Remote Testing of an Augmented Reality System

Marja Liinasuo
VTT Technical Research Centre of Finland Ltd
marja.liinasuo@vtt.fi

Timo Kuula
VTT Technical Research Centre of Finland Ltd
timo.kuula@vtt.fi

Vladimir Goriachev
VTT Technical Research Centre of Finland Ltd
vladimir.goriachev@vtt.fi

Kaj Helin
VTT Technical Research Centre of Finland Ltd
kaj.helin@vtt.fi

ABSTRACT

Remote test settings have become more common due to COVID-19. Our paper presents two user tests focusing on the usability and user experience of an augmented reality-based solution, i.e., augmented reality system. We describe the proceeding of the tests from the perspective of what party has participated in the test in the same location as the test participant, i.e., locally, and what party remotely. The importance - or unimportance - of physical presence is contemplated from the perspective of the successfulness of the test. The physical presence of a person providing technical support to the test participant during the testing proved vital for the augmented reality related testing; the location of other test organisers appears more indifferent in this context.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **HCI design and evaluation methods**; • **User studies**;

KEYWORDS

Augmented reality, Testing methodology, Usability, User experience

ACM Reference Format:
Marja Liinasuo, Timo Kuula, Vladimir Goriachev, and Kaj Helin. 2022. Remote Testing of an Augmented Reality System. In *33rd European Conference on Cognitive Ergonomics (ECCE 2022), October 04–07, 2022, Kaiserslautern, Germany.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3552327.3552337

1 INTRODUCTION

Scientific testing or evaluation usually takes place in dedicated facilities such as in a laboratory or in some more or less public (opposed to private) locations. Information collection is also being done in other premises. Remote testing is defined as testing in a situation where the test leader or moderator is separated in space and/or time from test participants (following loosely the definition by Andreasen et al. [1]). The idea of conducting usability testing remotely emerged already at the early 90's [2]. The increased number of commonly available software for collaborative activities has nowadays enabled remote testing in a large scale. Recently, COVID-19 has forced scientific testing to be done remotely.

Testing can be conducted in various locations. For instance, remote sensory testing has been successfully carried out at assessor's home or workplace [3], instead of a laboratory where it is usually is located. This has required the live online supervision of the test leader. In the review of Holland et al. [4], it was found the people with chronic lung disease have conducted an exercise test successfully with supervision at home or supported by remote administration.

Remote testing can set requirements to the test participants. Remote sensory testing proved to be feasible with trained panellists and was suggested to be useful also with consumers [3]. This probably means that the testing procedure in question was not too complicated and could be facilitated also via a videoconference or the like, with limited view of sight and without the possibility to guide in a hands-on manner.

Test subject may be sensitive by nature. Somewhat surprisingly, both literature and practice support the usage of telemedicine assessments for patients with cognitive impairment, even if guidance is lacking [5]. Probably the need for such testing combined with the availability of appropriate technical devices and the acceptance by both patients and caregivers have promoted the concept of remote testing in this context. The main deficiency seems to be the actual testing palette – the specific norms for remote testing must be set and the validity of such tests must be assessed.

Testing may include artifacts, which require professional handling. In testing focusing on sensing, samples should be handled and shipped without any hazard for participant safety and without biasing effects on the sample itself [3]. In human-robot interaction (HRI) studies, robot is sometimes controlled remotely. Studying is challenging if physical proximity between these key parties cannot be enabled [6].

Testing procedure can also be difficult, hampering remote testing. All tests of lung functioning are not appropriate to be conducted at home as not all tests document accurately desaturation with walking [4]. Consequently, it was recommended in that study that patients at risk of desaturation should be prioritised for centre-based testing when possible.

On the other hand, remote testing can provide benefits not existing during "normal" testing. For instance, conducting consumer testing remotely, close to the consumer, can be more persuasive for the test participant candidate as there is then no need to travel to testing facilities (e.g., [3]).

New guidelines have been called for in remote testing in many domains, such as education [7] or clinical assessment related to

ECCE 2022, October 04–07, 2022, Kaiserslautern, Germany

ACM ISBN 978-1-4503-9808-4/22/10.
https://doi.org/10.1145/3552327.3552337

cognitive impairment [5]. Appropriate guidelines are easier to produce after having gathered enough experiences of remote testing. That way challenges are identified and a way to deal with them can be invented.

This paper presents two similar usability and user experience tests with Microsoft HoloLens 2 mixed reality smart glasses, conducted partly remotely and partly as face-to-face, and scrutinises the appropriateness of the used setting.

Microsoft HoloLens 2 is built for interaction with three-dimensional (3D) models, a feature needed in both test applications. The purpose of ESA test [8] was to conduct a preliminary user review of the first integrated prototype of the developed system. The motivation for BIMprove test [9] was to acquire first usability and user experience related opinions and experiences for developing further the preliminary version of the augmented reality (AR) system.

ESA (European Space Agency) projects AROGAN (Augmented Reality-based Orbit and ground ApplicatioNs) and VirWAIT (Virtual Workplace for AIT & PA Training and Operations Support) provided an opportunity to develop a mixed-reality system for assembly, integration, testing, and verification activities in the space [10].

In the EU project BIMprove (Improving Building Information Modelling by Realtime Tracing of Construction Processes) [11], we developed an augmented reality (AR) system to be used in the building trade by various professionals in the field [9].

After the background presented in 'Introduction', we describe the methods for testing in the AR context in 'Methods'. In 'Appropriateness of research methods' we go through the appropriateness of the methods, from the perspective of remote vs physically present testing. In 'Discussion', we contemplate the results as such and from the perspective of the related scientific literature. Finally, we draw conclusions in the last section 'Conclusions'.

2 METHODS

We performed testing mainly remotely in two separate projects but in a similar manner. The methods of thinking aloud, observation, questionnaire and interview were used for studying usability and user experience pertaining to an AR solution (see BIMprove study [9] for a detailed description of the used methods). Testing was preliminary by nature, to remove the most obvious usability flaws before the actual testing with end users. In ESA study [8], instructions to be presented to astronauts in space, shown in AR, were tested. The instructions on how to proceed with different tasks in space were delivered to the test participant ("astronaut") using augmented reality, so no oral instructions were needed in the testing. In BIMprove study [9], digital twin of an imagery building information model (BIM), shown in AR, was tested, to be used by various professionals in the construction phase. Thus, in BIMprove test, guidance on what to do next was necessary. Testing session consisted of separate tasks to conduct with the model in augmented reality.

The term 'technical expert', used in ESA study [8], is here replaced with 'test instructor', because it includes both the technical support during the testing session, realised in both studies, and the guidance in the testing session and expert involvement in the post-test interview, the two latter activities pertaining to BIMprove study only.

2.1 Test participants

There were 2 test participants in ESA test and 4 in BIMprove test. In ESA, the test participants were ESA experts, and one of them was familiar with the application beforehand. In BIMprove, the test participants were researchers from various fields. One had supported the development of the AR application without seeing it and was highly familiar with the HMD (head-mounted display) device; one was not familiar with the application but highly familiar with the HMD device; and other participants were not familiar with the application nor with the HMD device.

2.2 Test organisers

In addition to the test participants, also other roles were needed in the test. They are coined here as test organisers. The test leader dictated when and how to proceed during the testing session and conducted the test interview. The role of the test instructor was to act as technical support, aiding the test participant in using the AR system and, additionally in BIMprove test, informing the test participant about the next task. The test assistant wrote down the verbal expressions of the test participant during the testing session (thinking aloud) and the responses to interview questions.

2.3 Proceeding of the test

Testing proceeded similarly in both tests. The only difference is in the way the test is executed, explained in the list below (in point 4). The test was performed separately to each participant. The whole testing session (points 2-5 in the following list) was audible and visible also to the test organisers not physically present in the test room (the test leader and the test assistant), through a videoconference. In the following, the test is described in more detail (Figure 1).

- Before the test, we sent a usability questionnaire (System Usability Scale, SUS) to the test participants by e-mail.
- The actual test session starts. The test leader introduces the agenda, test goals and methods to the test participant in the test room, utilising videoconference. The test assistant is also present in the videoconference but with a muted microphone. The test instructor is physically present with the test participant in the test room.
- Test leader shows a video about how to use the AR application through a videoconference. The test assistant is also present in the videoconference with a muted microphone. The test instructor has no role at this point.
- The test is executed. The test participant expresses verbally his/her thoughts and experiences during the testing of the system (the think-aloud method). The behaviour of the test participant is mediated to the test leader and the test assistant by a videoconference. The task of leading the test is divided between the test leader and the test instructor. The test leader prompts the test participant to think aloud when it seems to be forgotten. The test instructor helps the participant in using the AR system when needed. The test assistant writes down the test participant's verbal expressions. Additionally,

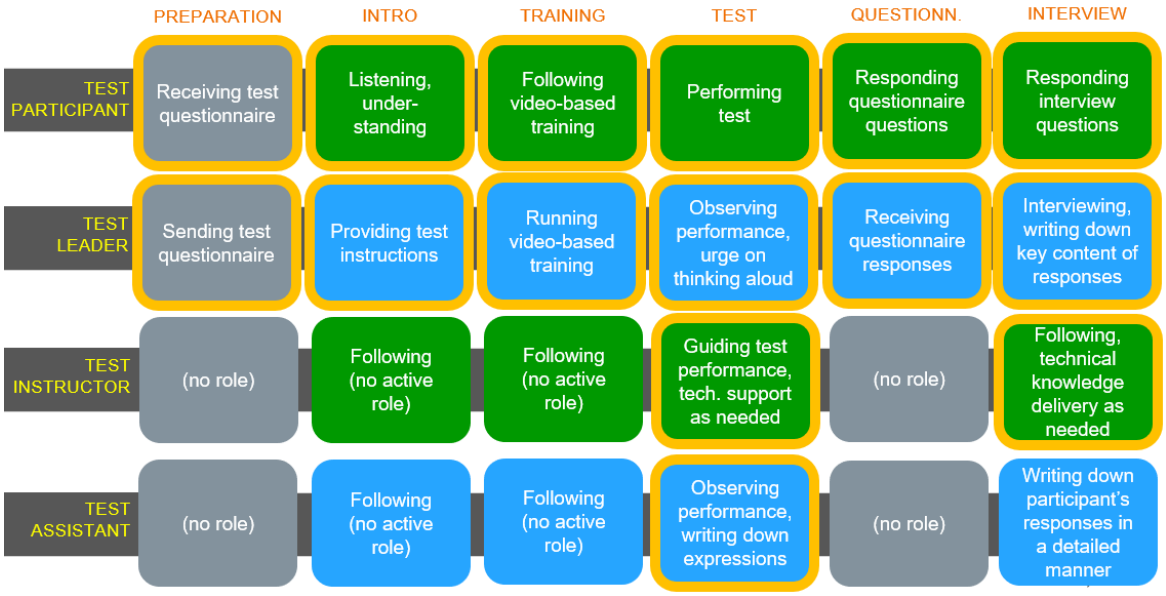

Figure 1: Overview of our hybrid test setting. The roles participating in the testing are indicated in the left and the test phase is written upmost in each column. Green means presence in the test facilities, pertaining practically to the test participant and the test instructor only, blue means telepresence (videoconference), valid for the test leader and the test assistant throughout the testing, and grey means that the location is indifferent from the perspective of the test methodology, pertaining to the test preparation and whenever the role has no task in the phase in question. Key roles in each phase, from the research perspective, are indicated with an outline (yellow).

in BIMprove test [9], the test instructor, physically present in the same room with the test participant, informs the test participant about the next task with the AR application, one task at a time.

- The test participant fills in the usability questionnaire (SUS) after the test.
- The test leader interviews about usability and user experience using a videoconference. The test assistant, also present online, writes the responses down and asks additional questions as needed. The test instructor, hearing the interview, clarifies technical matters related to the interview questions and answers when needed.

3 DISCUSSION

We have conducted two studies pertaining to usability and user experience of augmented reality-based solutions; (1) instruction, to be used by astronauts in space [8], and (2) an editable digital twin of a building information model to be used by construction professionals [9]. In this paper, we contemplate the efficiency of remote testing vs testing with physical presence in these tests. We classify the testing as remote in these studies as the test lead was physically separated from the test participant. The setting can also be called as remote moderated [12] or human-moderated remote user testing [13] as irrespective of physical separation, the test lead and the test participant communicated directly during the test. However, as the test instructor co-located with the test participant, the setting can also be coined as hybrid.

In both tests, relevant and interesting results were obtained, supporting the further development of AR solutions (see the related publications [8], [9]). Next, the effect of the location - remote or face-to-face - is contemplated, method by method.

Questionnaire related activities – sending/receiving, filling in, sending back – are not location sensitive.

When testing the AR solution, the location of the test leader and the test assistant appeared relatively indifferent but technical support was needed for using the AR system. The technology-clarifying comments by the test instructor facilitated the interview, ensuring that incorrect assumptions were rectified and unclear matters clarified. Perhaps the presence of the test instructor also provided the needed human closeness. Also, the sharing of the main interview results in a videoconference possibly added the trustworthiness of the test for the test participant.

All in all, hybrid test setting was proved appropriate in these studies, key elements being the physical presence of the test instructor providing technical support during the test, the relative indifference of physical location pertaining to other testing organisers, and the possibility to share main results during the interview through videoconference.

Regarding usability testing, already in 1994 Hammontree et al. [2] suggested technology for remote testing. Since then, supporting means such as protocols (e.g., [13]), accessibility (e.g., [14]), and tools ([12], [15], [16]) for remote testing have been contemplated and tested. As methodological tools are being continuously developed and because the benefits of remote testing have become more familiar to researchers, it is possible that remote and hybrid testing

will become even more common, also after COVID. In literature, hybrid testing is referred to in rather complex test settings (e.g., [17], [18]). Seems to be that the concept is not used in cognitive ergonomics, even if that kind of testing would have been conducted. Perhaps hybrid tests are categorised as belonging to remote testing, such as when performing consumer testing close to the consumer [3]. It can also be that such studies are seldom performed or published.

There is no general advantage of one setting over another [19] so researchers need to be able to make informed decisions among the possible options. Based on literature, the successfulness of different test settings seems to be mainly evaluated by the number and quality of results (as well as the time the testing requires, if appropriate) (e.g., [1], [20], [21], [22]) like in our paper.

4 CONCLUSIONS

The combination of a face-to-face and remote test setting proved to be successful for our AR related studies [[8],[9]. The importance of having a technical expert physically present and in an active role, when demanding technology is used in testing, was identified. The location of other test organisers did not appear as important in our case. Videoconferencing provided valuable means to manage testing and share main interview results online.

Cognitive-ergonomics-based literature seems to emphasise only the existence of the remote and face-to-face settings. Our results show that a more fine-tuned, hybrid approach can be fruitful when designing a test.

The results would have been more convincing if a comparison test had been conducted, for instance, with a completely remote or complete onsite setting. Also, the appropriateness of the methods should be systematically assessed instead of partly concluding it, as was performed in this study. This kind of study remains to be done in the future.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme as the project BIMprove, under grant agreement No 958450.

REFERENCES

[1] Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. What happened to remote usability testing? An empirical study of three methods. In Proceedings of the SIGCHI conference on Human factors in computing systems, 1405-1414.

[2] Monty Hammontree, Paul Weiler, and Nandini Nayak. 1994. Remote Usability Testing. interactions, 1(3), 21-25.

[3] Caterina Dinnella, Lapo Pierguidi, Sara Spinelli, Monica Borgogno, Tullia GallinaToschi, Stefano Predieri, Giliana Lavezzi, Francesca Trapani, Matilde Tura, Massimiliano Magli, Alessandra Bendini, and Erminio Monteleone. 2022. Remote testing: Sensory test during Covid-19 pandemic and beyond. Food quality and preference, 96, 104437. https://www.sciencedirect.com/science/article/pii/S0950329321003190

[4] Anne E. Holland, Carla Malaguti, Mariana Hoffman, Aroub Lahham, Angela T. Burge, Leona Dowman, Anthony K. May, Janet Bondarenko, Marnie Graco, Gabriella Tikellis, Joanna Y.T. Lee, and Narelle S. Cox. 2020. Home-based or remote exercise testing in chronic respiratory disease, during the COVID-19 pandemic and beyond: a rapid review. Chronic respiratory disease, 17, 1-18. https://doi.org/10.1177/1479973120952418

[5] Maiya R. Geddes, Megan E. O'Connell, John D. Fisk, Serge Gauthier, Richard Camicioli, Zahinoor Ismail, and Alzheimer Society of Canada Task Force on Dementia Care Best Practices for COVID-19. 2020. Remote cognitive and behavioral assessment: Report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID-19. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 12(1), e12111. https://doi.org/10.1002/dad2.12111

[6] Curtis L. Gittens. 2021. Remote HRI: A methodology for maintaining COVID-19 physical distancing and human interaction requirements in HRI studies. Information Systems Frontiers, 1-16. https://link.springer.com/article/10.1007/s10796-021-10162-4

[7] Karen Sando, Melissa S. Medina, and Karen Whalen. 2021. The Need for New Guidelines and Training for Remote/Online Testing and Proctoring Due to COVID-19. American Journal of Pharmaceutical Education. Commentary. https://www.ajpe.org/content/ajpe/early/2021/01/28/ajpe8545.full.pdf

[8] Kaj Helin, Jaakko Karjalainen, Timo Kuula, Paul Kiernan, Gianluca Casarosa, and David Martinez Oliveira. 2021. The Remote User Test of Mixed Reality System for AIT/AIV in the Space Domain. In EuroXR 2021: Proceedings of the Virtual EuroXR Conference, ed K Helin, J Perret and Arlati S (VTT Technical Research Centre of Finland Ltd) https://sarjaweb.vtt.fi/pdf/technology/2021/T395.pdf, 11-13.

[9] Marja Liinasuo, Timo Kuula, Kaj Helin, Vladimir Goriachev, and Matthias Aust. 2022. Building's digital twin model shown as first-person and third-person views. CIB WBC2022 conference on 27-30 June 2022, Melbourne, Australia.

[10] ESA 2022. Website of the European Space Agency. Retrieved April 20, 2022 from https://www.esa.int

[11] BIMprove 2021. Website for the EU project BIMprove. Retrieved January 28, 2022 from https://cordis.europa.eu/project/id/958450

[12] Felix Giroux, Pierre-Majorique Léger, David Brieu gne, Francois Courtemanche, Frédérique Bouvier, Shang-Lin Chen, Salima Tazi, Emma Rucco, Marc Fredette, Constantinos Coursaris, and Sylvain Sénécal. 2021. Guidelines for collecting automatic facial expression detection data synchronized with a dynamic stimulus in remote moderated user tests. In International Conference on Human-Computer Interaction, 243-254.

[13] Asimina Vasalou, Brian David Ng, Peter Wiemer-Hastings, and Lidia Oshlyansky. 2004. Human-moderated remote user testing: Protocols and applications. In 8th ERCIM Workshop, User Interfaces for All, Wien, Austria, Vol. 19

[14] Estella Oncins.. 2021. Accessibility in Online User Testing. Journal of Audiovisual Translation, 4(2), 6-22.Kathrine E. Thompson, Evelyn P. Rozanski, and Anne R. Haake. 2004. Here, there, anywhere: Remote usability testing that works. In Proceedings of the 5th conference on Information technology education, 132-137.

[15] Kevin Chen, and Haoqi Zhang. 2015, April. Remote paper prototype testing. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 77-80.

[16] Aurélie Vasseur, Pierrre-Majorique Léger, Francois Courtemanche, Elise Labonta-Lemoyne, Vanessa, Geroges, Audrey Valiquette, David Brieugne, Emma Rucco, Constantinos Coursaris, Marc Fredette, and Sylvain Sénécal. 2021 Collection for UX evaluation: A pilot project. In International Conference on Human-Computer Interaction, 255-267.

[17] Andrei M. Reinhorn, Mettupalayam V. Sivaselvan, Zach Liand, and Xiaoyun Shao. 2004. Real-time dynamic hybrid testing of tructural systems. 13th World Conference on Earthquake Engineering, Vancouver, B.C:, Canada, August 1-6, 2004, paper No. 1644.

[18] D. P. McCrum and M. S. William. 2016. An overview of seismic hybrid testing of engineering structures. Engineering Structures, 118, 240-261.

[19] Juergen Sauer, Andreas Sonderegger, Klau Heyden, Jasmin Biller, Julia Klotz, and Andreas Uebelbacher. 2019. Extra-laboratorial usability tests: An empirical comparison of remote and classical field testing with lab testing. Applied Ergonomics, 74, 85-96.

[20] Deborah A. Abowitz, and T. Michael Toole. 2010. Mixed method research: Fundamental issues of design, validity, and reliability in construction research. Journal of construction engineering and management, 136(1), 108.

[21] David R. Johnson, Christopher P. Scheitle, Elaine Howard Ecklund. 2019. Beyond the In-Person Interview? How Interview Quality Varies Across In-person, Telephone, and Skype Interviews. Social Science Computer Review, 39(6), 1142-1158. https://doi.org/10.1177/0894439319893612

[22] Aurélie Vasseur, Pierre-Majorique Léger, Francois Courtemanche, Elise Labonte-Lemoyne, Vanessa Georges, Audrey Valiquette, David Brieugne, Emma Rucco, Constantinos Coursaris, Marc Fedette, and Sylvain Sénécal, S. 2021, July. Distributed remote psychophysiological data collection for UX evaluation: a pilot project. In International Conference on Human-Computer Interaction, 255-267.