



✦ Anthony Steed, University College London, Francisco R. Ortega, Colorado State University, Adam S. Williams, Colorado State University, Ernst Kruijff, Bonn-Rhein-Sieg University of Applied Sciences, Wolfgang Stuerzlinger, Simon Fraser University, Anil Ufuk Batmaz, Simon Fraser University, Andrea Stevenson Won, Cornell University, Evan Suma Rosenberg, University of Minnesota, Adalberto L. Simeone, KU Leuven, Aleshia Hayes, University of North Texas

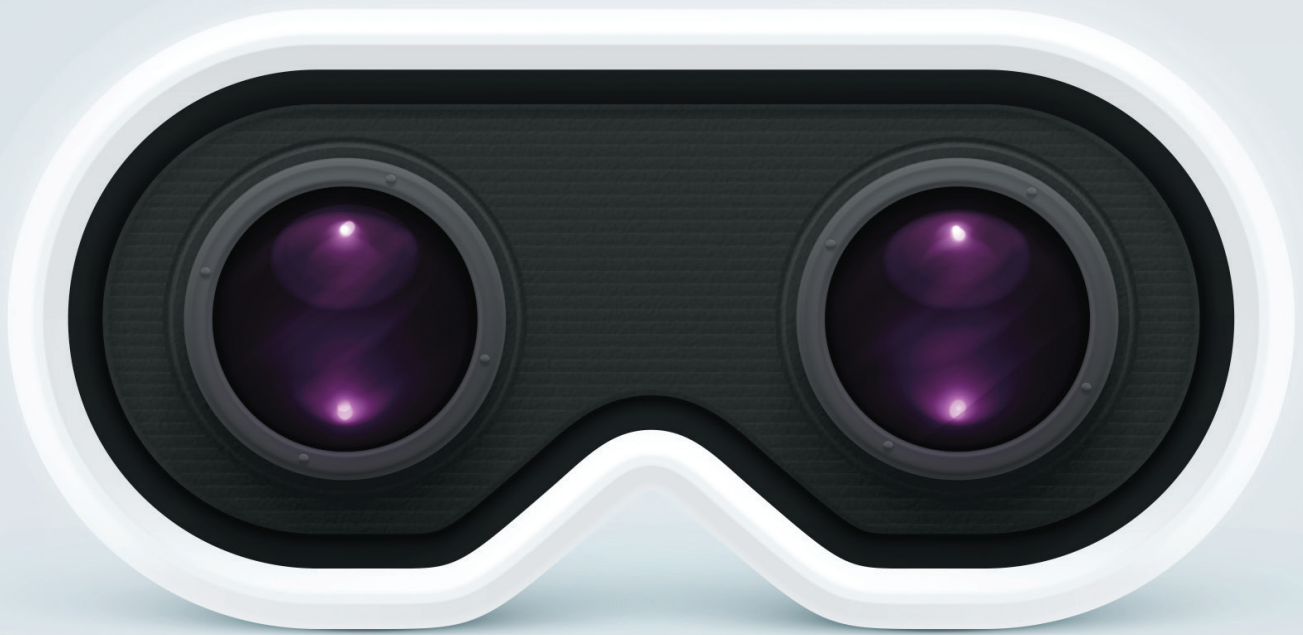
Evaluating Immersive Experiences During Covid-19 and Beyond

The Covid-19 pandemic has disrupted our daily lives. The safety and well-being of people are paramount, and there is no exception for the human-computer interaction (HCI) field. Most universities and research labs have closed non-critical research labs. With that closure and the student populations having left campus, in-person user studies have been suspended for the foreseeable future. Experiments that involve the usage of specialized technology, such as virtual and augmented reality headsets, create additional challenges. While some head-mounted displays (HMDs) have become more affordable for consumers (e.g., Oculus Quest), there are still

multiple constraints for researchers, including the expense of high-end HMDs (e.g., Microsoft HoloLens), high-end graphics hardware, and specialized sensors, as well as ethical concerns around reusing equipment that comes in close contact with each participant and may be difficult to sterilize. These difficulties have led the extended reality (XR) community (which includes the virtual reality (VR) and augmented reality (AR) research communities) to ask how we can continue to practically and ethically run experiments under these circumstances. Here, we summarize the status of a community discussion of short-term, medium-term, and long-

term measures to deal with the current Covid-19 situation and its potential longer-term impacts. In particular, we outline steps we are taking toward community support of distributed experiments. There are a number of reasons to look at a more distributed model of participant recruitment, including the generalizability of the work and potential access to target-specific, hard-to-reach user groups. We hope that this article will inform the first steps toward addressing the practical and ethical concerns for such studies [1].

There are currently no strong ethical guidelines for designing and running experiments in VR and AR.



Most of the VR and AR studies in HCI are conducted in research institutions, where researchers must follow local laws and the directions of the local institution's ethics board. VR and AR systems allow researchers to control the virtual environment and collect detailed user data in ways that might not be familiar to participants, so careful consideration of participant privacy is especially important. Further, some experiments might require direct supervision through an experimenter while the user interacts with the virtual environment, for example, to watch for behaviors that circumvent the objectives of the experiment. The rules and laws for

remote data collection and direct supervision of experiments, which can vary between different countries and regions, become an issue.

SHORT-TERM SOLUTION: USE LAB PERSONNEL AND INFRASTRUCTURE

The most immediate solution to performing remote experiments is to collaborate between labs to provide participants for each other's experiments. The subjects are likely to be lab members, or people associated with the labs in some manner, who have the correct equipment at hand.

A well-known concern for most work with human subjects is the

issue of working with populations of convenience. The problem can be particularly acute in this case. Groups of lab members may have too much knowledge about the field to react "naturally," or in a way unbiased by domain knowledge. They may guess the experimenter's aims and intentionally or unintentionally behave in accordance with or in opposition to them. They also may have strong existing opinions about interaction or visualization techniques, which can bias the outcomes. Finally, their experience with XR—either AR or VR—may make it difficult to generalize their data to the general population. Specifically, their expertise in the use of these platforms can confound the

> A RESPONSIVE KIND OF DESIGN

outcomes of usability testing on new tools and experiences.

However, there are also circumstances in which distributed studies across labs could be better than the usual population of convenience. Rather than a mix of participants who are familiar with XR and new to XR, a population of lab members would all be familiar with the equipment. This might be more generalizable to populations who might actually engage with the research because those populations already use XR. Assuming that lab members in general are less susceptible to simulator sickness (e.g., through self-selection), it could also mean that systems that aren't optimized for consumer release (e.g., because they use very heavy compute resources) might be investigated in the lab. In short, it is important to carefully consider which tasks can be best run with experienced participants.

MEDIUM-TERM SOLUTION: RECRUIT EXTERNAL USERS WHO HAVE THE NECESSARY HARDWARE

To develop a more sustainable participant pool, a more organized effort is needed to start recruiting outside of research labs. This phase would still be limited to participants who have the equipment required for a given user study. However, given that an estimated six million people currently own a VR headset, there is clearly the potential to reach out to these individuals. Unfortunately there are no easy-to-use tools to run VR experiments online, and there are various technical issues with implementing and distributing experiments to consumer devices. A few early works, though, have demonstrated the possibility of controlling enough aspects of the design to produce usable results (e.g., [2,3]).

An initial step would be a website that allowed participants to register for recruitment and research labs to advertise their experiments. This system could use crowdsourcing

websites to recruit participants, who would then be redirected to the site. This in itself brings many challenges. Simultaneous efforts by different regions (e.g., the EU, the U.K., the U.S., Canada, Japan, and Brazil) would be required to grow the system, by collaborating and seeking regional funds. For the site to be successful, different regional needs would need to be considered. For example, a study approved by an ethics board in the U.S. might not be acceptable to a panel in another country. This involves not only ethics but also local laws, such as the European General Data Protection Regulation (GDPR).

LONG-TERM SOLUTIONS: GENERATE POOLS OF USERS THROUGH FUNDED HARDWARE DISTRIBUTION

While the medium-term solution will improve the way we do remote distributed experiments, one suggestion for the long term is to provide equipment to a pool of subjects. Based on the experiences learned in the previous phase, this solution will continue to improve the tools and methods created, but it needs to identify ways of finding participants who can be lent equipment in the hope that they then use it to participate in multiple experiments. This in itself is an expensive goal, but we believe it is possible because the equipment might become cheaper. Some governmental scientific funding bodies (e.g., the National Science Foundation in the U.S.) could provide funds to acquire the required infrastructure to expand the pool. This would offer an opportunity beyond Covid-19. First, it would allow us to have expert users test a VR application while also having access to naive users when needed. It also would enable us to validate research results with different subject pools from multiple regions, and would remove the need to bring participants into the lab—unless this were required due to specialized hardware needs

or other restrictions imposed by an experiment's design.

In other words, we could seed a community of participants with specialized technologies to allow for a more diverse subject pool. This would ideally look like the distribution of HMDs or similar technologies to volunteers around the world, who in exchange would agree to participate in experiments. These seeded participants would be registered through a citizen science crowdsourcing site. The benefit of having these seeded users would be a new level of diversity among participants. The distributed HMDs will be new technologies (at least for some time) and these participants would represent novice users. The research lab would still be able to provide incentives (e.g., additional compensation) to maintain interest among these users. This style of crowdsourcing has some precedents. Some people make a portion of their income from crowdsourcing sites such as Amazon Mechanical Turk and Prolific or by participating in paid medical trials. In some cases, these participants are not ideal test subjects, because they may be very experienced with common experimental tasks or situations (e.g., overfamiliarity with the “trolley problem” in psychological experiments). However, having an official pool of representative, well-compensated participants could also address issues of undercompensation and unrepresentative samples.

ETHICAL CONSIDERATIONS

While ethical (and legal) considerations may vary depending on the country, review board, and institution, the following are some points to consider:

- *Pooling students to run each other's experiments.* While this seems to be an attractive idea, there is the problem that faculty members might induce students to take part in the pool. Thus, it is essential to have strong requirements about there not being an inducement to take part; for example, it cannot affect grades, funding, or progress toward degrees. A possible solution is to add a “non-inducement” clause while checking that it is enforced.

- *Desktop sharing.* If desktop screen sharing is used (e.g., to make it easier for the experimenter to control the remote apparatus), this poses potential

To develop a more sustainable participant pool, a more organized effort is needed to start recruiting outside of research labs.

ethical risks for the participants. For example, the experimenter may then see personal notifications. One potential solution is to use only window sharing for the VR application, if this is viable. Still, it is vital that these risks are assessed and correctly managed by the experimenter and are mentioned in the ethics review process.

- *Running studies on social VR platforms.* This poses several data-processing problems, as the confidentiality and security of emerging platforms is not assured. The platforms may require personal information to sign up, and as users we cannot be certain what data is being collected. While these risks may be alleviated through careful design (e.g., recruiting participants within the platform), they pose new concerns compared to, say, collecting data on social media platforms.

- To that end, *open platforms do exist and are gaining ground.* Platforms can be hosted on a server secured by the operator, such as Mozilla Hubs, while custom-made solutions solve many data-protection problems. We expect exemplar or template systems to emerge in the next few months.

- While *using videoconferencing and screen sharing to assist with remotely operating equipment are attractive, they present new challenges.* In particular, they may be hosted or relayed by servers in different countries and may not be secure. This is one area where institutions may have policies driven by contract agreements with existing providers.

HEALTH CONSIDERATIONS

Safety issues may constrain some types of experiments. For example, while labs are often wide-open spaces, domestic environments used for VR might be small and/or cluttered. Though applications can be coded to fall within the “guardian” space that the user configures for the system, the guardian systems are fallible due to environment change or system glitches. Thus, while games that encourage exaggerated movements are commonplace, we suggest not involving dynamic expansive gestures. Further, we suggest making sure that the experiment operates within a modest amount of space, and if, say, locomotion is important, that this is a key filter in the recruitment of participants.

Another issue, especially if hardware

is shared among participants, is “hardware quarantine.” If a headset is used by only a single person, hygiene may be less of an issue. However, once hardware is shared, obviously hygiene has to be taken into account. Hardware should always be cleaned thoroughly between participants, but extra precautions will need to be taken to limit the possibility of, for example, spreading an infection. Using disinfectant wipes and additional masks that the user can wear underneath an HMD can be valuable, but they offer only limited protection. Recently, decontamination systems have become available that make use of ultraviolet light and nano coatings that offer additional benefits, yet they will not cover every nook and cranny. Current research seems to suggest that contamination on surfaces may cease to exist after 72 hours for Covid-19 [4]. As such, cycling through HMDs that have been put away for some time may be an additional precaution. Currently, a combination of hygienic measures seem most appropriate, and users should always be informed about potential risks (e.g., in informed consent forms).

VALIDITY OF THE RESULTS

Running remote experiments makes keeping a uniform apparatus difficult. In a lab study, typically there is only a single apparatus. Yet remote participants might have varying hardware and, if not explicitly requested by the experimenters, even different headsets. Clearly, controlling for the uniformity of setups helps in isolating other factors that could affect the results. If, in order to reach a sufficiently large number of participants, it becomes necessary to relax the conditions for exclusion (e.g., by allowing users with different headsets to participate), it remains an open question how to consider the validity of results obtained in such a manner.

It can be argued that one of the goals of this type of research is to devise novel methods that can be applied to a wide range of setups, which then could be expected to continue providing comparable performance, as indicated by the empirical results. However, some types of experiments can be too dependent on the specific combination of headset and controllers. Thus, if remote experiments with heterogeneous hardware become an acceptable

platform to run experiments, how far of a divergence in hardware can be accepted? For example, an interaction technique designed to be operated with the now-standard trigger button of a specific VR controller might still work the same way with a controller designed by another maker, even though the ergonomics of the device will differ. Promoting a culture of replicating studies might provide a solution to these challenges.

REMOTE EXPERIMENT DESIGN GUIDELINES

Participants might be recorded completing the experiments over Zoom or other videoconferencing platforms. However, as mentioned earlier, these platforms have associated security risks. Another approach is for researchers to observe participants through a videoconferencing platform, without recording. This would also provide more control of the consistency of the procedure between participants. Because recording a participant and their home adds additional privacy concerns, researchers should weigh the benefits of recording versus observing the participant in real time as they participate in the research study. Labs can still use research assistants to run consistent studies remotely without recording a participant’s home, personal space, and/or unwilling family members in their environment.

Some general remote experiment design recommendations can be found in [5]. Researchers should remove or minimize as many accessibility barriers as possible. This can be achieved by adding feedback systems such as text, voice, and interface prompts. Researchers should also make sure that the language used is accessible to their intended audience. Lacking a live audio or video connection, it might not be possible to further explain instructions after an experiment has started. Reminders can help to ensure that participants complete remote experiments. We also suggest that experimenters take cultural and regional differences into consideration. For instance, for a VR/AR driving simulator designed for left-hand driving countries, user performance and experience might vary in right-hand driving contexts. Similarly, for VR text-entry studies, authors might need to consider the

➤ A RESPONSIVE KIND OF DESIGN

many different keyboard layouts around the world.

The required mental workload of the participant should be reduced to a minimum. This can be achieved by removing setup steps and automating parts of the experiment. Screen or input switching during the experiment should be limited. If possible, all surveys should be displayed within the HMD, or at least in an application that is automatically started on the desktop. Since current text-entry user interface (UI) solutions in VR are not as efficient as keyboards, experimenters might want to choose drop-down menus, sliders, radio buttons, or even voice-recording options to collect survey data in virtual environments. In a perfect setup, the entire experiment would be run from one executable, with consent, instructions, task, and post surveys all completed while wearing the HMD.

Experimenters should collect data on the device used, the specs of that device, the computer used, and the frame rate at which the experiment ran (ideally not just the mean, but also the standard deviation or other meaningful statistical information). The instructions of the experiment could be recorded in advance as a video or an animation in VR and shown to the participants. This could also include any training or context necessary to complete the experiment.

Before launching the experiment, researchers should solicit feedback from their own (and potentially other) labs. Ideally, experiments should be piloted with participants from outside the research group. Such feedback will help solve any setup challenges or other potential sticking points during the experiment and highlight potential safety concerns. This will also allow for a more accurate prediction of experiment completion times, which can then be communicated to participants. During such pilot studies,

any applicable screen-capturing protocol can also be tested.

If the experiment is to be run over a video call, the connection speed for both parties should be tested. This will inform whether it is possible to record the video on the participant's computer or the experimenter's computer, as needed.

Any data logged by the application ideally should then be automatically uploaded over the network to avoid using methods that could de-anonymize data, such as asking participants to send log files to the experimenter via email. For this, the system needs to be able to detect whether or not the upload of the data has happened successfully. Should it have failed, the application would indicate where these files are stored and how to upload them anonymously, for example, via a file-upload web form that does not collect any data besides the log file. This, however, might open the potential of "fake data" being uploaded maliciously. Thus, experimenters should consider solutions to verify whether the data being uploaded is genuine. Experimenters also should consider different end-to-end encryption methods to protect the participants' data.

With this new style of remote experimentation, it also is advisable to ask participants questions about their entire experience of the experiment after completion, to allow for continuous improvement. These questions could include inquiring about their overall satisfaction, levels of immersion, the ease of use of the system, and how intuitive or clear the process was.

While VR and AR HMDs are more popular than ever, they are currently not as widely used as TVs, monitors, or smartphones. As a result, most research on such systems is being conducted in research laboratories within specially designed environments. For instance, tracking base stations may need to be

positioned according to the purpose of the experiment to achieve the needed accuracy. Windowless laboratories may be necessary to avoid incoming sunlight and increase tracking reliability. Meeting such environmental constraints might not be possible in participants' homes, so design decisions may be different in distributed experiments. Experimenters have to consider these differences and design their experiments accordingly.

REMOTE EXPERIMENT OPEN QUESTIONS

- What is the best protocol for the transmission of projects and experimental data?
- How can payment be sent to participants?
- How much of a limiting factor is participants' bandwidth for streaming video and results?
- What are the ethical considerations needed to ensure the privacy of participants' data?
- What is the trade-off between acceptable quality of a recording versus its size?
- Should we expect that participants have space on their computers to record to?
- Is having experimental results streamed to computers outside of a lab a potential ethics issue?
- What is the best way to monitor for simulator sickness when the experimenter is not present?

ADVICE TO REVIEWERS

Regardless of how this new wave of human-subjects experiments is handled, reviewers must be aware of the changing nature of article submissions. In response to the concerns around user studies during Covid-19, many conferences in HCI and VR/AR have sent out calls for participation, highlighting the appropriateness of contributions from systems, design, methodology, literature review, or other contributions focused less on user studies. It falls upon the program chairs to communicate these new criteria down the line to program committee members and reviewers. This change in mindset will be a collective process across all of HCI and AR/VR research. Authors must clearly describe in the submissions the exact way the experiment was administered and also discuss the pool from which the

The required mental workload of the participant should be reduced to a minimum. This can be achieved by removing setup steps and automating parts of the experiment.

applicants were recruited. For example, if an experiment predominantly uses members of VR/AR labs or enthusiasts, this has the potential to distort the outcomes. This is due to existing biases or perceptions coming from frequent exposure to VR/AR systems, relative to participants who have rarely or never experienced VR/AR. Such a biased participant pool thus could represent a limitation for a study. Yet, given Covid-19, these limitations may not invalidate the work being presented, as long as the authors are clear in their description of the participant pool and the reviewers are encouraged to work with the understanding that this is currently one of the very few options for running VR/AR studies. In other words, transparency of the process is one of the best ways that we can usher in this new wave of publications.

CONCLUSION

In situations such as the current pandemic, the use of the short-, medium-, and long-term solutions discussed here enables the fields of HCI and XR to continue to forge forward with experimental work. A secondary benefit of the use of members of other labs in the community is that it increases the amount of transparency in the field by making people more aware of the exact nature of each other's experiments. It could potentially improve the external validity of experiments by increasing the diversity of platforms and participants used for a given task. At the very least, Covid-19 has strengthened this community and inspired new collaborations between researchers. While there are both ethical and practical concerns for distributed user studies, solutions for XR likely will be useful for other areas of HCI and, indeed, any field that relies on human experimentation. This article provides a starting point. We hope other articles will follow with more specific information, either expanding topics presented here or offering new ideas.

We invite people to join the discussion on Slack. The current community originated in a workshop and discussion launched at the online IEEE Virtual Reality conference in March 2020. Email fortega@coloste.edu to be added to the discussion.

ENDNOTES

1. Steed, A., Friston, S., Lopez, M.M., Drummond, J., Pan, Y., and Swapp, D. An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (2016), 1406–1414.
2. Ma, X., Cackett, M., Park, L., Chien, E., and Naaman, M. Web-based VR experiments powered by the crowd. *Proc. of the 2018 World Wide Web Conference*. ACM, 2018, 33–43.
3. van Doremalen, N., Bushmaker, T., Morris, D.H., Holbrook, M.G., Gamble, A., Williamson, B.N. et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England Journal of Medicine* 382, 16 (2020), 1564–1567.
4. Cooper, M. and Ferreira, J.M. Remote laboratories extending access to science and engineering curricular. *IEEE Transactions on Learning Technologies* 2, 4 (2009), 342–353.
5. Nickerson, J.V., Corter, J.E., Esche, S.K., and Chassapis, C. A model for evaluating the effectiveness of remote engineering laboratories and simulations in education. *Computers & Education* 49, 3 (2007), 708–725.

1 Anthony Steed is head of the Virtual Environments and Computer Graphics group in the Department of Computer Science at University College London. He has over 25 years' experience in developing virtual reality and other forms of novel user interface. He received the IEEE VGTC's 2016 Virtual Reality Technical Achievement Award.
→ A.Steed@ucl.ac.uk

1 Francisco R. Ortega is an assistant professor at Colorado State University and director of the Natural User Interaction Lab (NUILAB). His main research area focuses on improving user interaction in 3D user interfaces by eliciting (hand and full-body) gesture and multimodal interactions, developing techniques for multimodal interaction, and developing interactive multimodal recognition systems. His secondary research aims to discover how to increase interest for CS in non-CS entry-level college students via virtual and augmented reality games.
→ f.ortega@colostate.edu

1 Adam Williams is a Ph.D. student in computer science at Colorado State University. His research is on multimodal inputs for augmented reality, specifically, user-elicited gesture and speech interactions. His research goals are to create novice-friendly interactions for 3D learning environments.
→ adamwil@colostate.edu

1 Ernst Kruijff is a professor of human-computer interaction at the Institute of Visual Computing at Bonn-Rhein-Sieg University of

Applied Sciences and adjunct professor at SFU-SIAT in Canada. His research looks at the usage of audio-tactile feedback methods to enhance interaction and perception within the frame of AR view management, VR navigation, and hybrid 2D/3D mobile systems.
→ ernst.kruijff@h-brs.de

1 Wolfgang Stuerzlinger is a professor at the School of Interactive Arts + Technology at Simon Fraser University in Vancouver, Canada. His work aims to gain a deeper understanding of and to find innovative solutions for real-world problems. Current research projects include better 3D interaction techniques for virtual and augmented reality applications, new human-in-the-loop systems for big-data analysis, and the characterization of the effects of technology limitations on human performance.
→ w.s@sfu.ca

1 Anil Ufuk Batmaz has a Ph.D. in biomedical engineering from the University of Strasbourg. He is currently affiliated with Simon Fraser University as a postdoctoral fellow working on human-computer interaction and virtual and augmented reality.
→ abatmaz@sfu.ca

1 Andrea Stevenson Won is an assistant professor in the Department of Communication at Cornell University. She directs the Virtual Embodiment Lab, which focuses on how mediated experiences change people's perceptions, especially in immersive media. Research areas include the therapeutic applications of virtual reality, and how nonverbal behavior as rendered in virtual environments affects collaboration and teamwork.
→ asw248@cornell.edu

1 Evan Suma Rosenberg is an assistant professor in the Department of Computer Science and Engineering at the University of Minnesota. His research interests are situated at the intersection of virtual/augmented reality and HCI, encompassing immersive technologies, 3D user interfaces, and spatial interaction techniques.
→ suma@umn.edu

1 Adalberto L. Simeone is an assistant professor in the Department of Computer Science at KU Leuven in Belgium. His research lies in the intersection of 3D interaction and virtual reality with human-computer interaction. He is motivated by a deep interest in making immersive experiences more accessible by everyone.
→ adalberto.simeone@kuleuven.be

1 Aleshia Hayes is an assistant professor at the University of North Texas. She is passionate about developing, evaluating, and iterating on technology used for learning in formal and informal environments. She runs the SURGE XR Lab, where she has led interdisciplinary research with partners from manufacturing, defense, psychology, and education.
→ aleshia.proff@gmail.com