Wizard of Oz experiment for Prototyping Multimodal Interfaces in Virtual Reality

Blaž Gombač
blaz.gombac@student.upr.si

Matej Zemljak
matej.zemljak@student.upr.si

Patrik Širol
parik.sirol@student.upr.si

Damir Deželjin
ddezeljin@student.upr.si

Klen Čopič Pucihar
klen.copic@famnit.upr.si

Matjaž Kljun
matjaz.kljun@upr.si

University of Primorska, FAMNIT
Glagoljaška 8
Koper, Slovenia

ABSTRACT
In recent years the field of virtual reality has witnessed a rapid growth with significant investments in both hardware and software development. It has several potential applications for entertainment, education and enterprise where users benefit from being immersed into virtual worlds. VR headsets are available in several forms and price ranges from simple and inexpensive Google Cardboard to more complex products such as Oculus Rift. Nevertheless, designing fully operational virtual reality applications for researching new complex multimodal interaction possibilities (e.g. mid-air gesture, voice, haptics, etc.) may be difficult to implement, costly and time consuming. For this reason we have looked into ways of rapidly prototyping virtual reality interactions. Our approach consists of the Wizard of Oz experiment in which subjects interact with a computer system believing to be autonomous, but is in reality operated by researchers. The presented system allows non-technical designers to explore various multimodal interactions with rapid prototyping of VR environments.

Categories and Subject Descriptors
H.5.2 [**Information interfaces and presentation**]: Multimedia Information Systems—*Prototyping*

Keywords
Wizard of Oz, virtual reality, rapid prototyping, multimodal interaction

1. INTRODUCTION
The majority of big computer companies recently identified a big potential in Virtual and Augmented Reality (VR, AR) technology. This has lead to massive investments in hardware and software development, such as, Facebook takeover of Oculus Rift[1], Google's investment in MagicLeap[2] and Expeditions[3], Samsung's development of Galaxy Gear[4], and Microsoft's development of HoloLens[5].

Virtual reality offers immersion into virtual environments capable of producing a stereoscopic view into a virtual world that is usually coupled with audio. The stereo image is delivered by a head-mounted display (HMD) with sensors that track users' movements allowing the system to change the view accordingly. There are two types of HMDs: (i) the fully featured HMDs designed for use with gaming consoles or PCs and (ii) composite HMDs designed to hold a smart phone or a tablet computer. Fully featured devices are expensive and can cost between a couple of hundred to a couple of thousand euros excluding the cost of a console or PC. While in composite HMDs the mobile device (commonly accessible among the population) acts as a display and processing unit, which reduces the cost of HDMs bellow hundred euros.

Both types of HMDs offer various VR experiences with a varying degree of immersion. The latter partly depends also on the quality of the VR environment being projected on the screen and partly on other data produced for other senses. However the illusion most often remains incomplete, in that not all senses are catered for and natural ways of interacting in real world, such as with spoken and body language, are not supported. The work presented explores ways of supporting non-developeres to explore various multimodal interactions (including for example mid-air hand gestures, voice, haptics) in rapidly prototyped VR environments. For this purpose we designed and built a VR test-bead based on the Wizard of Oz (WoZ) metaphor. The test-bead enables screen sharing between desktop computer and HMD where the researcher acts as a wizard detecting and executing users' commands (e.g. hand gestures) on a desktop computer creating the illusion of a working prototype. In order to evaluate the test-bead the paper presents a short user study which was carried out using our fast prototyping

[1]https://www.oculus.com/
[2]https://www.magicleap.com/
[3]https://www.google.si/edu/expeditions/
[4]http://www.samsung.com/global/galaxy/gear-vr/
[5]https://www.microsoft.com/microsoft-hololens/en-us

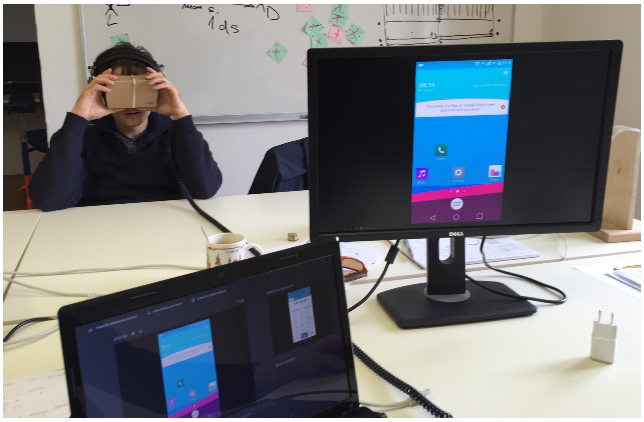

Figure 1: ***The conduct of the experiment. The experimenter controls the stream to the HDM based on participant's mid-air hand gestures or voice controls.***

technique.

2. LITERATURE REVIEW

The Wizard of Oz (WoZ) experiments in human-computer interaction have a long tradition and were pioneered in 1983 by Gould et. al. who simulated an imperfect listening typewriter to explore users' reactions when giving dictations [4]. The method was used in numerous studies since. It was for example used in prototyping speech user interfaces when AI agents were not so capable [5] or for studying a mixed reality application for enriching the exploration of a historical site with computer generated content [2]

WoZ is nowadays commonly used for rapid prototyping of systems that are costly to build, or as means of exploring what people require or expect from systems that require novel or non existent technology [8]. However, Maulsby et. al. have warned that researchers need to understand what limitations need to be implemented on the Wizard's intelligence, and need to base behavior and dialog capabilities on formal models in order to ensure consistent interaction, keep simulation honest, and to prevent inappropriate, optimistic results [7]. Nevertheless, as demonstrated by numerous studies employing WoZ, the observation of users' using such systems can lead to qualitative data that could not be otherwise acquired.

Furthermore, the HCI community has pointed out that there is a great need for easy to use, rapid prototyping tools (such as WoZ) [6, 3] and that any medium (including VR and AR) can reach its potential only when put into the hands of designers who can further develop it and define its popular forms. Such tools have already been developed to research AR usability and interactions [1]. Our contribution to existing work is providing an affordable, easy to use and intuitive set of tools and procedures for rapid prototyping user interfaces in VR. We evaluate the prototyping tool by running a small user study comparing voice and mid-air gesture interface while wearing HDM.

3. PROTOTYPE DESCRIPTION

There are three main hardware components used in our prototype: Android based smartphone, Google Cardboard, and a Windows based computer. The user interface is then streamed in real time to the phone from a desktop computer using TrinusVR[6] application as seen in Figure 2. Depending on the configuration, either the full screen or only the active window is streamed to the HMD device. The application was designed to transform any Android or iOS device into an affordable HMD to be used by gamers when playing 3D games on their computers. The application also features a lens correction system aimed to improve user experience by minimising distortion inducted by Google Cardboard's lenses. The communication between desktop computer and used mobile devices works both via USB cable or WiFi. The later is particularly interesting as it enables researcher to place the wizard into another room observing users via web cam and executing users commands.

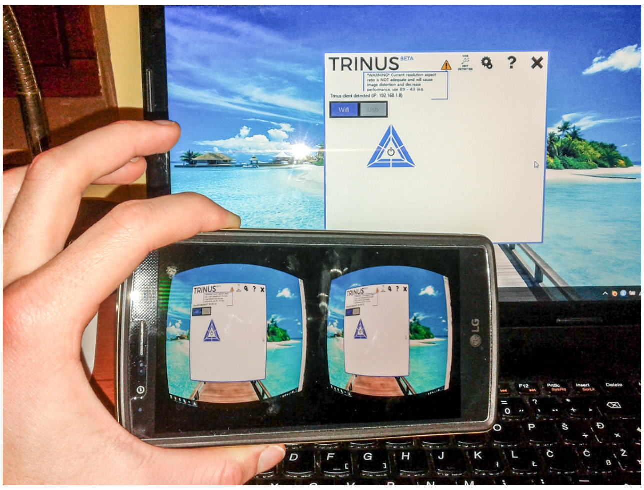

Figure 2: ***TrinusVR streaming the computer desktop to a mobile phone to be used in Google Cardboard.***

4. METHOD

To test our test-bed we have designed an interaction scenario comparing two different interaction techniques: namely mid-air finger gesture and voice based interaction. For this purpose we have created minimum viable product – two simple linear presentations in a presentation program. Each slide of the presentation featured a screenshot of the user Interface (UI) for a particular step towards completing a task. Users performed generic phone tasks such as initiate a call, take a picture, browse files. In order for the linear presentation to work in each step participants had only one possible option to choose from. In Figure 3 both gesture based (left) and voice based (right) user interfaces are displayed. In gesture based UI users had to bend the appropriate finger to trigger one of the available actions (e.g. bending pointer finger opened a folder named "Camera" as seen on the left in Figure 3) while in voice based UI users had to name available options visible on the screen (e.g. reading aloud the name of the folder "Camera" framed in red (right in Figure 3) opened it). After users initiated an action, the exper-

[6]http://trinusvr.com/

imenter switched to the next slide in the presentation in order to show the next screen on the HDM.

One of the issues we had to deal with is how to provide instructions for mid-air gesture interaction. The provision of gesture controls is almost indispensable at the beginning until users get familiar with interaction. This is also the case with current HDM controllers that come in sets of two (one for each hand) and each has several buttons and ways to control the VR worlds and tools. Until one gets accustomed to controls in a particular VR environment, the instructions can be overlaid over the controllers. In our study all available options were always visible on the screen. The limitation of our mid-air finger gesture set is bending five fingers only, which limited us to have five options only in each step. However, as w had a linear presentation with two options at most (back and forward) this was enough for our study. Users have also not had any troubles using the system and did not find instructions intrusive.

While mid-air (hand, finger) gesture interfaces are not so common (yet) on mobile devices, voice recognition and intelligent agents or personal assistants (such as Siri, Cortana and Alexa[7]) are a part of all major mobile operating systems and many users are accustomed to use them to complete certain tasks or services. Exploring natural language interactions thus did not present the same issues as mid-air gesture interactions. In our scenario users just had to read aloud the text on the screen or use controlling words such as “up”, “down”, “left”, “right”, “select”, etc.

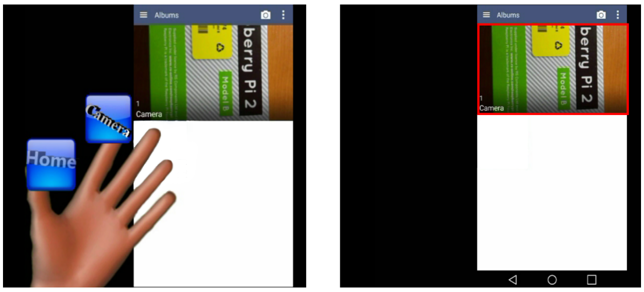

Figure 3: ***A sample screen from the scenario. Left: a mid-air finger gesture based interface where available options are visible over fingers and bending a finger with available option triggered the appropriate action. Right: a voice based interface where available options are framed in red and reading aloud their names triggered the appropriate action.***

We have used a convenience and snowball sampling to recruit participants. We have recruited 10 participants. The average age was 22.3 (SD = 3). All participants were either students (8 participants) or faculty members (2 participants) from our department. Each participant has tried both voice and gesture based interactions where the order was randomised. Before commencing the designed scenario participants tested each interaction mode in order to make sure they understood the principles of how to control each navigation. After completing the scenario of each interaction technique (see Figure 1) users had to answer SUS questionnaire.

5. RESULTS AND DISCUSSION

As mentioned, our scenario was a minimum viable study to test our test-bed. It involved handling the Android operating system UI by presenting users with sequence of images including browsing photos and controlling music player. We have thus not used a 3D virtual world, which is a limitation of our evaluation and which makes it difficult to generalize the results in context of virtual world interactions. Due to virtual representation of mobile device in our study, it is also not possible to generalize results in the context of mobile phone interaction. Nonetheless, the pilot provides practical insights into how the designed test-bed could be effectively used as a rapid prototyping tool for exploring different interaction possibilities within VR environments.

Since anything can be streamed from a desktop computer to the HDM, designers and non-technologists can use any available software to create such interactive environments. For example, navigating a 3D information environment can be simulated in non-linear and zooming presentation software such as Prezi[8], or 3D worlds could be simulated by creating them in computer-aided design (e.g. AutoCAD) or 3D computer graphics software (e.g. Blender[9]).

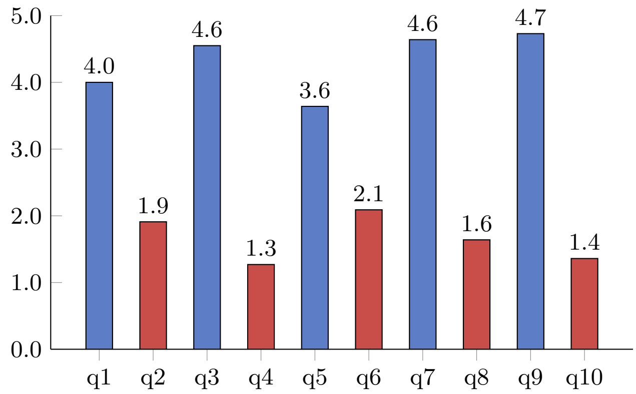

Figure 4: ***SUS scores by question for the gesture controlled interaction.***

The results of SUS questions[10] from our questionnaire for each interaction technique are visible in Figures 4 and 5. Even questions (colored in blue on the graphs) are about positive aspects of the system, while odd questions (colored in red on the graphs) regard negative aspects. We can see that in both cases negative aspects scored low while positive aspects scored high. Average SUS score for mid-air gesture interaction was 83.18 (SD 13.04), whilst voice interaction scored 81.46 (SD 8.08). Both scored above the threshold where users are more likely to recommend the product to friends. However, since this was just a minimum viable study we can just say that users were intrigued with how a phone’s UI can be interacted with and SUS scores are provided for informative purposes only.

[7] https://en.wikipedia.org/wiki/Intelligent_personal_assistant

[8] https://prezi.com/

[9] https://www.blender.org/

[10] See http://www.measuringu.com/sus.php

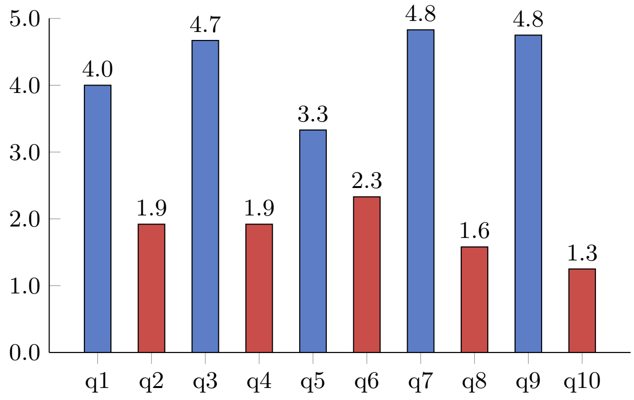

Figure 5: ***SUS scores by question for the voice controlled interaction.***

Despite the fact, that we have not used a virtual world in our study, we have tested the prototype as a test-bed for VR interaction with a minimum viable product. We believe that our approach can open novel possibilities to explore, further develop and define popular forms of such medium since there is no need for designers to know any programming language except how to use designing software, which they should be familiar with already.

6. CONCLUSION

Mid-air gesture and voice interaction provide for richer experience than touch screen user interfaces (UI). Especially in virtual and augmented environments, where interaction with common paradigms (e.g. mouse + keyboard or touch screen) is challenging or inadequate. This introduced a need for new interaction metaphors designed particularly for these new worlds. One such example are mid-air gesture and voice interaction which can facilitate greater immersion into virtual environments. While there are fairly inexpensive depth camera and gesture sensors available for end users, programming for these can be challenging and time consuming particularly for non-technical people such as designers limiting their ability to contribute, further develop and define popular forms of such medium.

In this paper we present an affordable easy to use rapid prototyping tool for creating VR environment and explore different interactions with the Wizard of Oz (WoZ) experiment. With the introduction of wizard we remove the need for additional hardware setup such as wired gloves, depth aware or stereo cameras, gesture based controllers, etc. Experimenters can use any software designers are familiar with to create VR worlds, such as standard non-linear presentation, CAD or 3D graphics software. Or can simply create a sequence of UI screens that users can navigate through with interactions beyond mouse and keyboard. In the future we plan to further evaluate the test-bed (i) by running a user study in a 3D VR environment involving more participants, (ii) including other metrics such as timing tasks, interviews, coding transcriptions of filmed evaluations, etc. and (iii) by placing the wizard into a separate room creating a more convincing illusion of a working system.

7. REFERENCES

[1] G. Alce, K. Hermodsson, and M. Wallergård. Wozard: A wizard of oz tool for mobile ar. In *Proceedings of the MobileHCI '13*, MobileHCI '13, pages 600–605, New York, NY, USA, 2013. ACM.

[2] S. Dow, J. Lee, C. Oezbek, B. MacIntyre, J. D. Bolter, and M. Gandy. Wizard of oz interfaces for mixed reality applications. In *CHI '05 Extended Abstracts*, CHI EA '05, pages 1339–1342, New York, NY, USA, 2005. ACM.

[3] M. Gandy and B. MacIntyre. Designer's augmented reality toolkit, ten years later: Implications for new media authoring tools. In *Proceedings of the UIST '14*, UIST '14, pages 627–636, New York, NY, USA, 2014. ACM.

[4] J. D. Gould, J. Conti, and T. Hovanyecz. Composing letters with a simulated listening typewriter. *Communications of the ACM*, 26(4):295–308, apr 1983.

[5] S. R. Klemmer, A. K. Sinha, J. Chen, J. A. Landay, N. Aboobaker, and A. Wang. Suede: A wizard of oz prototyping tool for speech user interfaces. In *Proceedings of the UIST '00*, UIST '00, pages 1–10, New York, NY, USA, 2000. ACM.

[6] B. MacIntyre, M. Gandy, S. Dow, and J. D. Bolter. Dart: A toolkit for rapid design exploration of augmented reality experiences. In *Proceedings of the UIST' 04*, UIST '04, pages 197–206, New York, NY, USA, 2004. ACM.

[7] D. Maulsby, S. Greenberg, and R. Mander. Prototyping an intelligent agent through Wizard of Oz. In *Proceedings of the CHI '93 and INTERACT '93*, pages 277–284. ACM, 1993.

[8] J. Wilson and D. Rosenberg. Rapid prototyping for user interface design. In M. Helander, editor, *Handbook of Human-Computer Interaction*, pages 859–875. Elsevier Science Publishers, 1988.