Augmenting Remote Interviews through Virtual Experience Prototypes

Almohannad Albastaki
Design Lab, School of Architecture, Design and Planning
The University of Sydney
Sydney, Australia
almohannadk@aol.com

Marius Hoggenmueller
Design Lab, School of Architecture, Design and Planning
The University of Sydney
Sydney, Australia
marius.hoggenmueller@sydney.edu.au

Frederic Anthony Robinson
Creative Robotics Lab
University of New South Wales
Sydney, Australia
frederic.robinson@unsw.edu.au

Luke Hespanhol
School of Architecture, Design and Planning
The University of Sydney
Sydney, Australia
luke.hespanhol@sydney.edu.au

Figure 1: Qualitative research on urban robot, Woodie, in-the-wild *and* online: researcher interviewing people at a public festival (left), and remote interview facilitated through a virtual experience prototype (right).

ABSTRACT

This paper presents insights from a pilot study in which we used a Virtual Experience Prototype (VEP) to gather qualitative feedback in an online evaluation context. For our study, we created a VEP consisting of a non-immersive virtual reality simulation of an urban robot to inform the design of robotic expressions. Through the means of video conferencing software, we were able to collect qualitative data through the think-aloud protocol while participants interacted with the VEP, followed by subsequent in-depth interviews. By comparing our data to findings from previous in-the-wild deployments, we report on aspects which were comparable to in-person evaluations. Reflecting on our approach, we present a preliminary list of lessons learnt and examine how VEPs can support researchers and practitioners to gain rich feedback from participants in synchronous remote user testing. While bearing in mind the limitations over physical prototypes, we argue that VEPs can be used as a lightweight tool to engage participants in remote interviews through interactive spatial experiences.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
OzCHI '20, December 2–4, 2020, Sydney, NSW, Australia
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8975-4/20/12…$15.00
https://doi.org/10.1145/3441000.3441057

CCS CONCEPTS

• **Human-centered computing** → **Systems and tools for interaction design**; **Human computer interaction (HCI)**.

KEYWORDS

Experience Prototypes, Virtual Reality, Urban Robots, Human-Robot Interaction, Remote Qualitative Research

ACM Reference Format:
Almohannad Albastaki, Marius Hoggenmueller, Frederic Anthony Robinson, and Luke Hespanhol. 2020. Augmenting Remote Interviews through Virtual Experience Prototypes. In *32ND AUSTRALIAN CONFERENCE ON HUMAN-COMPUTER INTERACTION (OzCHI '20), December 2–4, 2020, Sydney, NSW, Australia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3441000.3441057

1 INTRODUCTION

Remote user research has become an increasingly popular mode of inquiry to obtain feedback from users at various stages in the design process [13, 27, 39]. Compared to in-person lab and field studies, remote studies provide several advantages, such as access to participants in various locations and time zones, as well savings of time and costs. With the establishment of crowdsourcing services such as Amazon MTurk, remote studies have become a widely established research tool in academia and industry. Most of these studies are conducted in the form of unmoderated online surveys, where quantitative data is collected, for example, through likert scales, or qualitative feedback through text input. For evaluating specific aspects of a design, these surveys can be augmented through videos [4], click-though mock-ups [2] or interactive simulations of the prototype in question [44]. While this is now a common approach for experimental studies in human-computer interaction (HCI), less is known about the remote evaluation of prototypes following a qualitative descriptive research design, for example through observations or semi-structured interviews. However, this type of study is important to gain rich contextual insights about how users perceive and interact with a prototype in a certain situation, and it is also applicable to new and emerging areas of research where coherent theories might not exist yet [3].

At the same time, with the availability of easily accessible game development platforms, such as Unity[1], computer-generated simulations have become a popular method among HCI-researchers and practitioners for the design and evaluation of early design concepts. These simulations can be deployed and experienced across various platforms, for example through an immersive virtual reality (VR) headset [33] or on a conventional computer screen [21]. They provide several benefits in terms of technical effort and skills required to build prototypes [33], as well as cost reduction and robustness [29, 44]. By incorporating interactive components and some level of functionality, designers and prospective users can actively engage with an envisioned future product, system, or mediated environment. Buchenau and Suri refer to this form of early concept explorations as *experience prototyping* [8], whereby the prototype does not necessarily take on a physical form, but can be deployed in any medium, including VR, as long it supports understanding, exploring and communicating the envisioned experience.

In this paper, we build on the notion of experience prototyping and investigate how this form of early concept explorations can be carried out in remote design exploration sessions. We therefore propose the use of Virtual Experience Prototypes (VEPs), which we define as a specific instance of an experience prototype: a lightweight application that is designed to enable rapid prototyping and remote evaluation of an interactive experience in a non-immersive VR environment. As VEPs mimic the general appearance and functions of a prototype in the real world, we argue that they can be of value to reveal insights on how users may perceive the envisioned product, even if they are not able to interact with it in a real-world setting. To illustrate our approach, we present preliminary insights from a synchronous remote user study [39]; we observed and interviewed participants while they were interacting with an urban robot named Woodie, manifested through a VEP. Our analysis indicates that some of our findings match those from previous studies that we conducted on Woodie [19, 20], as well as similar HRI studies [10, 26, 42, 43, 49], which were executed in a real-world context: for example, how the robot was perceived, as well as how participants intended to interact with the robot. Based on our approach, we present a preliminary list of lessons learnt in designing and employing Virtual Experience Prototypes to collect qualitative data in remote evaluations. Further, we discuss challenges in interpreting feedback on non-immersive virtual experiences, which risk missing out on aspects such as collective sense-making or tangible interactions, when compared to their real-world counterparts. We believe our findings can be useful to researchers in the field of Human-Robot Interaction (HRI) and HCI who wish to employ lightweight remote prototyping tools in the design and evaluation of robotic applications and interactive spatial experiences.

2 RELATED WORK

Our work on VEPs builds on and combines experience prototyping [8] and remote user testing [39], two methods which have been extensively used by the HCI community to evaluate early design concepts. We believe that the use of non-immersive virtual environments may accommodate for the combination of these two methods. Therefore, we build on insights from previous studies conducted with virtual environments to inform our own.

We present an overview of studies which have used VR for evaluating HCI-systems, many of which were conducted in a lab environment and employed quantitative data collection techniques. As our research context was concerned with an urban robot we also refer to findings from HRI evaluation studies using non-physical prototype representations, such as video and VR (see Figure 2).

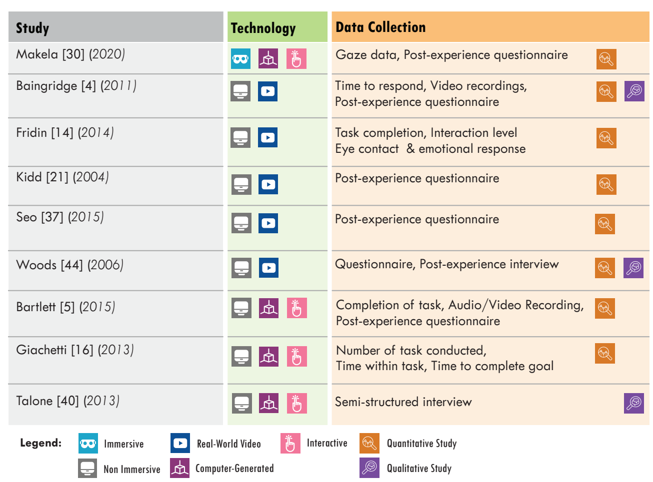

Figure 2: Overview of related VEP studies in HCI.

2.1 Experience Prototyping

The creation of prototypes is an essential activity in the design process of interactive systems [9]. Thereby, prototypes can fulfil various purposes: for example, prototypes can be used to evaluate multiple design concepts and select the most promising before moving on to the next development stage. Further, prototypes can be used as a communication tool to help other designers, clients

[1]https://unity.com

or potential users to understand a design concept that is manifested through the prototype, thereby also acting as an enabler to generate new design ideas [8, 30]. Buchenau and Suri introduced the concept of experience prototyping [8], emphasising the active engagement of participants to convey the experience when interacting with an artefact. Further, they highlight the importance of considering contextual factors, such as social circumstances and environment. Experience prototyping has been widely applied in the HCI-community: for example, Henderson et al. described the use of rapid modular prototypes to augment user interviews in a study on parking meters [17]. Lim et al. emphasised the importance that designers should carefully consider the manifested form of prototypes (e.g. material, level of detail), which in turn filters the design qualities of interest [30].

2.2 VR for User Testing

For the purpose of this review, we will put VR (using computer-generated representations) into two broad categories. The first, *immersive*, implies a certain level of immersion, for example through the use of head mounted displays (HMD) with associated controllers [33], or a four-sided Cave Automatic Virtual Environment (CAVE) [40] where the user's bodily movements are translated into the virtual world. The other category is *non-immersive*, which comprises settings where users experience the virtual world through a computer screen or tablet. Movement and rotation is done with a keyboard and mouse or through rotation of the device [36].

Immersive VR provides researchers with the opportunity to evaluate prototypes without the need of physically building and deploying them into the real-world. Thereby, it offers a high level of interactivity, and can immerse participants into the environment and context the prototype is situated in [33]. Studies have investigated the efficacy of immersive VR in various contexts, such as evaluating pedestrian safety [12], locomotion behaviour [1], wayfinding [40], and interactions with public displays [33]. All of those studies emphasised the higher ecological validity of immersive VR in comparison to controlled lab studies, while mentioning that it is still weaker than field studies conducted in the wild. However, with HMDs just entering the consumer market and CAVE systems being expensive to build, immersive VR testing is still often conducted in lab environments.

On the other hand, non-immersive VR allows researchers to evaluate their prototypes without the need for specialised hardware equipment, thus increasing the feasibility of remote user testing, where participants are not required to come to a lab. A study by Madathil & Greenstein [32] evaluated the efficacy of remote collaborative non-immersive VR for usability testing by comparing it to a controlled lab and video conferencing contexts. The results showed little difference between the three cases except in terms of ease of use and mental load in part of the participant cohort, with controlled lab cases scoring best.

Hollaender et al. [21] used a non-immersive virtual environment in a remote study to collect pedestrian behaviour data when virtually crossing a road. They reported that the virtual environment evoked similar behaviour to that observed in the real world, thus implying that their method enables collection of pedestrian behaviour data at large-scale and not posing any danger to study participants.

2.3 Evaluating HRI through Video and VR

Although conducted in the same environment as the researcher, a number of studies have investigated interactions with humanoid robots through video-based playback in comparison to real robots. Researchers evaluated engagement and perception [23, 47], empathy presented by the participant when something bad happened [41], response of preschool children when instructed to perform motor tasks [15], and judgement and trust when presented with instructions [4]. Although there was a lack of interaction, the studies suggested that physical and video-based robots were treated the same in most cases. However, the physical presence of the robot played a role in social interaction and participants were more responsive in these cases [4, 15, 41].

Other virtual robot studies investigated non-humanoid utilitarian robots for military operations [16, 44] or urban search and rescue [5, 29, 46], utilising interactive non-immersive VR. The key benefit is that virtual robots are better suited for repeated trials. This makes it easier to test various HRI behaviour beyond the general functioning of the robot [29, 44]. Further, physical robots may have some level of unreliability due to issues with power supply or misaligned sensors, which can cause physical testing to be cumbersome and failure is therefore more detrimental to the study in a physical set up [44]. Additionally, virtual robots cost less to build, reduce data collection time and effort, and are easier to modify, thereby allowing a wide range of research objectives to be addressed [44].

To sum up: Experience prototyping is a common approach in HCI and there is a wealth of examples on how to create physical prototypes (e.g. using tinkering platforms) for evaluating envisioned products. Furthermore, previous studies have shown that VR can be also a viable technology for collecting user data. While remote testing provides several advantages and is widely used for the collection of quantitative data, little is known about the use of VR to create virtual experience prototypes for the collection of qualitative data remotely. In our study we address this gap and present preliminary findings on evaluating an urban robot through a non-immersive VR environment in an online context.

3 METHODOLOGY

3.1 Research Context

As non-humanoid robots are expected to become pervasive in the near future - also in urban environments - researchers have began to investigate how to increase their social acceptance, for example through the ability to express emotional states [10, 26]. Empirical studies have proved the effectiveness of emotional expressions through various modalities, such as light [45], sound [24, 48], movement [10] or varying combinations of each [19, 42, 49]. However, most of these studies have been carried out in a controlled lab environment. Our recent in-the-wild study [19] conducted with an urban robot, named Woodie, on the contrary, showed that contextual aspects can influence the perception of the robot and its emotional expressions. This makes it more difficult for people to interpret the emotional expressions compared to their lab counterparts. Further, the study illustrated a need for coherent integration of emotional expression with a robot's function, as emotional expressions are not intuitive when presented in-the-wild [19]. Consequently, a robot's behaviour benefits from being coherent with its shape and function. The here presented research-through-design study [50] builds on this previous findings. By using the same robot, we sought to address these challenges by investigating a method for encoding functional and emotional expressions. Through the modalities of light and sound, the functional states we communicated were *move*, *draw* and *rest*, while the emotional states were *excited* and *tired*. An existing framework that proposed the communication of status, intent and awareness for self-moving automated systems guided our design considerations [35]. Our project progressed through the creation of a framework for designing robotic expressions based on existing literature, the iterative development of a VEP to test these expressions and, finally, collecting and analysing data from an expert user study.

3.2 Prototype Design and Development Process

Following Buchenau and Suri's definition of experience prototypes [8], we created a VEP that users can interact with through the Unity 3D game engine. We designed a setting that mimicked the robot's first deployment [20]: an alleyway at night time with a few neon lights scattered throughout the scene. We created a 3D-model of the robot, thereby retaining the same proportions as the physical counterpart. The robot was equipped with 64 individual lighting components to present the designed expressions (see Figure 1, right), thus simulating the interface qualities of the physical robot.

We used Processing [2] to create the visual expressions. At first, the output from Processing was sent to Unity using an UDP transmission/receiver script (see Figure 3, left) to allow for quick prototyping iterations. After confirming a final set of visual cues for the user study, we converted them into a set of videos. The videos were then imported into the Unity project, and a separate script was developed to extract the pixel information from the videos and map them onto the LED materials of the 3D-modelled robot[3]. To create the audio expressions, we used a Vocoder in Ableton Live[4]. Each file was uploaded to Unity and triggered when the visuals changed, signifying a state switch. We started with a number of audio expressions per state, later reducing it to one per state. Figure 3 (right) presents an overview of the second iteration of the prototype. The model was completely automated. We randomised movement to create the impression that the prototype acted autonomously. State switches occurred every 10 seconds. When users moved in front of the robot, it would stop, express awareness of the user and change direction after some time. For user movement, we opted for a first-person perspective to mimic a real world perspective. Users could use the keyboard buttons "WASD" to move around, and change their gaze direction using the mouse. Once we concluded the software development, we distributed the VEP through a standalone software application, which participants could download and run on their own computers for the purpose of the remote user study.

[2]https://processing.org/
[3]Follow the link to access the Gitlab repository containing all the Unity code for the VEP: *https://gitlab.com/almok/remote-interview-through-veps*
[4]https://www.ableton.com

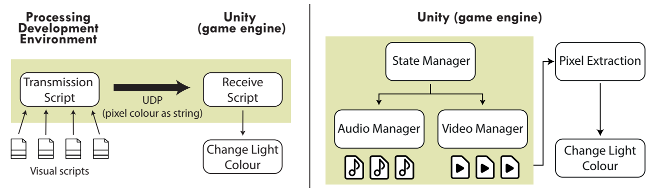

Figure 3: First iteration prototype (left): using Processing to create visuals and Unity to apply them to the virtual robot. Second iteration prototype (right): visual and audio expressions uploaded to Unity and applied to the model.

3.3 User Study

We conducted a remote study with eight experts of various design backgrounds with relevance to our study purpose: two professional sound designers, five interaction designers and one PhD-student in design robotics with a background in product design. The expert sessions were carried out individually via Zoom, each session lasting forty minutes in total. We opted for an evaluation with designers rather than potential users as our aim was less to spot usability issues with the current prototype, but rather support the ideation of alternative design approaches. All sessions were audio and video recorded for later transcription and analysis.

At the start of the testing session, we introduced the experts to the topic of the study and gave them a brief description of the research context. We described how we designed the functional and emotional expressions and asked the experts what they expected those expressions to look and sound like. Then, we asked participants to start the VEP which they had previously downloaded. After getting familiar with the controls (approx. one minute), we asked them to explore the virtual space while continuously thinking out loud [28]. We wanted to uncover the participants' immediate impressions of how they perceived and interpreted the designed expressions. After approximately 7-10 minutes of interaction - when the participants had explored all aspects of the VEP - we asked experts to close the prototype and conducted a 30-minute semi-structured interview. Our line of questioning was directed at how effectively the robot communicated functional and emotional cues, what kind of characteristics and intentions the participants would attribute to it, and what other approaches the experts could envision for the design of the robot's expressions. We kept our questions open-ended and let participants express themselves freely.

4 FINDINGS AND DISCUSSION

Using a VEP to remotely evaluate our designs through qualitative data collection techniques provided rich insights. We found that our participants' impressions of the virtual robot generally matched impressions of the physical robot in two previous in-the-wild studies [19, 20]. This suggests that the method has the potential to be used as an additional lightweight tool for the purpose of remote qualitative data collection. However, we acknowledge that the sample size for the study is small and conducted with only experts - who have more experience in the domain of human-centred design, therefore the data we present here is only indicative. Further studies need to involve larger number of participants, including experts and non-experts, to confirm our findings.

In the following section, we present and discuss the insights from our deployment of the VEP. Using thematic analysis [6] on transcripts and observations, we looked for data points related to findings from previous studies on the physical robot and other related HRI studies. We identified the following key themes:

The virtual robot seemed alive and had character. Participants readily perceived the VEP as animate, and would anthropomorphise it or relate it to other living things such as animals and foreign beings. Moreover, they perceived it as curious, inquisitive and cautious, whilst also being independent and purpose-driven, and projected positive characteristics such as cute, adorable, friendly and playful onto it. Even though they were interacting with a virtual robot on a screen, participants generally acknowledged and empathised with its emotional expressions. As shown in previous studies, emotionally expressive robots are socially more accepted [10, 26]. Crucially, all these associations and characteristics correspond to the ones elicited by the same robot in its physical form during previous deployments [19, 20].

Participants explored the robot's interactive behaviour. We observed that participants would thoroughly explore the rules around the robot's interactive behaviour and would try to get as close to the robot as possible. One potential reason for this was the clear communication of directionality through the virtual robot's eyes, which would dynamically appear on the area of the robot facing the movement direction, or disappear when the robot was in a resting state. Szafir et al. [43] point towards the advantages of having non-humanoid robots communicate direction, noting how it allows users to better predict the robot's behaviour and intentions. In our case, the communication of directionality emphasised the presence of behaviour and intentions, and encouraged participants to discover the specific rules around the robot's operation. A second possible factor was the robot's awareness of the user's presence and actions, which was communicated through various colour patterns and audio cues. Once participants noticed the communication of direction and moved in front of it, they would repeatedly trigger this interaction. This behaviour matches responses during the physical robot's real-world deployment where participants would move very close to the robot to see if it would react to their presence [20]. However, one participant who only moved *once* in front of the virtual robot mentioned: "It was quite obvious with the eyes and [that] it would be moving in this direction... I felt it had a purpose and I did not want to disrupt its purpose in life". This behaviour was also occasionally observed in the first deployment of the physical robot, where some participants felt the robot had a purpose and would therefore stay clear of its path and observe, rather than try to interact with it.

Participants moved around to gain a new perspective. Even though the robot was spherical in shape and looked similar on all sides – except for the side communicating its direction through the low-res display – participants would move around the scene to gain a new perspective. After an initial learning phase of getting familiar with the controls, they made full use of the VEP's spatial possibilities, rather than passively observing the robot like in a video recording. Participants explained, that they wanted to get a clear view where it would go and what would happen if it ran into an object or into them. This again corresponded well with participant behaviour during the robot's physical deployment, when people would walk around it and observe it from different angles in an attempt to fully understand its purpose and behaviour [20].

The prototype's multimodal cues effectively conveyed its various states. All participants were able to identify emotional and functional cues from the visuals, corroborated by the audio. The visual cues were reported to be indicative of emotional states, although we received differing opinions from participants on what those emotions were. This is expected as colours are "situation, history [and] personal dependant" [34]. Participants pointed at the implied meaning of the visuals based on their shade and brightness, whereas the audio cues explicitly communicated the emotional states, hence confirming their expectations. Conversely, the prototype's functional cues were explicitly communicated through visual means, conveying behavioural information such as status or intent, while the audio cues implicitly communicated these. Previous studies on multimodal expression in HRI emphasise the accumulative effects of multimodality, noting how communication via a combination of several modalities can be more effective than via a single one [31, 42, 49]. We found that participants readily saw the modalities used in our VEP – low-resolution light patterns and audio cues – as part of the robot's multimodal communication.

Sense of presence was not satisfied. Participants were constantly aware that they were interacting through a computer screen rather than with a physical robot and reported this as a limitation. This has also been found in prior studies that utilised video-playback for assessing virtual robots [4, 15, 23, 41, 47]. One participant expressed how interacting with the virtual robot through a computer screen was restricting the robot's communication skills. In reference to the robot's visual and audio cues during the awareness interaction, she stressed that "[...] with the computer screen it's a little bit more difficult to express that". We believe that the VEP's lack of spatial sound localisation [37] made it difficult to clearly assign the various audio cues to the robot. The diminished sense of presence in virtual prototypes also led to a lack of tangible rapport with the robot. Physical robots, for example, provide a tactile sensation that cannot be satisfied in VR. One participant in our study mentioned that, "it's like a dog, you just want to pet it". Participants in the physical robot's first deployment [20] would often touch the robot, and children often tried to hug it. Our VEP was not able to provide this sensation and we did not attempt to approximate this kind of interaction using participants' mouse or keyboard input.

5 LESSONS LEARNT

Our motivation for this study was to investigate the feasibility of using VEPs to rapidly prototype interactive systems to collect qualitative data in a remote context. Our findings indicate that the data collected through this method can match those of a physical setup. The overall design process, deployment and subsequent testing revealed some interesting opportunities for further investigation and improvement. By discussing these lessons learnt we hope the HCI community may benefit from the notion of VEPs to collect qualitative data in an online evaluation context. The lessons are presented and discussed throughout the following two sections:

VEP Design and Deployment, and Using VEPs to Augment Remote Interviews.

5.1 VEP Design and Deployment

Virtual prototypes require less effort to build and are easier to maintain. A variety of features may be tested without the necessary hardware, different deployment settings may be designed, and they have the benefit of being more robust for repeated trials [29, 33, 44]. In the context of our research project, the actual physical robot (which we modelled in VR) is yet to have any components installed that enable situational awareness and audio output. This is a limitation, which we were able to circumvent through the use of our VEP. Additionally, we were able to easily program the behaviour of our virtual prototype, including state switches, movement trajectories and obstacle detection. A large number of HRI studies involving physical robots, including the previous deployments of Woodie [19, 20], have employed the Wizard of Oz (WoZ) technique [22], thereby controlling robot behaviour through a human operator [44]. While the implementation of this is simpler than creating fully autonomous behaviour, it still requires operator training, coordination with the moderator and access to, and understanding of, a robot's movement controls. In contrast, our VEP could be realised with little knowledge of robotic systems and VR development, as the necessary software tools, such as Unity, Blender, and Processing, are widely available and provide high-quality educational resources at no cost. Combined with a large and active user base, this made our VEP very accessible both in terms of expertise and cost.

Noticeable gaps and unrelated items should be avoided. We found that several participants tried to discover technical limitations in the virtual environment or were intrigued by objects unrelated to the focus of the study. As we did not enclose the virtual space (see 1, right), participants often tried to move off the scene to explore the limits of the virtual environment. Although movement was restricted to avoid this, some participants needed to be reminded to maintain focus on the robot rather than wander off. To mitigate this issue, we recommend designing a completely enclosed space without any observable gaps. Similarly, some participants started analysing objects within the scene that captured their attention. Hence, minimising the number of objects that are in close proximity of the artefact in question may help with maintaining participant focus, even if this might make the virtual environment less realistic.

5.2 Using VEPs to Augment Remote Interviews

VEPs can have ecological validity for certain research questions. As suggested by our preliminary findings, many similarities were drawn from existing studies with physical robots. Nonetheless, we encountered some challenges specific to VR, the lack of a tactile sensation and the social aspect surrounding urban deployments. Therefore, this form of prototyping may not reach the validity of in-the-wild studies, but can be used as a lightweight tool to answer specific research questions, for example related to the perception and interpretation of a robot's behaviour and embodiment. Further research is required to investigate in a more systematic manner which type of common research questions in HRI can be evaluated through a VEP. Additionally, as our study was restricted to expert participants, it would be worthwhile conducting further investigation of this prototyping approach on target end users. To achieve results more akin to that of real world urban deployments, where interactions are spontaneous [18], it would be appropriate to withhold information on the research to participants. Instead a short description of the situation to simulate a natural scenario should be provided to give participants context before entering the virtual environment [33].

Multi-user support may provide the grounds for a collective sense-making process. The physical robot's first deployment [20] stimulated a honeypot effect [7] and would often have a multitude of people interacting with it at the same time. Therefore, a passer-by could learn of the interactions by observing an active participant [18]. It can be deduced that not only the user's experience, but also the comprehension process is different in a field study scenario in comparison to our virtual prototype. An alternative would be an online virtual world implementation, such as Second Life, which has been applied for learning purposes [11], or the programming of a crowd interacting with the robot in the virtual environment [33], which the user would encounter when arriving at the scene. In order to simulate social aspects and leverage collective sense-making processes, we are planning to further develop our VEP implementation and approach to support multiple remote participants interacting with the robot simultaneously. This also opens up new questions that needs to be addressed: for example, what avatar representations (i.e. low vs. high visual realism) best to be used, and how the current evaluation protocol of using think-aloud and semi-structured interviews needs to be adapted.

Location-independent evaluation. Our research team for this study was distributed across the globe, however through the built-in collaboration capabilities in Unity we were able to develop the prototype simultaneously. Participants also came from various locations and distribution of a software application enabled us to synchronously evaluate the prototype through video conferencing. As such, VEPs provide the grounds for location-independent design teams and the reach of a larger, more diverse participant pool, thereby also providing the means for studying aspects related to cultural differences. This is, we would argue, one of the key benefits of VEPs over physical and immersive VR prototypes. Physical prototypes require the participant to be in the same vicinity. This may involve a lab environment or moving the prototype to a new environment, which has its own set of difficulties [29, 33, 44]. Immersive VR does have the benefit of being lower in cost - depending on the level of immersion, easy to maintain and is suitable for repeated trials when compared to physical prototypes. However, HMDs for immersive VR are just entering the market and only few people own an HMD. Therefore, testing is usually conducted in a lab environment or in a safe space clear of obstructions [1, 12, 33, 40].

Think-aloud protocol and interviews seamlessly applied. We found that participants were able to employ the think-aloud protocol with ease. Participants would often comment on the changes in visuals and audio and how they perceived these changes, whilst also attempting to understand where the robot would go and what would happen if they blocked its path. Data extracted from the technique allowed us to assess participants' impressions and how they went about understanding the VEP. Therefore, we were able to obtain rich insights for our designs. This technique is not commonly used in immersive VR studies. Instead researchers are opting to collect qualitative data through post-experience interviews and questionnaires [12, 25], or quantitative data through eye tracking and task completion time [1, 33, 40], in order to not interrupt participants' immersive VR experience.

On the other hand, when testing in an urban context, using the think-aloud protocol while the participant is interacting with the prototype is more difficult. Due to the transient nature of public spaces, interactions with urban interfaces are more fluid and interaction duration may be short [18], therefore employing a think-aloud protocol is often not viable. Instead, a participant's actions are observed at first, followed by a post-experience interview. However, participants recruited on site may not be interested in giving an in-depth interview, which is therefore often kept short [19, 20]. In contrast, with our VEP, we were able to obtain rich insights into a participant's thoughts and impressions during interaction. Further, participants may feel more comfortable interacting and providing feedback in a remote study setup, as it prevents perceived risk of social embarrassment [38].

Video conferencing for observing and recording participant interactions and expressions. Participants were asked to share their screen during interaction with the prototype, so that we could observe their actions in the virtual environment. At the same time, a video stream of the participant's camera was still available to us during screen-sharing. A benefit here over physical testing is the reduced amount of equipment required to record user interactions. While physical testing of self-moving interfaces and spatial experiences may require multiple cameras and microphones, this form of testing only needs a single webcam and microphone, which most computers are equipped with. Facial expressions and speech are the main focus, allowing for easy application of transcription software. Recordings in a physical space may miss some cues due to non-optimal positioning of recording equipment. This said, virtual testing can miss out on bodily expressions, which are also key for interpretation of user impressions [14]. Another issue with conferencing software is the requirement for an internet connection to be maintained throughout testing. One participant's connection dropped while he was interacting with the prototype. This slight pause in the interaction can be detrimental to the quality of results as participants are required to backtrack. Additionally, some level of knowledge of the conferencing software is expected on part of the participant cohort. Participants who aren't well acquainted with conferencing software are expected to face more difficulties with this form of testing, hence leading to longer session.

6 CONCLUSION

In the context of a study on functional and emotional cues for urban robots, we described the design, deployment and evaluation of what we refer to as a Virtual Experience Prototype (VEP). We designed a virtual instance of a previously deployed urban robot and conducted a remote study, in which participants interacted within the virtual environment through a game engine. During the participant's experience, we collected qualitative data using the think-aloud protocol, followed by semi-structured interviews.

Our study offers insights into the remote collection of qualitative data for the evaluation of interactive spatial experiences and human-robot interaction. It draws from previous work that highlights how remote data collection can in some cases provide a valid alternative to in-lab and in-the-wild studies. While bearing in mind the limitations over physical experience prototypes and evaluations in a real-world environment, VEPs employed in remote study setups provide several benefits: they are lightweight in development and deployment, and can be evaluated location-independent with a broad pool of potential participants.

By drawing connections between our findings and previous deployments of the robot in its physical form, we showed how several of our participants' experiences, impressions and behaviours largely match those observed during in-person evaluations. The emerging themes included readily perceiving the robot as animate and assigning a persona, exploring the robot's interactive capabilities through trial-and-error, closely observing the robot's activities in order to understand its purpose, and perceiving the robot's individual modalities as part of its multi-modal communication.

To sum up, VEPs present a specific instance of experience prototypes [8], which come with new opportunities for remote evaluation, however also a range of considerations to be made when creating them. Unlike often the case with physical experience prototypes, virtual experience prototypes require not only to design the artefact itself, however also the surrounding context and environment in which the artefact is situated in. Based on our study, we presented a range of lessons we have learned along the design and evaluation process. Further we indicated areas of potential future work, such as supporting multi-user remote evaluations to enable collective sense-making processes and simulate multi-user interactions with public interfaces. We hope our study provides a starting point for other designers and researchers who are aiming to evaluate interactive spatial experiences remotely in a virtual setting.

ACKNOWLEDGMENTS

Many thanks to all the participants for taking part in this research. We would also like to thank Lian Loke for her valuable feedback on the initial research. We thank the anonymous reviewers of OzCHI '20 for their constructive feedback on how to make this a stronger contribution. A special thanks to the co-authors of this paper for their constructive feedback and input.

REFERENCES

[1] Philipp Agethen, Viswa Subramanian Sekar, Felix Gaisbauer, Thies Pfeiffer, Michael Otto, and Enrico Rukzio. 2018. Behavior Analysis of Human Locomotion in the Real World and Virtual Reality for the Manufacturing Industry. *ACM Trans. Appl. Percept.* 15, 3, Article 20 (July 2018), 19 pages. https://doi.org/10.1145/3230648

[2] William Albert and Thomas Tullis. 2013. *Measuring the user experience: collecting, analyzing, and presenting usability metrics.* Elsevier Science, Burlington, MA.

[3] Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. 2012. How to Evaluate Public Displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays* (Porto, Portugal) *(PerDis '12)*. Association for Computing Machinery, New York, NY, USA, Article 17, 6 pages. https://doi.org/10.1145/2307798.2307815

[4] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52.

[5] Cade E Bartlett and Nancy J Cooke. 2015. Human-robot teaming in urban search and rescue. *Proceedings of the Human Factors and Ergonomics Society Annual*

Meeting 59, 1 (2015), 250–254. https://doi.org/10.1177/1541931215591051
[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3 (01 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[7] Harry Brignull and Yvonne Rogers. 2003. Enticing People to Interact with Large Public Displays in Public Spaces.. In *Proceedings of INTERACT*. IOS Press, Brighton, UK, 17–24.
[8] Marion Buchenau and Jane Fulton Suri. 2000. Experience Prototyping. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (New York City, New York, USA) *(DIS '00)*. Association for Computing Machinery, New York, NY, USA, 424–433. https://doi.org/10.1145/347642.347802
[9] Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
[10] Jessica Rebecca Cauchard, Kevin Y. Zhai, Marco Spadafora, and James A. Landay. 2016. Emotion Encoding in Human-Drone Interaction. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Christchurch, New Zealand, 263–270.
[11] Andrea De Lucia, Rita Francese, Ignazio Passero, and Genoveffa Tortora. 2009. Development and evaluation of a virtual campus on Second Life: The case of SecondDMI. *Computers & Education* 52, 1 (2009), 220–233. https://doi.org/10.1016/j.compedu.2008.08.001
[12] Shuchisnigdha Deb, Daniel W Carruth, Richard Sween, Lesley Strawderman, and Teena M Garrison. 2017. Efficacy of virtual reality in pedestrian safety research. *Applied ergonomics* 65 (2017), 449–460. https://doi.org/10.1016/j.apergo.2017.03.007
[13] Susan Dray and David Siegel. 2004. Remote Possibilities? International Usability Testing at a Distance. *Interactions* 11, 2 (March 2004), 10–17. https://doi.org/10.1145/971258.971264
[14] Julius Fast. 1970. *Body language*. Souvenir Press, London.
[15] Marina Fridin and Mark Belokopytov. 2014. Embodied robot versus virtual agent: Involvement of preschool children in motor task performance. *International Journal of Human-Computer Interaction* 30, 6 (2014), 459–469. https://doi.org/10.1080/10447318.2014.888500
[16] Ronald E Giachetti, Veronica Marcelli, José Cifuentes, and José A Rojas. 2013. An agent-based simulation model of human-robot team performance in military environments. *Systems Engineering* 16, 1 (2013), 15–28. https://doi.org/10.1002/sys.21216
[17] Hamish Henderson, Martin Tomitsch, and Tuck Wah Leong. 2018. Tools to Think with: Augmenting User Interviews with Rapid Modular Prototypes. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne, Australia) *(OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 256–260. https://doi.org/10.1145/3292147.3292209
[18] Luke Hespanhol and Martin Tomitsch. 2015. Strategies for Intuitive Interaction in Public Urban Spaces. *Interacting with Computers* 27, 3 (2015), 311–326. https://doi.org/10.1093/iwc/iwu051
[19] Marius Hoggenmueller, Jiahao Chen, and Luke Hespanhol. 2020. Emotional Expressions of Non-Humanoid Urban Robots: The Role of Contextual Aspects on Interpretations. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays* (Manchester, United Kingdom) *(PerDis '20)*. Association for Computing Machinery, New York, NY, USA, 87–95. https://doi.org/10.1145/3393712.3395341
[20] Marius Hoggenmueller, Luke Hespanhol, and Martin Tomitsch. 2020. Stop and Smell the Chalk Flowers: A Robotic Probe for Investigating Urban Interaction with Physicalised Displays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376676
[21] Kai Holländer, Luca Schellenberg, Changkun Ou, and Andreas Butz. 2020. All Fun and Games: Obtaining Critical Pedestrian Behavior Data from an Online Simulation. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382797
[22] John F Kelley. 1985. *Cal-A Natural Language Program Developed with the Oz Paradigm: Implications for Supercomputing systems*. IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. 238–248 pages.
[23] Cory D Kidd and Cynthia Breazeal. 2004. Effect of a robot on user perceptions. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, Vol. 4. IEEE, Sendai, Japan, 3559–3564.
[24] Takanori Komatsu. 2005. Toward Making Humans Empathize with Artificial Agents by Means of Subtle Expressions. In *Affective Computing and Intelligent Interaction*, Jianhua Tao, Tieniu Tan, and Rosalind W. Picard (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 458–465.
[25] Dannie Korsgaard, Niels Christian Nilsson, and Thomas Bjørner. 2017. Immersive eating: evaluating the use of head-mounted displays for mixed reality meal sessions. In *2017 IEEE 3rd Workshop on Everyday Virtual Reality (WEVR)*. IEEE, Los Angeles, CA, USA, 1–4.
[26] Miriam Koschate, Richard Potter, Paul Bremner, and Mark Levine. 2016. Overcoming the Uncanny Valley: Displays of Emotions Reduce the Uncanniness of Humanlike Robots. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Christchurch, New Zealand, 359–365.
[27] Frederick Scott Hunter Krauss. 2003. Methodology for remote usability activities: A case study. *IBM Systems Journal* 42, 4 (2003), 582–593.
[28] Clayton Lewis and John Rieman. 1993. *Task-Centered User Interface Design: A Practical Introduction*. University of Colorado, Boulder, Department of Computer Science, Boulder, CO, USA.
[29] Michael Lewis, Jijun Wang, and Stephen Hughes. 2007. USARSim: Simulation for the study of human-robot interaction. *Journal of Cognitive Engineering and Decision Making* 1, 1 (2007), 98–120. https://doi.org/10.1177/155534340700100105
[30] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The Anatomy of Prototypes: Prototypes as Filters, Prototypes as Manifestations of Design Ideas. *ACM Trans. Comput.-Hum. Interact.* 15, 2, Article 7 (July 2008), 27 pages. https://doi.org/10.1145/1375761.1375762
[31] Diana Löffler, Nina Schmidt, and Robert Tscharn. 2018. Multimodal expression of artificial emotion in social robots using color, motion and sound. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, USA, 334–343.
[32] Kapil Chalil Madathil and Joel S Greenstein. 2017. An investigation of the efficacy of collaborative virtual reality systems for moderated remote usability testing. *Applied Ergonomics* 65 (2017), 501–514. https://doi.org/10.1016/j.apergo.2017.02.011
[33] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376796
[34] Niels A. Nijdam. 2005. Mapping emotion to color.
[35] Chelsea Owensby, Martin Tomitsch, and Callum Parker. 2018. A Framework for Designing Interactions between Pedestrians and Driverless Cars: Insights from a Ride-Sharing Design Study. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne, Australia) *(OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 359–363. https://doi.org/10.1145/3292147.3292218
[36] Claudia Repetto, Serena Germagnoli, Stefano Triberti, and Giuseppe Riva. 2018. *Learning into the Wild: A Protocol for the Use of 360° Video for Foreign Language Learning*. Springer International Publishing, Cham, Switzerland, 56–63. https://doi.org/10.1007/978-3-030-01093-5_8
[37] Frederic Anthony Robinson, Oliver Bown, and Mari Velonaki. 2020. Implicit Communication through Distributed Sound Design: Exploring a New Modality in Human-Robot Interaction. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 597–599. https://doi.org/10.1145/3371382.3377431
[38] Yvonne Rogers and Harry Brignull. 2002. Subtle ice-breaking: encouraging socializing and interaction around a large public display. In *CSCW'02 Workshop Proceedings*. Citeseer, New Orleans, LA, USA, 1–6.
[39] Jean Scholtz. 2001. Adaptation of traditional usability testing methods for remote testing. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE, Maui, HI, USA, 8–pp.
[40] Helmut Schrom-Feiertag, Volker Settgast, and Stefan Seer. 2017. Evaluation of indoor guidance systems using eye tracking in an immersive virtual environment. *Spatial Cognition & Computation* 17, 1-2 (2017), 163–183. https://doi.org/10.1080/13875868.2016.1228654
[41] Stela H Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E Young. 2015. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Portland, OR, USA, 125–132.
[42] Sichao Song and Seiji Yamada. 2017. Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 2–11. https://doi.org/10.1145/2909824.3020239
[43] Daniel Szafir, Bilge Mutlu, and Terry Fong. 2015. Communicating Directionality in Flying Robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, Oregon, USA) *(HRI '15)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2696454.2696475
[44] Andrew Talone, Thomas Fincannon, David Schuster, Florian Jentsch, and Irwin Hudson. 2013. Comparing physical and virtual simulation use in UGV research: lessons learned from HRI research with two test beds. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 57. SAGE Publications Sage, Los Angeles, CA, USA, 2017–2021. https://doi.org/10.1177/1541931213571451
[45] Kazunori Terada, Atsushi Yamauchi, and Akira Ito. 2012. Artificial emotion expression for a robot by dynamic color change. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*.

IEEE, Paris, France, 314–321. https://doi.org/10.1109/ROMAN.2012.6343772

[46] Ning Wang, David V Pynadath, KV Unnikrishnan, Santosh Shankar, and Chirag Merchant. 2015. Intelligent agents for virtual simulation of human-robot interaction. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, Cham, Switzerland, 228–239.

[47] Sarah Woods, Michael Walters, Kheng Lee Koay, and Kerstin Dautenhahn. 2006. Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In *9th IEEE International Workshop on Advanced Motion Control, 2006*. IEEE, Istanbul, Turkey, 750–755.

[48] Selma Yilmazyildiz, David Henderickx, Bram Vanderborght, Werner Verhelst, Eric Soetens, and Dirk Lefeber. 2011. EMOGIB: Emotional Gibberish Speech Database for Affective Human-Robot Interaction. In *Affective Computing and Intelligent Interaction*, Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 163–172.

[49] Selma Yilmazyildiz, David Henderickx, Bram Vanderborght, Werner Verhelst, Eric Soetens, and Dirk Lefeber. 2013. Multi-Modal Emotion Expression for Affective Human-Robot Interaction. In *Proceedings of the Workshop on Affective Social Speech Signals*. Research Gate, Grenoble, France.

[50] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through Design as a Method for Interaction Design Research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704