



HAL
open science

Design and Evaluation of Three Interaction Models for Manipulating Internet of Things (IoT) Devices in Virtual Reality

Günter Alce, Eva-Maria Ternblad, Mattias Wallergård

► **To cite this version:**

Günter Alce, Eva-Maria Ternblad, Mattias Wallergård. Design and Evaluation of Three Interaction Models for Manipulating Internet of Things (IoT) Devices in Virtual Reality. 17th IFIP Conference on Human-Computer Interaction (INTERACT), Sep 2019, Paphos, Cyprus. pp.267-286, 10.1007/978-3-030-29390-1_15 . hal-02877677

HAL Id: hal-02877677

<https://inria.hal.science/hal-02877677v1>

Submitted on 22 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Design and evaluation of three interaction models for manipulating Internet of Things (IoT) devices in Virtual Reality

Günter Alce¹, Eva-Maria Ternblad¹ and Mattias Wallergård¹

¹ Lund University,

Abstract. More and more things are getting connected to the internet, including lights, speakers, and refrigerators. These connected things are an example of what a smart home system that is part of the Internet of Things (IoT) can incorporate. IoT enables advanced services by interconnecting physical and virtual things. But, building interactive prototypes for smart home systems can be difficult and costly, since it involves a number of different devices and systems with varying technological readiness level. Virtual reality (VR) is a technology that can create computer-generated environments and has been used as a design tool in many different domains, such as architecture, city planning, and industrial design. However, the focus has traditionally been on visualizing design proposals rather than letting the intended users directly interact with them. Recently, we have seen an intensified development of VR headsets such as HTC Vive and Oculus Rift. These headsets come with relatively well-developed hand controllers, which can be used to interact with the virtual environment. This opens up opportunities to develop and evaluate interactive virtual smart home systems.

This paper presents three interaction models developed and evaluated using the new generation of VR technology. The interaction models were then compared in a user study with 18 participants. Some statistically significant differences and subjective preferences could be observed in the quantitative and qualitative data respectively.

The main contribution of this paper is to elucidate knowledge about using VR as a prototyping tool to explore IoT interaction. Moreover, this study implies that you can collect and analyze data for statistical analysis using VR.

Keywords: Virtual Reality, Internet of Things, Method, Interaction.

1 Introduction

Virtual reality (VR) has been used as a design tool in many different domains, such as architecture, city planning, and industrial design [1]. However, the focus has traditionally been on visualizing design proposals for the users rather than letting them directly interact with them. The main problem has been the maturity of VR

technology; it has been either too expensive or difficult to build VR applications that allow for immersive embodied interaction.

This has changed with the latest generation of commercial VR hardware (e.g. HTC Vive and Oculus Rift) which comes with tracking of headset and hand controllers in sub-millimeter precision. Moreover, they allow the user to walk around and interact with the virtual environment (VE) in a relatively realistic manner.

A general benefit of using a VE to build prototypes of interactive systems is that it allows researchers to test complex systems or hardware that does not actually exist in a controlled manner. A smart home system is an example of a complex system, which is difficult to develop and test [2, 3]. The complexity of a smart home system increases since more and more things are being connected to the Internet, including lights, dishwasher and refrigerator. Everything that has a unique id and is sending data over a network can be considered as part of the Internet of Things (IoT) [4].

Several smart home frameworks are currently being developed, with an application running on a mobile device that can control things in a smart home. Examples of such frameworks are the Samsung SmartThings [5], the Apple HomeKit [6] and Google Weave [7]. However, having yet another application to control things is perhaps not the best solution and does not utilize the potential of IoT interaction in a smart home environment. Building interaction prototypes with these applications can be difficult and costly, since it involves a number of different devices and systems with varying technological readiness level [3]. In particular, it is difficult to achieve prototypes that offer an integrated user experience and show the full potential of IoT interaction concepts. The ideal prototyping methodology would offer high-fidelity at a relatively low cost and the ability to simulate a wide range of IoT use cases.

This paper presents three interaction models, which were developed and evaluated in a controlled experiment, using the new generation of VR technology such as HTC Vive [8] together with the game engine Unity [9].

The main contribution of this paper is to elucidate knowledge about the method of using VR as a prototyping tool to explore IoT interaction for a smart home environment.

2 Related Work

Using VR as a prototyping method is an area that has been well studied. This section reviews previous related research in using VR as a prototyping method including earlier user experiments and evaluation methods.

2.1 Using VR to simulate smart home systems

A number of researchers have been using different simulation tools to prototype smart home systems. The main argument of using simulation to prototype IoT environments is to reduce development time and cost [2]. There are many different smart home

simulators with a variety of fidelity. SHSim [2] is built on a dynamic mechanism that allows the user to configure the system and test different use cases. Other examples of smart home simulators are UbiWise [10] that can show a close-up view of virtual devices, TATUS [11] that can simulate adaptive systems, UbiReal [12] that have functions to facilitate deployment of virtual devices, and Furfaro et al. [13] which tried to illustrate how VEs can be a valuable tool to assess security properties. However, these tools are desktop based tools that mainly focus on the installation and configuration of a smart home system and lack embodied user interaction.

Other simulation tools take in account context awareness and some user interaction such as Nguyen et al. [14] proposed Interactive SmartHome Simulator (ISS) that models the relationship between the environment and other factors and simulates the behavior of an intelligent house. CASS [15] is a context-aware tool which can generate the context information associated with virtual sensors. Armac and Retkowitz [16] proposed eHomeSimulator, that can be used to simulate multiple environments with different spatial structures. Hu et al. [17] proposed a web-based tool to check the home status and control devices with a 3D interface. However, the focus of this work is more on a system level and hence they do not offer high fidelity user interaction with IoT devices.

2.2 Using VR for interaction

There are several examples of research attempts to develop immersive VR user interaction in different domains. For example, Bowman and Wingrave [18] used VR to design various types of menu systems to be used within the VE i.e. 3D graphical user interface. de Sá and Zachmann [19] investigated the steps needed to apply VR in maintenance processes in the car industry. In the paper, they present several interaction paradigms e.g. how to assemble the front door of a car, and other functionalities, which a VR system needs to support. The results from their study show the users being very optimistic of how VR can improve the overall maintenance process. Alce et al. [20] used VR for simulating IoT interaction with glasses-based AR. This had the advantage of creating a realistic experience in terms of AR display resolution and tracking. However, it might be hard for the user to discriminate the augmented stimuli from the VE and it can be difficult to interact with the simulated environment in an easy and realistic manner. Furthermore, the movements of the user are restricted by the range of the VR tracking system, which makes some use cases difficult or impossible to simulate. More recently, Ens et al. [21], introduced Ivy which is a spatially situated visual programming tool using immersive VR. Ivy allows users to link smart objects, insert logic constructs and visualize real-time data flows between real-world sensors and actuators. However, Ivy focus on how to configure a smart home system and not on how users can discover and control with the IoT devices.

2.3 Using VR with Wizard of Oz method

Wizard of Oz (WOZ) is a well-known method where a human operates undeveloped components of a technical system. Above all, the WOZ method has been widely used in the field of human-computer interaction to explore design concepts. Carter et al. [22] state that WOZ prototypes are excellent for early lab studies but do not scale to longitudinal deployment because of the labor commitment for human-in-the-loop systems. Recently Gombac et al. [23] used WOZ for prototyping multimodal interfaces in VR. They compared voice and mid-air gesture interface while interacting with a computer system. WOZ is often used for voice interaction and one of the interaction model presented in this study used WOZ to simulate voice interaction together with head-gaze.

In summary, over the past 20 years, researchers have developed a range of different VR simulators to prototype different systems. User studies show the benefits of VR as a prototyping tool regarding simulation to save time and cost. However, it seems that there has been little research on using VR to prototype IoT interaction. The described approaches have their merits in the view of using VR as a prototyping tool but lack exploring user interaction models, which can be a problem if it is discovered late in the project when the smart home system is put into operation. VR facilitates the development of IoT applications since it is cheaper and easier to add a myriad of virtual devices compared to real devices [3]. Therefore, we have focused on utilizing the new VR hardware such as HTC Vive as a prototyping tool for immersive embodied VR IoT interaction.

3 Building the prototype

One of the main goals with the presented work was to design and test a set of embodied IoT interaction models by exploring the possibilities and technical advantages with the VR environment. This prototyping method implicates that relatively futuristic models can be explored, and technical obstacles can be avoided in favor of human preferences, natural behavior, and cognitive capacities. Consequently, the process of designing the prototype was performed in an iterative approach, starting with user preferences and a wider concept of IoT interaction, followed by exploring and testing implementation possibilities in VR until the final prototype was developed and evaluated.

3.1 Low-fidelity interaction and user observations

For choosing appropriate types and modes of interaction, six participants (one male and five women) were recruited within the project members' social network. The participants were between 17 to 52 years old, had different backgrounds and different experiences concerning smart home systems. Each session last about 15 minutes. The participants were invited to a real but small living room, where they were told to imagine a variety of day-to-day objects as being hyper-intelligent and connected to

the Internet. They were then asked to freely “interact” with these objects with the use of their own bodies and modalities, without any help from devices, remote controls or traditional interfaces. During the test a concurrent think-aloud protocol (CTA) was utilized. CTA is a common procedure within the field of usability testing that is considered to be both reliable and cost efficient [24, 25]. Being well aware of that the question about whether CTA affects user performance or not has been debated over several years [26, 27] we concluded that the benefits from this method outweighed the disadvantages. Asking the participants to share their ideas and thoughts with us made it possible to retain valuable tacit information.

After the introduction, the participants were asked to interact with four devices:

- **TV**, turning it on/off, change channel, increase volume
- **Light bulbs**, turning on/off and changing the luminance
- **Music player**, to turn it on/off, select favorite song and increase the volume
- **Coffee machine**, start the coffee machine which is not in the room.

Although the interaction was completely imaginary and performed without any type of feedback, this highly explorative test led to very useful findings, making it possible to select a limited sample of plausible and testable types of IoT-interaction. Interestingly, the participants showed similar preferences concerning the choice of interaction modalities including voice commands and gestures. However, their personal expressions of precise manipulations of hand gestures and verbal utterances had a larger variation (see Fig. 1a-d). One participant would use voice for all interaction except for selecting a device which pointing was preferred. Four participants used pointing with finger, one used open hand and one used the fist and wanted to turn/on/off by opening the fist (see Fig. 1c and 1d). The same participant would also like to be able to increase and decrease the volume or the luminance of the light bulbs depending on how much the fist was open, so fully open hand would be max volume or luminance while half open would be half volume or luminance depending on which device was selected.



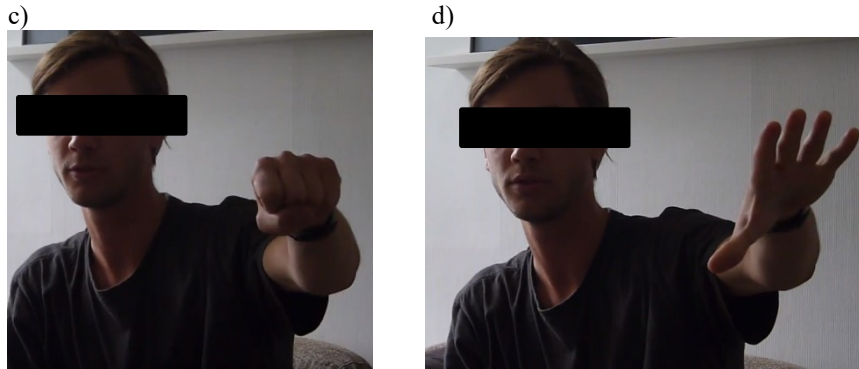


Fig. 1. a) Pointing to start b) swiping gesture c) close lights or TV d) Open lights or TV.

After discussing, retrying and analyzing these outcomes a set of interaction types were chosen and categorized according to Ledo et al.'s [28] four different stages of interaction: Discover, Select, View status and Control. This classification helps to identify necessary components that have to be fulfilled in an IoT environment. Finally, after subsequent low-fidelity testing, the following interaction patterns were decided to advance into the stage of VR prototyping and implementation:

1. To discover objects: Raise one hand
2. To select objects: Point with your hand; or head-gaze; or proximity (walk towards an object)
3. To view the status of objects: Feedback from the objects themselves and/or through a virtual smartwatch device
4. To control objects: Point and click; or head-gaze and voice; or simple hand movements

3.2 VR Implementation

The interaction with the IoT objects was decided to take place within a realistic but sparsely furnished virtual living room. Moreover, for making it possible to study and evaluate IoT-interaction in relation to traditional object manipulation such as the use of switches for turning on/off lamps, this interaction type was added. Since the difference between selecting and controlling objects could not easily be defined, and since a strict and uniform use of the two stages could be perceived as both strange and unnecessarily complicated for several of the IoT objects, these two stages of interaction were merged into one. Finally, the following interaction models were implemented:

1. To discover objects: Raise one hand
2. To view the status of objects: Feedback from the objects themselves and/or through a virtual smartwatch device showing the status of the object
3. To select and control objects: Point and click; head-gaze together with voice; and physical manipulation (the participant had to walk to the wall and press on the switches in which they got haptic feedback)

The virtual living room consisted of four wall-mounted lamps, two pot plants and a TV as IoT objects (see Fig. 2). The size of the virtual living room was experienced as ten by ten meters large while the physical space were nine square meters due to the limited tracking space of the VR system. The VE was implemented with Unity ver. 5.5.0f3, and the VR hardware was an HTC Vive with resolution 2160x1200 (1080x1200 per eye) and > 90 frames per second frame rate, rendering the user a sufficient realistic experience without delays or visual discomfort.

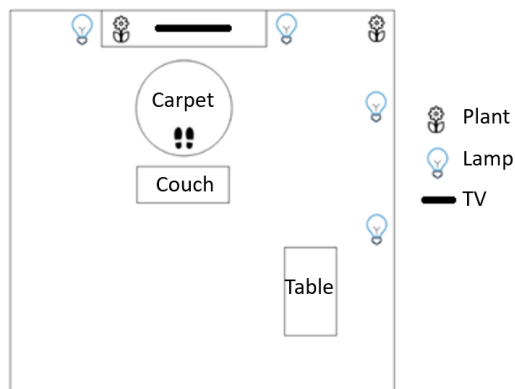


Fig. 2. A drawing of the virtual living room with four wall-mounted lamps, two pot plants and a TV as IoT-objects.

Discover devices

To discover IoT objects, the user had to raise the right hand above the head (see Fig. 3a). Since the gesture is common and natural, it was presumed to be used with ease and expertise. As long as the participant's hand remained above the head, all IoT objects replied by casting a beam of yellow light, indicating their belonging to the IoT-family and showing that they were ready to be used and controlled accordingly.

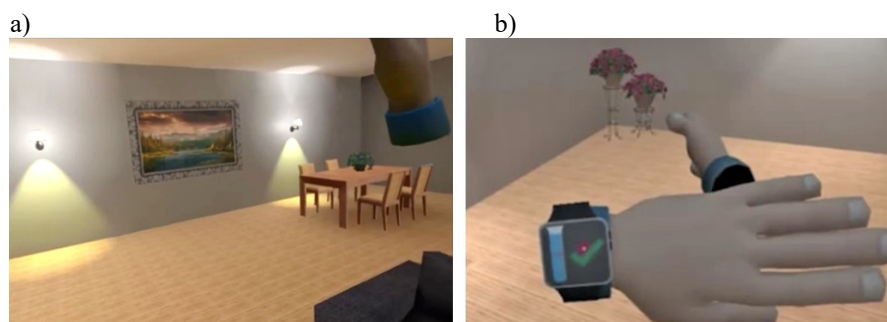


Fig. 3. a) IoT objects could be discovered by rising the hand. b) Looking at feedback from one of the pot plants.

View status

In the prototype, some of the objects (such as the TV and the lamps) changed and revealed their status themselves when the participant was controlling them i.e. turning them on/off. Another way to see the status of the IoT objects was through the virtual smartwatch situated on the participant's left wrist (see Fig. 3b).

Select and control a device

To select and control an IoT device, the participant could:

- a) **Point and click**, pointing was performed by directing the right or the left-hand control towards a chosen object and pressing a control button. This command animated the active virtual hand into a pointing gesture (see Fig. 3b) and used an invisible ray-casting technique (i.e. removed the line segment attached to the participant's hand which would have represented the direction of the ray) towards the actual object. For the chosen object to be activated, the direction of the ray-cast (although invisible) had to hit the item within a defined area, equal to or slightly larger than the object's virtual boundaries. Since, it is not easy to select a small object distant from the participant without a visible ray-cast neither in VR nor in reality.
- b) **Head-gaze together with voice**, the participant could also use head-gaze and voice for activating IoT devices. The reason for merging these communicative tools is mainly due to their nature. First, to use gaze as a single behavioral pattern for selecting or controlling objects would not be easy. The human eye is almost never still, blending longer fixations and saccades into complex patterns [27], and it is quite possible to show an interest in an object without wanting to interact with it. Secondly, voice interaction in general is unable to discover what you can interact with and is unable to make all possible actions visible to the user. You could perhaps ask a "virtual voice assistant", but you might have trouble remembering what was listed and what the different devices were called. Moreover, as pointed out by Norman and Nielsen [29], user interfaces built on gestures and speech interaction lack several fundamental principles of interaction design. These are principles that are completely independent of technology, such as visibility (affordances or signifiers), feedback, consistency, non-destructed operation (undo), discoverability, scalability and reliability [29]. This is, of course, less of a problem in a familiar home environment, where the user knows what devices and services are available, and where they are located. However, in an unknown environment, such as a new workplace, it could be difficult for a user to discover nearby devices and their capabilities. Hence, we decided to use head-gaze together with voice for turning on and off objects, making the head direction the selecting part and the voice the controlling part. However, since the used VR equipment does not support eye tracking or voice control, this interaction model was simulated with a WOZ solution. We chose identical and extremely easy commands for all implemented smart objects, namely "ON" for turning objects on and "OFF" for turning objects off. One of the test leaders were acting as a wizard and used predefined keys on the computer to

turn on and off each virtual device in the virtual living room. For example, the “1” key was used to alternate between turning on and off the TV. The Wizard could follow what the test person was looking at on the computer display and could hear the test person during the whole test session. Prior the test session the wizard trained to avoid mistakes during the real test.

- c) **Traditional switch buttons**, switch buttons mounted on the virtual wall also controlled the four light bulbs in our VR-prototype. The vibration from the hand controller gave haptic feedback when the participant pressed the button. This implementation made it possible to compare traditional and futuristic ways of interaction even though it is important to point out the fact that all interactions were virtual.

4 Experiment

A comparative evaluation was conducted in a VR laboratory environment to compare the proposed interaction models: a) point and click; b) head-gaze together with voice; and c) traditional switch button for selecting and controlling IoT devices. Both quantitative and qualitative data were collected. The purpose of this test was mainly to explore the participants’ preferences and to identify possible differences between the interaction models in regard to physical or cognitive load. At the same time, we wanted to gather valuable information concerning participants’ immersion and feel of reality.

As dependent variables, we used NASA TLX values, and individual ratings for the interaction patterns. The two main null hypotheses was that neither the NASA TLX values nor the individual ratings would differ in regard to the type of interaction model. We also decided to analyze qualitative data concerning any stated difficulties with certain types of interaction and comments on feedback, and to measure user presence with the use of a standard questionnaire.

4.1 Setup

The evaluation was conducted in a VR laboratory environment with audio and video recording facilities. One single session involved one participant and two test leaders, where one was in charge of the HTC Vive equipment and one guided the participants through the test and performed interviews (see Fig. 4). All test sessions were recorded.

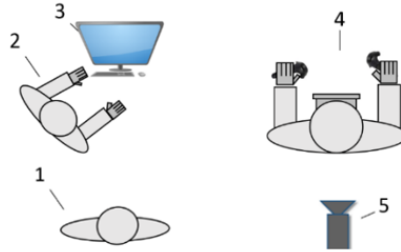


Fig. 4. The experiment setup. 1) test leader (introducing), 2) test leader (HTC Vive equipment), 3) computer running the HTC Vive, 4) test participant, 5) camera.

4.2 Participants

18 participants were recruited by notifications on Facebook and through advertisements on public billboards at university faculties and cafés. The participants consisted of twelve males and six females, between 19 and 51 years old ($M = 25.1$) and from various backgrounds (although 12 of them were students at the University). Nine of them had previous experience of VR while nine of them had none.

4.3 Procedure

All participants were given a brief introduction to the project and its purpose. Next, all participants filled in a short questionnaire together with informed consent regarding their participation and the use of collected data. Thereafter they were introduced to the HTC Vive where they performed a quick training session in a very basic and minimalistic VE. The purpose of this exercise was for the participants to get familiarized with the fictive world and the virtual interactive patterns of pressing buttons and pointing at objects (see Fig. 5).

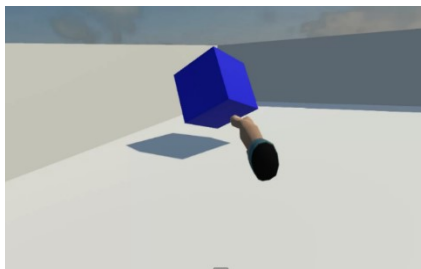


Fig. 5. Pointing is exercised by aiming at virtual cubes and spheres in a minimalistic environment.

After one or two rounds of training (mainly depending on the participants' capacity of successfully pointing at objects), the participants were sent into the virtual living room. Here they were introduced to the raising-hand-gesture for discovering IoT devices as well as to the smartwatch device for feedback, and they were asked to get familiarized with the environment and talk about their experience. Subsequently, they

interacted with all of the four lamps in the room by turning them on and off from left to right. This was done three times, one time for each interaction model.

In an attempt to understand and describe the users' perceived workload NASA TLX was used as an assessment tool. Although, it is normally reported in writing we let the participants respond orally. Moreover, we simplified the NASA TLX scales, the scores were calculated out of 10 instead of 100. NASA TLX is commonly used to evaluate perceived workload for a specific task. It consists of two parts. The first part is referred to as raw TLX (RTLX) and consists of six subscales (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) to measure the total workload. The second part of the NASA TLX creates an individual weighting of the subscales by letting the subjects compare them pairwise based on their perceived importance. However, as reported by Hart [30], using the second part of the NASA TLX might actually decrease experimental validity. For this reason, it was not used in this experiment.

To make the physical effort equal in all interaction models, each one of the scenarios were initiated by directing the participant to a specific starting point marked on the virtual carpet by two-foot prints (see Fig. 2). To avoid sequential effects, the order of the interactive types were balanced to obtain all possible orders (see Table 1).

Table 1. The Interactive types were balanced accordingly.

Test Person	1 st Interaction	2 nd Interaction	3 rd Interaction
1, 7, 13	Switchers	Pointing	Gaze & Voice
2, 8, 14	Pointing	Gaze & Voice	Switches
3, 9, 15	Gaze & Voice	Switches	Pointing
4, 10, 16	Switches	Gaze & Voice	Pointing
5, 11, 17	Pointing	Switches	Gaze & Voice
6, 12, 18	Gaze & Voice	Pointing	Switches

After this, the participants were allowed to interact with all smart objects in the room, choosing one or several interactive patterns of their own liking. They were also asked to verbally report the visual feedback from the smartwatch. The VR-session ended with a short interview regarding the subject's individual interactive preferences, perceived difficulties under the test and his or her general VR experience. The questions asked during the interview:

1. If you got to rank your interaction methods. Which one was the best, the worst and the one between them?
2. What did you experience as hardest in the entire test?
3. What did you experience as easiest in the entire test?
4. Was there any moment that surprised you?
5. Do you have any comments on the VR environment?
6. Do you have any comments on the test itself?

Finally, the participants filled in the Slater-Usoh-Steed Presence Questionnaire (SUSPQ), which is a common tool for measuring the presence a user experiences in a VE [31], and received coffee and cake as a reward for participating. Each session last

about 30 min. The whole procedure of the test session is visualized in a block diagram (see Fig. 6).

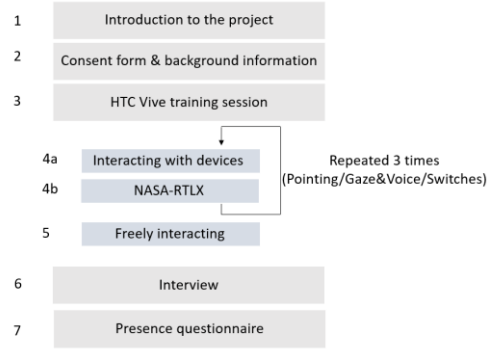


Fig. 6. Test session procedure.

4.4 Result

The measurements were successful and all participants performed all moments without incidents or major problems. The training session was generally performed twice (six participants did it only once) and eight of the participants needed to exercise the pointing gesture with the help of additional instructions and tips. The most common problem in this session was to point at a distant object, and two of the participants had to practice with a visible ray-casting technique, i.e. a line segment that would extend out to touch the virtual object being pointed to. The visible directed ray-cast helped to learn how to correctly hold and direct the hand controller for a successful hit.

The interaction model “Raise-hand gesture” was used for discoverability of IoT devices, and half (nine) of the participants were able to locate and identify all of the smart devices quickly and without errors. Some users mistook the loudspeakers for IoT objects, while others missed the pot plants or lamps. However, the gleam of light from the objects was correctly interpreted, and the raise-hand gesture was easily remembered. At the end of the test, all participants except for one recalled the gesture without any further instructions.

Cognitive and physical load

The overall RTLX scores generated relatively low average values (see Table 2), and a one-way ANOVA for dependent measures showed a significant relation: $F(2,51) = 3.40$, $p = .041$. Multiple pairwise-comparison showed significant difference between the reported RTLX_Point and RTLX_gaze_voice with an adjusted p-value of $p = 0.034$ (see Fig. 7).

Table 2. The RTLX scores for the different interaction models. Means and standard deviations.

	RTLX button	RTLX point	RTLX gaze voice
Mean	12.28	14.39	8.50
SD	5.65	10.02	3.01

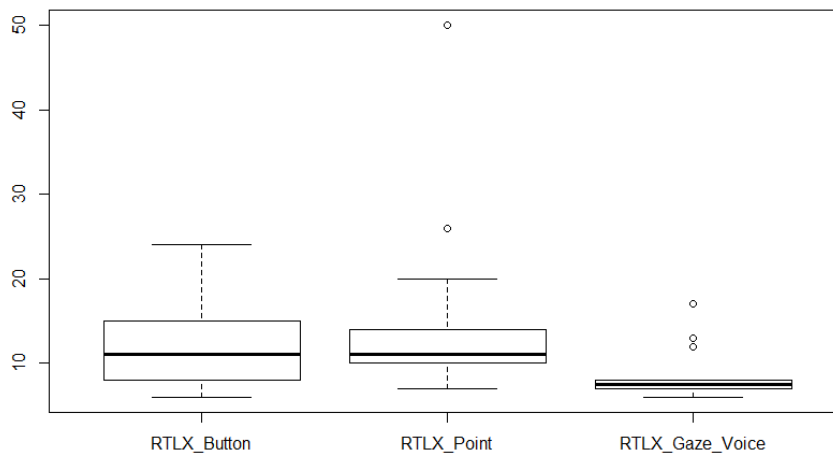


Fig. 4. The RTLX scores illustrated in a boxplot.

Preference

When it comes to the individual rankings, the differences between the interaction models were similar to the perceived workload, the same interaction type were preferred as the one which was perceived as the lowest workload (head-gaze and voice) (see Table 3). A one-way ANOVA for dependent measures showed a significant relation between the reported “ranked interaction-value” (RINT-value) and the corresponding interactive pattern: $F(2,51) = 7.4, p = .001$. Multiple pairwise-comparison showed significant difference between RINT_gaze_voice and RINT_button with an adjusted p-value of $p = .001$. Moreover, it was close to the margin of statistical significance between RINT_Point and RINT_gaze_voice with an adjusted p-value of $p = .072$. The interaction type with voice and head-gaze were preferred prior to the traditional pressing on switch buttons (see Fig. 5). However, in the free interaction phase, the main part of the participants (11 of 18) actually used pointing as their main model of interaction, even though this was not always easily performed. Six of these 11 participants also ranked pointing as their primary choice, while five of them stated head-gaze and voice as their main preference. This inconsistency indicates a curiosity for testing new or challenging interaction models. The result equally implicates a need for complementary evaluation methods when prototyping unfamiliar IoT-environments.

Table 3. Order of priority (values of 1, 2 or 3) for the different interaction models. Means and standard deviations (note that high values correspond to high rankings, and vice versa).

	RINT button	RINT point	RINT gaze voice
Mean	1.6	1.9	2.5
SD	0.62	0.80	0.79

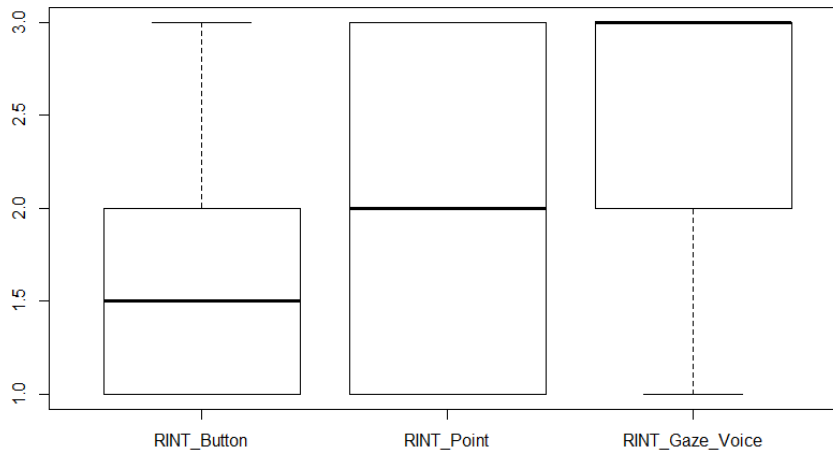


Fig. 5. Preferred interaction models in the free interaction phase.

Presence

The results obtained from the SUSPQ present a total mean value of $M = 5.4$ ($SD = .98$). Since the maximum value is 7 for all of the 6 questions, this has to be considered as relatively high rating (the actual usefulness and validity of the SUSPQ measure has, however, been a subject for discussion [31]). To analyze if the user experience could depend on the previous use of VR equipment, the group of participants was divided into VR novices (completely novice and with no experience of VR) and VR veterans (with at least one experience of virtual worlds). The resulting mean values were $M = 6.0$ ($SD = .87$) and $M = 4.9$ ($SD = .78$) respectively (see Fig. 6).

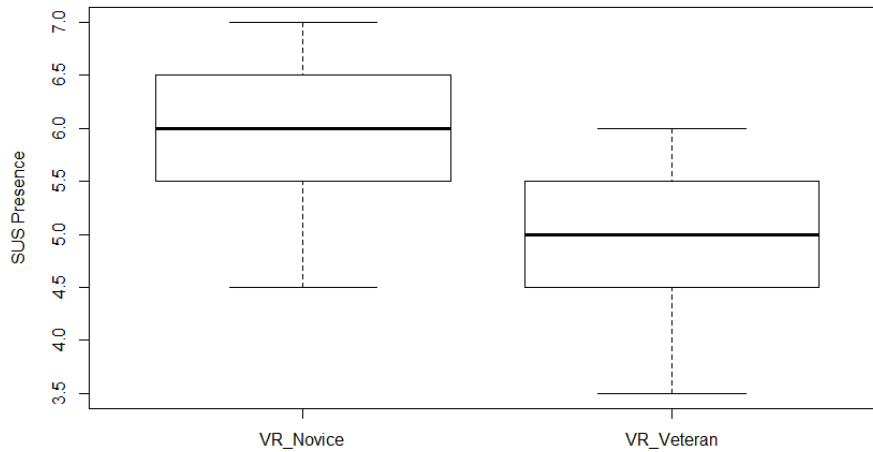


Fig. 6. SUSPQ for different users.

A T-test verified that this difference between the groups were significant: $t(8) = 2.49, p = .038$.

Qualitative findings

On the question, “What did you experience as hardest in the entire test?” The majority of the participants replied “To point at distant objects.” Other responses referred to feedback or interactive possibilities in general (see Table 4). The easiest aspect of the test was, according to nine of the participants, to use head-gaze and voice as interaction pattern (not surprisingly since this never failed due to the WOZ-solution). For other responses regarding this question, see Table 5.

Table 4. Responses regarding difficulties in the test.

What did you experience as hardest in the entire test?	Number of participants
Point at distant objects	10
Match switch buttons to lamps	3
Understand interactive possibilities	2
Notice feedback from smartwatch	1
Understand depth	1
Nothing (everything was easy)	1

Table 5. Responses regarding what was easy in the test.

What did you experience as easiest in the entire test?	Number of participants
To use head-gaze and voice as interactive pattern	10
Everything was quite easy	3
Point at TV or lamps	2
To use the switches for turning on lamps	2
Hands-up-gesture	1

The feedback from the smartwatch was generally quite well understood, although several of the participants forgot to look at the watch when interacting with the pot plants. One reason for this could be the lack of feedback from the plants themselves. Since these did not confirm the interactive gesture by changing their visible state (similar to the TV and lamps), several participants got confused and interpreted their interaction as unsuccessful. Another consideration is that the participants did not expect the plants to be smart at all, and seven participants named the smartness and feedback from the pot plants as the least expected element during the test (see Table 6). On the question, “Was there any moment that surprised you?” some participants replied that traditional interaction with switches and haptic feedback was not expected, while other expressed their astonishment over the successful interaction in general.

Table 6. Responses regarding surprising moments.

Surprising moments	Number of participants
The smartness of the flowers	4
The haptic feedback from the wall switches	3
That the interaction actually worked	2
Nothing was surprising	2
Smartwatch feedback	2
The test environment	2
No feedback when pointing correctly	1
Being able to use “normal switches	1
Turning on the TV	1

5 Discussion

As a whole, VR could be considered as an interesting and valuable tool for prototyping and evaluating IoT-interaction, mainly due to the immersive user experience and the possibility of evaluating non-existing interaction technologies with good ecological validity.

5.1 Comparative study

Overall, the RTLX scores of all three-interaction models were relatively low. The perceived workload of the head-gaze and voice had the lowest score and was significantly lower than the point and click. One can argue that it is not fair to compare head-gaze and voice with the other two interaction models. Since it was based on a WOZ solution. However, the commands were really simple “Turn on/off” and most of the available voice assistants such as Siri and Google Assistance can handle much more advanced commands and the participants were not aware that a human was operating the specific interaction model.

Usability evaluation in VR requires the participant to step out of the VE to answer questionnaires, or have the questionnaire available in the VE. If the virtual environment is equipped with sound effects it could of course also be hard to instruct or talk to the user during interaction. The experiences drawn from this study were that it was possible to state the NASA TLX questions orally during the virtual interaction. Moreover, statistically significant differences could be observed regarding the preferred interaction model, which was the same as the one having the lowest perceived workload i.e. head-gaze and voice. However, in order to evaluate and investigate users’ cognitive workload, the experiment would have benefited of having eye tracking integrated in the HMD. Zagerman et al. [32], for example, “encourage the use of eye tracking measurements to investigate users’ cognitive load while interacting with a system.” The fact that several manufacturers of eye tracking equipment, e.g. Tobii and SMI, currently are integrating their products in popular VR systems.

One of the major advantages with VR as an interactive environment is the fictive but realistic setting, and many users in our test did not only experience relatively high presence but also quickly accepted the hand controls as replacements for their own hands. However, current VR systems only allow relatively coarse hand gestures without the variation and nuance offered by finger gestures e.g. pinch to zoom.

5.2 VR limitations

As have been mentioned earlier, the physical space obtainable for the user was very limited in relation to the virtual space in the prototype. Even though the size of this area (9 m²) partly depended on practicalities in the VR laboratory environment, the discrepancies between these two spaces are one of the limitations with current room-scale VR systems. Not only could it be hard to simulate large IoT-environments, such as warehouses, parks, squares or entire buildings, it equally complicates the testing of spaces with a lot of objects. If virtual objects cannot be reached or explored from different angles, they risk concealing each other, which in turn hinders interaction. To facilitate for the user to explore a larger area, it would therefore be necessary to introduce some sort of locomotion technique. Today, the two most common locomotion techniques for VR headsets are “teleportation” and “trackpad-based locomotion.” Both have their pros and cons and the locomotion technique of choice should therefore depend on the specific IoT use case. However, it is important to note that adding an artificial locomotion upon the room-scale tracking potentially could give rise to

higher cognitive load due to a less natural user interface. Higher cognitive load in turn might impair the user's performance and render a task more difficult than it would have been in real life. An alternative to using locomotion techniques could be to exploit new VR locomotion hardware, such as the Cyberith Virtualizer [33], which facilitates walking in place locomotion.

5.3 VR as a prototyping tool

Results from a similar study by Alce et al. [20] suggested that using VR as a tool for IoT interaction has potential but that "several challenges remain before meaningful data can be produced in controlled experiments." Tracking was identified as one such challenge: the 3DOF tracking of the Oculus Rift DK1 used in that study limited the usefulness of the method. However, the HTC Vive with its sub-millimeter precision and 6DOF tracking constitutes a huge step in the right direction, which could be observed in this study. For example, in Alce et al. [20] study the authors were not able to find any statistically significant differences between the evaluated interaction concepts while in this study we could. However, this should be compared with similar IoT interactions in with physical prototypes to provide more evidence of using VR for prototype IoT interactions.

One can ask, why not use augmented reality (AR) glasses, which combines virtual and real objects such as Microsoft HoloLens or Magic Leap to evaluate prototypes? Alce et al. [34] developed three basic AR interaction models that focused on similar aspects as this paper, discovering and selecting devices. However, the current AR glasses comes with even more limitations, for instance, the field of view is very small compared with VR headsets, the interaction is very limited, and you need to track and detect things in order to make e.g. plants smart. The VR headsets are more mature in these cases, but the vast development of AR glasses will solve some of the limitations in the near future.

Finally, when evaluating prototypes in VR, it is important to remember the discrepancy between real and virtual interaction. Obviously, it is the virtual interaction that is evaluated, and the transfer of this to reality will most certainly not be perfect. Furthermore, if the users are not immersed enough, it could very well be that similarities between interactions are enhanced (that is to say, all interactions in the virtual environment feel equally awkward or strange), whereupon any differences become concealed and unnoticed.

6 Conclusion

This paper used VR to prototype IoT interaction in a smart home environment. Three IoT interaction concepts were compared in a controlled experiment. The results showed that statistically significant differences and subjective preferences could be observed. The participants preferred the combination of head-gaze and voice. Additionally, this study implies that VR has the potential to become a useful prototyping tool to explore IoT interaction for a smart home environment.

References

1. Davies, R. C. Applications of Systems Design Using Virtual Environments. In K. M. Stanney (Ed.), *Handbook of Virtual Environments - Design, Implementation, and Applications* (pp. 1079–1100). Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey (2002).
2. Lei, Z., Yue, S., Yu, C., Yuanchun, S. SHSim: An OSGI-based smart home simulator. In *Ubi-media Computing (U-Media)*, 2010 3rd IEEE International Conference on IEEE, pp. 87–90, IEEE (2010).
3. Seo, D. W., Kim, H., Kim, J. S., Lee, J. Y. Hybrid reality-based user experience and evaluation of a context-aware smart home. *Computers in Industry*, 76, 11–23 (2016).
4. Rogers, Y., Sharp, H., & Preece, J. *Interaction Design - beyond human-computer interaction* (Third Edition). A John Wiley and Sons, Ltd, Publication (2011).
5. Samsung SmartThings Homepage, <https://www.smarthings.com/>, last accessed 2019/01/15.
6. Apple HomeKit, Homepage, <https://www.apple.com/ios/home/>, last accessed 2019/01/15.
7. Weave homepage: <https://openweave.io/>, last accessed, 2019/01/15.
8. HTC Vive Homepage, <https://www.vive.com/eu/>, last accessed 2019/01/15.
9. Unity Technologies homepage, <http://unity3d.com/>, last accessed 2017/08/25.
10. Barton, J. J., Vijayaraghavan, V. UBIWISE, a simulator for ubiquitous computing systems design. Hewlett-Packard Laboratories Palo Alto, â AI HPL-2003-93, (2003).
11. O'Neill, E., Klepal, M., Lewis, D., O'Donnell, T., O'Sullivan, D., Pesch, D. A testbed for evaluating human interaction with ubiquitous computing environments. In *Testbeds and Research Infrastructures for the Development of Networks and Communities, 2005. Tridentcom 2005. First International Conference on IEEE*, pp. 60–69. IEEE (2005).
12. Nishikawa, H., Yamamoto, S., Tamai, M., Nishigaki, K., Kitani, T., Shibata, N., Ito, M. UbiREAL: realistic smartspace simulator for systematic testing. In *International Conference on Ubiquitous Computing*, pp. 459–476, Springer, Heidelberg (2006).
13. Furfaro, A., Argento, L., Parise, A., Piccolo, A. Using virtual environments for the assessment of cybersecurity issues in IoT scenarios. *Simulation Modelling Practice and Theory*, vol. 73, pp. 43-54, (2017).
14. Van Nguyen, T., Nguyen, H., & Choi, D. Development of a context aware virtual smart home simulator. *arXiv Preprint arXiv:1007.1274* (2010).
15. Park, J., Moon, M., Hwang, S., Yeom, K. CASS: a context-aware simulation system for smart home. In *Software Engineering Research, Management & Applications, 2007. SERA 2007. 5th ACIS International Conference on IEEE*, pp. 461–467, IEEE (2007).
16. Armac, I., Retkowitz, D. Simulation of smart environments. In *Pervasive Services, IEEE International Conference on IEEE*, pp. 322–331, IEEE (2007).
17. Hu, W., Zhou, H., Lin, C., Chen, X., Chen, Z., Lu, Y. Design of web-based Smart Home with 3D virtual reality interface. In *Control (CONTROL), 2012 UKACC International Conference on IEEE* pp. 223–228, IEEE (2012).
18. Bowman, D. A., Wingrave, C. A. Design and Evaluation of Menu Systems for Immersive Virtual Environments. In *Proceedings of the Virtual Reality 2001 Conference (VR'01)*, pp. 149–156, IEEE (2001).
19. de Sá, A. G., and Zachmann, G., Integrating virtual reality for virtual prototyping. *Proceedings of DETC'98*, pp. 1-12, Atlanta, Georgia. (1998).
20. Alce, G., Hermodsson, K., Wallergård, M., Thern, L., Hadzovic, T. A Prototyping Method to Simulate Wearable Augmented Reality Interaction in a Virtual Environment-A Pilot Study. *International Journal of Virtual Worlds and Human Computer Interaction*, 3, 18–28 (2015).

21. Ens, B., Anderson, F., Grossman, T., Annett, M., Irani, P., & Fitzmaurice, G., Ivy: Exploring spatially situated visual programming for authoring and understanding intelligent environments. In Proceedings of the 43rd Graphics Interface Conference, pp. 156-162, Canadian Human-Computer Communications Society. (2017).
22. Carter, S., Mankoff, J., Klemmer, S., Matthews, T. Exiting the Cleanroom: On Ecological Validity and Ubiquitous Computing. *Human-Computer Interaction*, 23(1), 47–99, (2008).
23. Gombac, B., Zemljak, M., Širol, P., Deželjin, D., Copic Pucihar, K., Kljun, M. Wizard of Oz experiment for Prototyping Multimodal Interfaces in Virtual Reality. (2016).
24. Barnum, C. M. Usability testing essentials: ready, set... test! Elsevier (2010).
25. Nielsen, J., Pernice, K. How to conduct eyetracking studies. Nielsen Norman Group, Fremont, CA. (2009).
26. Hertzum, M., Hansen, K. D., Andersen, H. H. K. Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28(2), 165–181 (2009).
27. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., de Weijer, J. Eye tracking: A comprehensive guide to methods and measures. OUP Oxford (2011).
28. Ledo, D., Greenberg, S., Marquardt, N., & Boring, S. Proxemic-Aware Controls: Designing Remote Controls for Ubiquitous Computing Ecologies. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services pp. 187–198, ACM (2015).
29. Norman, D. A., Nielsen, J. Gestural interfaces: a step backward in usability. *Interactions*, 17(5), 46–49 (2010).
30. Hart, S. NASA-task load index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2006). Homepage, <http://pro.sagepub.com/content/50/9/904.short>, last accessed 2019/01/15.
31. Usoh, M., Catena, E., Arman, S., & Slater, M. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5), 497–503 (2000).
32. Zagermann, J., Pfeil, U., Reiterer, H. Measuring cognitive load using eye tracking technology in visual computing. Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization. pp. 78-85. ACM (2016).
33. Cyberith Virtualizer Homepage, <https://www.cyberith.com/>, last accessed 2017/11/30.
34. Alce, G., Roszko, M., Edlund, H., Olsson, S., Svedberg, J., & Wallergård, M. [POSTER] AR as a User Interface for The Internet of Things—Comparing Three Interaction Models. In 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct) (pp. 81-86). IEEE (2017).