

[51] Marques, B, Silva, S, Dias, P, Santos, BS. A Toolkit to Facilitate Evaluation and Characterization of the Collaborative Process in Scenarios of Remote Assistance Supported by AR. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct). 2021, p. 336–337.

[52] Marques, B, Silva, S, Rocha, A, Dias, P, Santos, BS. Remote asynchronous collaboration in maintenance scenarios using augmented reality and annotations. In: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), IEEE VR 2021. 2021, p. 567–568.

[53] Madeira, T, Marques, B, Alves, J, Dias, P, Santos, BS. Exploring annotations and hand tracking in augmented reality for remote collaboration. In: Human Systems Engineering and Design III. Springer International Publishing; 2021, p. 83–89.