A. Weisbecker, M. Burmester & A. Schmidt (Hrsg.): Mensch und Computer 2015 Workshopband, Stuttgart: Oldenbourg Wissenschaftsverlag, 2015, S. 69-76.

Virtual Reality Crisis Simulation for Usability Testing of Mobile Apps

Kristian Rother, Inga Karl, Simon Nestler

Forschungsgruppe Mensch-Computer-Interaktion, Hochschule Hamm-Lippstadt

Abstract

Usability testing is expensive in some fields due to the resource requirements that go hand in hand with taking the context of use into account. Crisis related applications are one such field, typically requiring the reenactment of an extensive crisis scenario. To lessen the resource requirements crisis scenarios can be reconstructed as virtual reality simulations. This paper outlines the development of an initial prototype of such a simulation. A refined prototype could eventually allow usability testing of a mobile app in a virtual reality crisis simulation.

1 Problem Identification and Motivation

Crisis management is a complex domain. In complex domains, the context of use has to be taken into account for usability testing (DIS, 2009; Redish, 2007). Consequently, usability testing in the lab is necessary but not sufficient to improve the usability of crisis related interactive systems. Methods that focus on the context of use such as contextual design (Beyer & Holtzblatt, 1999) and field research methods (Kantner et al., 2003; Rosenbaum & Kantner, 2007) are typically conducted during common work processes. These methods by themselves are not suitable for usability tests of crisis related interactive systems because a crisis happens unexpectedly and is not part of the routine work. Even if a crisis would occur while these methods are used they could negatively affect the outcome of the crisis, for example by disturbing domain experts during their tasks. The tests should therefore be stopped if a crisis occurs and be turned into a post-crisis examination.

Due to the outlined problems, field exercises also known as *crisis simulations* (Boin et al., 2004; Kleiboer, 1997) are used for usability testing of crisis related interactive systems (Nestler, 2014). These *real crisis simulations* are resource intensive because they require actors, extras, vehicles, equipment and space (see figure 1). Additionally, changing variables during real crisis simulations, which is often desired for usability testing purposes, is not easy.

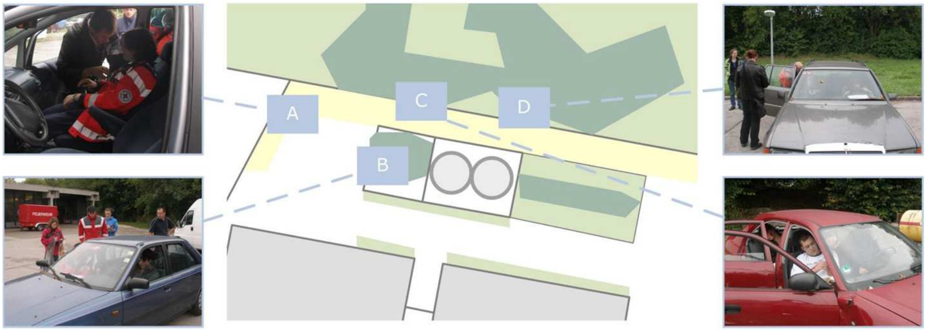

Figure 1: Excerpt of the overview of a real crisis simulation (Nestler, 2014) which shows some of the required vehicles and indicates the space requirements.

Mann et al. (2004) point out, that improvements in disaster management are needed. Improved usability of interactive systems in a crisis context can serve as a step in that direction. If the resource constraints of real crisis simulations could be relaxed, more in depth usability tests could be conducted on the same budget. Ceteris paribus this would lead to an improved usability of the tested interactive system.

2 Definition of the objectives for a solution

To counteract the resource requirements of real crisis simulations these simulations can be transferred into virtual worlds. The resulting simulations are *virtual reality crisis simulations (VRCS)*. The development and use of VRCS is associated with costs. To provide a benefit for usability testing these costs have to be lower than the resources saved by using the VRCS. It is currently assumed as a working hypothesis that this can be achieved. Under this assumption objectives can be identified based on the initial conditions (i.e., are real crisis simulations already used or are they not used at all so far).

1. Real crisis simulations are not used for usability testing
   - Problem: Real crisis simulations are too resource intensive and as a result only lab based usability tests are conducted. The crisis context is not taken into account.
   - *Objective 1*: If some additional resources are available but not sufficient to conduct an entire real crisis simulation they can be used to conduct VRCS and as a result the crisis context is taken into account.
2. Real crisis simulations are used for usability testing
   - Problem: Due to the resource requirements of running an entire real crisis simulation both the number of design solutions that can be tested and the scenarios in which they can be tested are limited.

- *Objective 2*: VRCS can serve as a pre-test to reduce the number of design solutions that have to be tested in the real crisis simulation. VRCS can also be used to pre-configure the real crisis simulation to fit the testing needs.
- *Objective 3*: VRCS replace the real crisis simulations entirely. Due to the reduced resource requirements more scenarios can be tested or scenarios can be tested more in depth and varied easily.

Since real crisis simulations are common practice and accepted as useful as a general tool outside the realm of usability testing (Perry, 2004) objective 3 is excluded as a candidate. The initial goal is to work towards objective 1 or objective 2.

The development of the VRCS and its integration into the usability testing process should be possible and feasible (Peffers et al., 2007). It is possible in principle because virtual reality has been used successfully for other purposes like training in different domains (Orr et al., 2009; Seymour et al., 2002) therapy (Riva, 2005) and way finding (Tang et al., 2009). Furthermore, virtual prototypes (Kuutti et al., 2001) and virtual worlds (Chalil Madathil & Greenstein, 2011) have been suggested as potential tools for usability testing. To ensure that the development of the VRCS is feasible the scope was limited by concentrating on a single crisis scenario and by creating this scenario ad hoc without the direct consultation of domain experts. The selected scenario is a *prolonged power outage* because it is described in literature (Petermann et al., 2014) and the scenario is used in the INTERKOM research project[1]. This ensures access to domain experts for future iterations of the VRCS. The interactive system for which the usability tests will be conducted within the VRCS was limited to a not further specified handheld mobile device that provides helpful information during the ongoing crisis.

3 Design and Development

The developed prototype served as a first test for some ideas and to get a general feeling for the feasibility of creating a VRCS. The two major design decisions were the transformation of the crisis scenario into a VRCS and the representation of the mobile device that will eventually be tested within the VRCS. The simulation was limited to a small city that was constructed from scratch by using preexisting city components such as buildings and streets. The city is in a state of power outage during the entire simulation with limited sound effects where appropriate and additional small visual indications of the power outage such as garbage that wasn't picked up. Obstacles that strategically limit the route to a predetermined one were used which means that one essentially walks from "start to finish" within the city while still retaining a feeling of free movement. With one minor exception, buildings cannot be entered. A playback component for a before-after effect of ten identified events (like in-

[1] See acknowledgements for further details.

creased accidents due to the lack of working traffic lights) was also added. These effects are triggered upon entering certain zones.

The mobile device for which the usability testing will eventually be conducted within the VRCS was prototyped as a very simple virtual tablet (see figure 2).

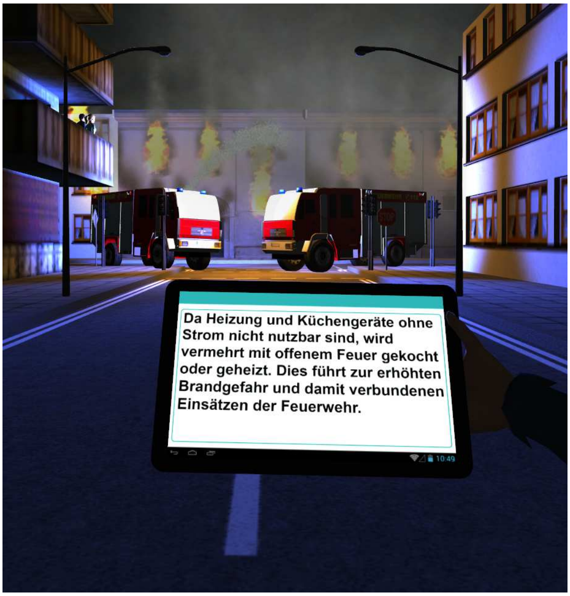

Figure 2: A scene from the VRCS depicting a power outage. The German text on the tablet translates to "Because electric heating and kitchen appliances cannot be used without electricity the use of open fires for cooking and heating increases. This leads to an increased risk of fires and corresponding increases in fire fighter deployments."

The tablet and a virtual hand that holds it show up at the bottom of the simulation when certain trigger points are passed. There is no interaction with the virtual tablet. The device simply shows text related to the ongoing crisis. For example, when the trigger point for an accident is passed, the tablet will show a note that due to the lack of functioning traffic lights, accidents increase significantly during the early stages of a power outage.

The two technology choices for the development of the VRCS were (a) picking a virtual reality technology and (b) picking a 3D-engine. While there are many virtual reality technologies an approach based on the Oculus Rift (OR) and an Xbox controller was selected because the OR and controllers were already available to us and integrated into the teaching process. Furthermore this setup can be used with a laptop, which makes the solution portable. Likewise there are many different 3-D engines. The Unity Engine was selected because it is free, wide spread and already used in other projects at the Hochschule Hamm-Lippstadt.

4 Outlook

This paper outlined the motivation for the development of a VRCS prototype, defined objectives for its construction and sketched a high level view of the design and technology choices. While it proved to be feasible to build the prototype within a reasonable timeframe the construction of the first version of the VRCS has already revealed some defects and ideas for further improvements. Thus the next step is to iteratively improve the VRCS and to develop experiments to test it before eventually moving on to conducting usability tests inside the VRCS. Input from both domain experts in crisis management and human-computer interaction (HCI) specialists is needed and welcome to achieve this. To kick-start this dialogue we list some identified weaknesses and some of our own ideas for further discussion.

Cybersickness: A major drawback is that the well-known problem of cybersickness (Davis et al., 2014; McCauley & Sharkey, 1992) occurred while using the VRCS. The next iteration of the prototype will thus focus on following best practices (Yao et al., 2014) that lead to a reduction in cybersickness. Even if this problem can be reduced it may still have influence on the design choices of future experiments as cybersickness gets worse with prolonged exposure (Kennedy et al., 2000). For example, within-subject designs require a longer exposure to the VRCS than between-subject designs, which could lead to a higher number of subjects dropping out during the experiment due to the experienced sickness.

Presence: An evaluation of presence and how different variations of the VRCS influence the feeling of presence is needed. The methods suggested by Witmer and Singer (1998) as well as Poeschl and Doering (2013) provide instructions.

Relationship between the VRCS and the interactive system that will be tested: The exact specifications of the mobile app that will be tested within the VRCS are currently being developed in the INTERKOM research project. For now a simplified virtual tablet served as a placeholder. However the actual app that will be tested influences the design of the VRCS scenario and as such knowing the exact system to be tested is a precondition for further advances.

Representation of mobile apps in virtual environments: There are multiple possible ways of representing an app in a VRCS. Mirroring the screen of the actual device onto a virtual representation of the device or a recreation of the app within the VRCS are two examples. One of the key problems is that it is hard to impossible to use the mobile device while wearing a head-mounted display (HMD). Even if that wasn’t the case the interaction with the device provides interesting challenges. A transfer of ideas from the use of virtual keyboards or gesture control (Cheney & Ancona, 2014) could prove fruitful. The problem of interacting with the mobile device can be mitigated by moving form a HMD to a CAVE (Cruz-Neira et al., 1993) which is planned in the near future. Even if a CAVE is used a HMD based VRCS can serve as a prototyping environment for the CAVE as long as the underlying technology (e.g. 3D-Engine) is compatible.

Crisis representation: The selected crisis scenario was built ad hoc without the input of domain experts based on the scenario description found in Petermann et al. (2014). Since most test subjects will not have experienced this crisis situation it is hard to measure how realistic the reconstruction actually was. The most obvious approach is to involve domain experts and rely on their feedback or create the crisis scenarios in cooperation with domain experts, which is currently being done for three scenarios in the INTERKOM research project. Other alternatives are seeking subjects that have lived through the specific crisis and relying on their memory of the past experiences (which is limited to types of crises that have already happened) or developing a generic questionnaire to evaluate if the crisis scenario felt real. Additionally content development was an afterthought and mostly based on what was available for free and some intuition regarding the construction of the city. A more rigorous approach following established principles (Isdale et al., 2002) is planned for a future iteration.

Lack of interaction: Currently users can only walk through the city by using a gamepad or keyboard in combination with the direction they look in. The simulation ends when the final destination is reached which leads to a fade to black. While this is acceptable for a first prototype the next steps need to focus on actual actions that are taken during a crisis. These interactions depend on both the app to be tested and the crisis scenario, which have to be developed. The addition of a walking device like the Virtuix Omni as a replacement for the need to walk via a gamepad is planned in the near future.

Usability testing of the VRCS: From an HCI point of view we jumped straight into the development step of the EN ISO 9421-11 process (2009). While this is a good compromise when developing software for your own use or to quickly see how much time it takes to build a prototype, a usability test of the VRCS itself has to be conducted especially if it is to be used by other researchers. Bowman et al. (2002), Gabbard et al. (1999) and Tromp et al. (2003) provide some insight into how this could be done. Sutcliffe and Gault (2004) provide some useful heuristics.

Comparison of VRCS and other methods of generating a crisis context: To test the assumed working hypothesis of resource savings VRCS have to be compared to other methods of creating a crisis context like non-VR 3D simulations, storytelling, paper based descriptions, low-fidelity crisis-simulation and real crisis simulations.

Acknowledgements

This research is supported by a grant from the German Federal Ministry of Education and Research (BMBF) as part of the INTERKOM project in the context of the program „Forschung für die zivile Sicherheit II“ (01.01.2014 – 31.12.2016, Förderkennzeichen 13N1005). We would like to thank Alexander Giesbrecht, Sandra Jürgensmeier, Inge Kling and Sinan Mert for their contributions.

Contact information

Kristian Rother	Inga Karl	Simon Nestler
kristian.rother@hshl.de	inga.karl@hshl.de	simon.nestler@hshl.de

Hochschule Hamm-Lippstadt, Marker Allee 76, 59063 Hamm, Germany

References

Beyer, H., & Holtzblatt, K. (1999). Contextual Design. *Interactions*, *6*(1), 32–42. http://doi.org/10.1145/291224.291229

Boin, A., Kofman-Bos, C., & Overdijk, W. (2004). Crisis simulations: Exploring tomorrow's vulnerabilities and threats. *Simulation & Gaming*, *35*(3), 378–393.

Bowman, D. A., Gabbard, J. L., & Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments*, *11*(4), 404–424. http://doi.org/10.1162/105474602760204309

Chalil Madathil, K., & Greenstein, J. S. (2011). Synchronous remote usability testing: a new approach facilitated by virtual worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2225–2234). ACM.

Cheney, J., & Ancona, D. (2014). Gesture Controlled Virtual Reality Desktop.

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (pp. 135–142). ACM.

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). A Systematic Review of Cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment* (pp. 8:1–8:9). New York, NY, USA: ACM. http://doi.org/10.1145/2677758.2677780

DIS, I. (2009). 9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. *International Standardization Organization (ISO). Switzerland*.

Gabbard, J. L., Hix, D., & Swan, J. E. (1999). User-centered design and evaluation of virtual environments. *Computer Graphics and Applications, IEEE*, *19*(6), 51–59.

Isdale, J., Fencott, C., Heim, M., & Daly, L. (2002). Content design for virtual environments. *Handbook of Virtual Environments: Design, Implementation, and Applications*, 519–532.

Kantner, L., Sova, D. H., & Rosenbaum, S. (2003). Alternative methods for field usability research. In *Proceedings of the 21st annual international conference on Documentation* (pp. 68–72). ACM.

Kennedy, R. S., Stanney, K. M., & Dunlap, W. P. (2000). Duration and exposure to virtual environments: Sickness curves during and across sessions. *Presence: Teleoperators and Virtual Environments*, *9*(5), 463–472.

Kleiboer, M. (1997). Simulation methodology for crisis management support. *Journal of Contingencies and Crisis Management*, *5*(4), 198–206.

Kuutti, K., Battarbee, K., Sade, S., Mattelmaki, T., Keinonen, T., Teirikko, T., & Tornberg, A.-M. (2001). Virtual prototypes in usability testing. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 2001* (p. 7 pp.–). http://doi.org/10.1109/HICSS.2001.926545

Mann, N. C., MacKenzie, E., & Anderson, C. (2004). Public health preparedness for mass-casualty events: a 2002 state-by-state assessment. *Prehospital and Disaster Medicine*, *19*(03), 245–255.

McCauley, M. E., & Sharkey, T. J. (1992). Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoperators and Virtual Environments*, *1*(3), 311–318.

Nestler, S. (2014). Evaluation der Mensch-Computer-Interaktion in Krisenszenarien / Evaluating human-computer-interaction in crisis scenarios. *I-Com*, *13*(1), 53–62. http://doi.org/10.1515/icom-2014-0008

Orr, T. J., Mallet, L. G., & Margolis, K. A. (2009). Enhanced fire escape training for mine workers using virtual reality simulation. *Mining Engineering*, *61*(11), 41.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77.

Perry, R. W. (2004). Disaster Exercise Outcomes for Professional Emergency Personnel and Citizen Volunteers. *Journal of Contingencies and Crisis Management*, *12*(2), 64–75. http://doi.org/10.1111/j.0966-0879.2004.00436.x

Petermann, T., Bradke, H., Lüllmann, A., Poetzsch, M., & Riehm, U. (2014). *What Happens During a Blackout: Consequences of a Prolonged and Wide-ranging Power Outage*. BoD–Books on Demand.

Poeschl, S., & Doering, N. (2013). The German VR Simulation Realism Scale–psychometric construction for virtual reality applications with virtual humans. *Stud Health Technol Inform*, *191*, 33–37.

Redish, J. (2007). Expanding usability testing to evaluate complex systems. *Journal of Usability Studies*, *2*(3), 102–111.

Riva, G. (2005). Virtual reality in psychotherapy: review. *Cyberpsychology & Behavior*, *8*(3), 220–230.

Rosenbaum, S., & Kantner, L. (2007). Field usability testing: method, not compromise. In *Professional Communication Conference*.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O’Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual Reality Training Improves Operating Room Performance. *Annals of Surgery*, *236*(4), 458–464.

Sutcliffe, A., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. *Interacting with Computers*, *16*(4), 831–849. http://doi.org/10.1016/j.intcom.2004.05.001

Tang, C.-H., Wu, W.-T., & Lin, C.-Y. (2009). Using virtual reality to determine how emergency signs facilitate way-finding. *Applied Ergonomics*, *40*(4), 722–730.

Tromp, J. G., Steed, A., & Wilson, J. R. (2003). Systematic usability evaluation and design issues for collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, *12*(3), 241–267.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240.

Yao, R., Heath, T., Davies, A., Forsyth, T., Mitchell, N., & Hoberman, P. (2014). Oculus VR Best Practices Guide. *Oculus VR*.