A Technical User Study on an Authoring Tool for Simplifying VR Study Setups in HRI Research

André Helgert
Institute of Computer Science
Ruhr West University of Applied Sciences
Bottrop, Germany
andre.helgert@hs-ruhrwest.de

Sabrina C. Eimler
Institute of Computer Science
Ruhr West University of Applied Sciences
Bottrop, Germany
sabrina.eimler@hs-ruhrwest.de

Tom Gross
Human-Computer Interaction Group
University of Bamberg
Bamberg, Germany
hci@uni-bamberg.de

***Abstract*—This paper presents a novel virtual reality (VR) authoring tool designed to simplify the creation of human-robot interaction (HRI) studies. The tool was developed to lower the technical barriers in VR development and allows researchers to simulate complex robot interactions without the need for advanced programming skills. It incorporates key HRI methods such as the Wizard of Oz (WoZ) technique, eye-tracking and motion tracking, and provides a comprehensive platform for data collection and study customization. A technical user study with computer scientists was carried out to evaluate the technical applicability of the tool. The participants confirmed the user-friendliness, flexibility and accessibility of the tool and furthermore found the tool efficient and expressed a strong interest in using it for future research. The results underline the tool's potential to expand the use of VR in HRI research, especially for non-technical researchers.**

***Index Terms*—authoring tool, virtual reality, human-robot interaction, study setup, wizard-of-oz**

I. INTRODUCTION

Virtual reality (VR) is a technology that allows users to immerse themselves in and interact with virtual environments. In the context of human-robot interaction (HRI) research, VR offers the possibility to create virtual clones of robots that serve as tools to study and simulate interactions between humans and robots. By replicating the actual functional capabilities of physical robots in virtual environments, VR offers researchers a flexible and cost-effective platform for experiments. In addition, the digital clones can be adapted in terms of their appearance, range of functions and behavior towards humans without having to adhere to physical boundaries or technical limitations of physical robots. This capacity for testing advanced features in a virtual environment speeds up development, as it eliminates the need for physical prototypes, enabling faster iteration and feedback cycles [1].

In addition to these technical advantages, VR also provides a reliable platform for studying social dynamics between humans and robots. Research has shown that human behavior in virtual environments closely mirrors real-world behavior [2], [3], ensuring the validity of data collected in VR simulations [4]. This makes VR especially valuable for investigating human responses and interactions with robots, providing insights into social dynamics and behavioral patterns that can be observed in controlled virtual settings [5].

Given this unique combination of technical flexibility and behavioral authenticity, VR emerges as a powerful tool for HRI research. The ability to test advanced robotic features while simultaneously studying human behavior makes it an ideal medium for a wide range of study applications. The advantages that VR offers as a research tool for the HRI community make it clear that this medium could become the method of choice for many future investigations.

However, despite its immense potential, the widespread adoption of VR in HRI research faces several challenges. While the benefits are evident, not all researchers have the necessary resources or technical expertise to fully utilize these advantages. The development of VR-based studies can be both time-consuming and costly [6]. Although some existing systems aim to streamline this process with modular and adaptable solutions [7]–[10], there is still no comprehensive tool that makes VR-based HRI studies easily accessible. Overcoming these barriers could make the field more sustainable, reducing the reliance on expensive physical robots.

Recognizing these challenges, we developed a VR authoring tool to optimize the creation of VR-based studies, offering researchers a simplified and modular approach. We collected the requirements, needs, and preferences of HRI researchers, after which the tool was designed and developed into a functional system [11]. Following the requirements of HRI researchers, the tool provides ready-to-use virtual environments, robots, measurement methods, and interaction options, enabling users to customize studies to meet their specific research needs. It also integrates key technologies and paradigms, including eye-tracking for data collection and a visual programming interface for easy robot behavior design. Additionally, it features a Wizard of Oz (WoZ) dashboard, a widely adopted method in HRI research. This paper conceptually outlines the integrated techniques and assesses the tool with input from computer science (CS) experts, to evaluate the applicability of the system from a technical side. In this user study, participants used the tool to create a study and then evaluated its usability, workflow, UI/UX, and other aspects through a qualitative interview. Since the usability of such a tool is the fundamental prerequisite for its acceptance, this user study serves as a first step in refining the tool, ensuring it meets standards for intuitive design and user-friendliness.

II. Related Work

VR is a widely used research tool in various disciplines. For instance, it is employed in medicine as a training tool [12], in educational settings to make learning more immersive and efficient [13], and for raising awareness about social issues, using interactive elements to engage people more effectively [14]. Specifically in the field of Human-Robot Interaction, VR applications range from virtual environments for training and simulations that would be dangerous in real life [15], to studies focusing on social interactions between humans and robots [16]. Recent studies also highlight VR's potential for robot teleoperation [17] and exploring human-robot collaboration in safe virtual spaces [18].

The widespread and diverse use of VR as a research tool can be attributed to its numerous advantages. Virtual environments allow researchers to immerse participants in simulations that create a sense of presence (e.g., ”being there”). Presence and immersion have been identified by Rebelo et al. [19] as key concepts for understanding the psychological and physical experiences of participants. These conditions make VR-based evaluations suitable for real-time assessments of clinical, affective, and social processing in a manner that closely resembles real-world conditions [20].

Additionally, other factors, such as the ability to maintain control over the experimental setting at all times [21], [22], are important selling points for VR as a research tool. Unlike field and laboratory studies, VR offers an ecologically valid platform because it enables studies to be conducted in realistic environments ”on-site” while providing a reproducible and controllable setting [21], [23], [24].

Furthermore, VR can integrate behavioral and physiological measures to track users' responses. It allows for the measurement of user behavior during the VR-study, e.g. via motion tracking systems, such as eye-tracking or head-tracking devices. These devices can monitor head and eye movements, facilitating the analysis of non-verbal cues of engagement [25]. In addition, physiological measurements can be monitored in a virtual environment through the VR-headset, which tracks heart rate variability, skin conductance, and skin temperature in real time during the virtual experience. Thus, VR can be considered an integrated measurement tool, offering a more ecological and controlled environment that enables the development of a more holistic and comprehensive psychological model of user experiences [6], [26], [27].

When looking at similar systems that aim to simplify complex processes for the VR universe, there are a few examples. Giglioli et al. have presented a framework capable of collecting psychological attributes in real-time during VR scenarios [7]. Brookes et al. have presented UXF, a system that offers various forms of data collection and can be easily adapted to independent variables, but which does not offer any stimulus material [8]. There are also attempts, e.g. by Vasser et al., to simplify the generation of 3D worlds through modularized functions [10]. This may be sufficient for many use cases, but especially for HRI it is important to provide

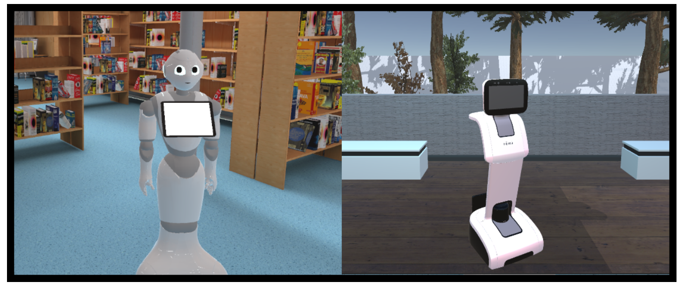

Fig. 1: Left: Pepper in library; Right: Temi in museum

realistic robots and environments. All of these systems are useful and legitimate attempts to simplify complex technical processes in VR development for studies. However, they are not specialized in the HRI area and therefore do not offer robot configuration options and various prefabricated, high-quality environments, nor do they offer measurement methods that are required for the HRI area.

III. VR Authoring Tool

We developed the VR authoring tool using the Unity game engine, incorporating various virtual robots, objects, and environments, which were used in previous iterations and projects [11], [28]. The authoring tool itself features a modular architecture, designed with a high degree of technical generalization to maximize customizability for a wide range of use cases. The version presented here is a standalone application that can be executed on any computer, eliminating the need for researchers to work with Unity directly, thus lowering technical barriers. This chapter briefly presents the initial needs and requirements of HRI researchers, which were collected in a previous iteration and developed into a catalog of functions [11]. It then describes the integrated techniques and the conceptual workflow of the VR authoring tool and shows how researchers can use it to create studies. In addition, this chapter explains the novelty and significance of the system for HRI research.

A. Needs and Requirements for the System

To establish the VR authoring tool in HRI research, it is crucial to identify the needs and requirements of HRI researchers for such a system. This approach ensures that the tool provides the necessary functions required by the HRI field. Since HRI is inherently interdisciplinary, encompassing researchers from technical disciplines to those from psychological backgrounds, special care was taken in the initial requirements gathering to include feedback from HRI researchers who work less within the technical spectrum. This approach aims to ensure that the VR authoring tool is accessible to a wide range of HRI researchers with varying technical expertise, especially given that the development of VR applications for studies is a complex process. The key needs and requirements identified by HRI researchers focused on a modular design for the tool, enabling its functionalities to be applied across various use cases. At the same time, these functionalities should be as simple as possible, allowing researchers to navigate the tool with ease. These needs emphasize the necessary flexibility and simplification of the tool. Additionally, HRI researchers expressed the importance of offering a simplified “Beginner“ version that, while limited in functionality, empowers researchers without implementation expertise to use the tool for their own research. Alongside this, they suggested the inclusion of an “Expert“ version without functional restrictions, allowing advanced users to leverage the tool’s full capabilities. This dual approach aims to accommodate researchers with varying levels of technical knowledge. Many HRI researchers also highlighted the need for paradigms that simplify human-robot interaction. As an example, some researchers suggested a block-based development of the robot interaction schema to streamline the process. They also expressed the wish for suitable documentation in video form, or demos. Further results regarding the perception of VR in HRI research and the measurement methods available in VR can be found in the previous iteration of the study [11].

B. Synthesis of Existing Techniques to the System

The VR authoring tool integrates established techniques and paradigms into a unified platform, making these resources easily accessible for researchers. Virtual versions of real robots, such as Softbank Robotics’ Pepper and Temi, have been digitized for use in VR, with functionalities mirroring their physical counterparts (see Figure 1 for the digital clones of Pepper and Temi in two different virtual environments). Customization options, such as adjusting robot colors, are available directly within the VR environment. For robot control, the system incorporates the Wizard of Oz (WoZ) technique, which is highly valued in HRI research for its flexibility [29], cost-effectiveness [30], and ability to bypass the need for fully autonomous robot functions [31]. In WoZ experiments, a hidden human operator, or “Wizard“, remotely controls the robot’s movements and speech, while participants remain unaware of the human intervention. The WoZ has already been presented in an earlier iteration [32]. Additionally, a visual programming interface inspired by tools like Scratch [33] and Blockly [34] has been implemented, enabling users to design robot behavior schemes with ease. This interface uses color-coded blocks, each representing pre-programmed functions, simplifying the creation of interaction patterns. The visual programming interface offers an alternative to the WoZ approach for controlling the robot’s interactions within the VR environment. The tool also accommodates established measurement methods, including eye-tracking, motion tracking, physiological data collection, and self-report tools such as questionnaires. These methods, widely used in HRI research, come pre-implemented for immediate use, streamlining data collection and analysis. For example, eye-tracking metrics like dwell time and fixation count can be applied to specific objects in the virtual environment, such as the robot, to capture detailed user interaction data. In motion tracking, the tool focuses on proxemic metrics, such as measuring the distance between the human and the robot, providing valuable insights into human-robot interaction dynamics. The advantage of these integrated measurement tools is that they are ready-to-use, eliminating the need for laborious implementation.

C. Conceptional Workflow

When launching the standalone VR authoring tool, the main menu is displayed. The main menu consists of the following options: starting previously created studies for data collection with participants, creating a new study, configuring an existing study, and providing help through demos, FAQs, and written documentation.

When creating a new study, various configurations must be defined, which follow the study creation process in HRI based on Bartneck’s method [35], a widely recognized approach in HRI research. These configurations include specifying whether the interaction is dyadic or group-based, selecting the robot (see Figure 2), choosing the virtual environment (currently library, museum, or industrial settings), deciding on the interaction mode (WoZ vs. visual programming interface), and selecting the preferred measurement methods.

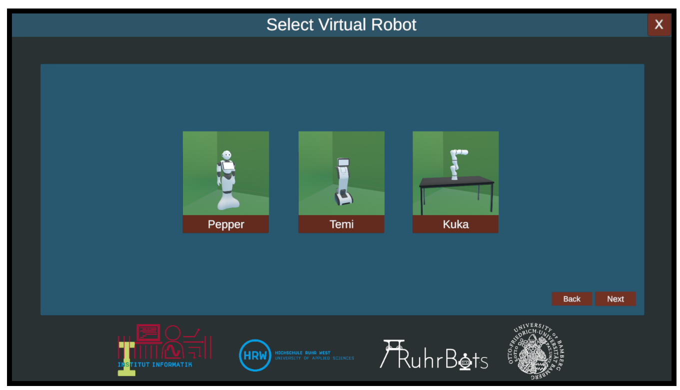

Fig. 2: Selection of the robot in the study creation process

Once all these steps are completed, the virtual environment is loaded with the specified settings and visualized in a new scene (see Figure 3 for an example of the visual programming interface). This environment reflects the configurations chosen during the study creation process. In the example shown in Figure 3, the Pepper robot is selected for a dyadic interaction in a library setting. Alongside the visual representation of the environment, users can toggle between Robot point of view (POV) and Player POV to view the study environment from different perspectives. Additionally, users can navigate the virtual environment using keyboard and mouse controls, as well as move and rotate the robot. The menu also offers options to customize the robot’s appearance, adjust participant metrics (e.g., the height of virtual participants and different locomotion mechanics), create interactions, modify or reselect measurement methods, and configure data export, which is available in various formats.

When creating an interaction, the remaining menu options disappear, and an interface appears where interaction steps can be defined sequentially (see Figure 4 on the left side). Interactions can be categorized into robot navigation, tablet

Fig. 3: Menu of the visual programming interface

content display, voice output, and robot gestures. These interactions can be enhanced with program logic elements such as wait, do-until, if-else, and repeat, allowing for more dynamic and responsive behavior patterns. In the example shown, two interaction steps have already been set. The interaction begins with a gesture where Pepper welcomes the participant, followed by a 4-second pause. Then, the robot delivers a verbal message to the participant. In Figure 4, the *movement* menu for the third interaction block is visible. Here, researchers can choose from predefined blocks for quick integration into the interaction flow, set navigation targets within the virtual environment for the robot, or define precise movements using the *move* and *turn* commands. Once the interaction creation process is complete, the study design can be saved as a .json file and initiated for data collection via the main menu.

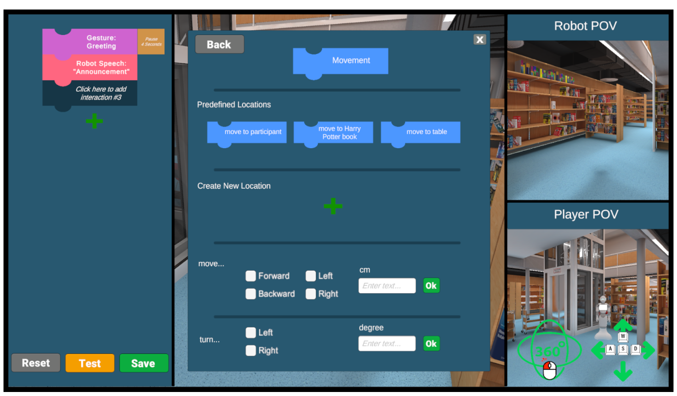

Fig. 4: Selection of navigation task for the robot

D. Novelty and Significance of the VR Authoring Tool

The VR authoring tool introduces several innovations aimed at addressing key challenges in HRI research, offering both technical advancements and practical usability. One of its main contributions is its potential to lower technical barriers by removing the need for in-depth programming or familiarity with complex VR development platforms like Unity. By providing a standalone application, the tool enables researchers from various backgrounds to create, configure, and conduct VR-based studies without requiring advanced technical expertise. This could make the research process more simple, encouraging broader adoption and experimentation with VR within the HRI community.

IV. Technical User Study

The VR authoring tool was designed to address the specific requirements of HRI researchers, culminating in its first functional iteration. For this technical user study, we utilized the "Beginner“ version, as we expect it to be the preferred choice for non-technical researchers due to its simplified design. To validate the tool's usability, workflow, UI/UX design, and overall simplicity, a technical user study involving computer science researchers was conducted to gather expert feedback and ensure it meets the necessary standards.

A. Sample

We were able to recruit a total of nine participants (two women and seven men) who worked as research associates and PhD students in the CS field. They had an average age of 33.00 years (SD = 11.05). Participants had to rate on a 7-point Likert scale (7 = very high; 1 = very low) their experience in programming (M = 5.11, SD = 0.99), UI/UX design (M = 3.44, SD = 1.64) and the development of technical systems (M = 4.67, SD = 1.33). The number of participants was sufficient for the technical user study, as data saturation could be determined. We found this by continuously analyzing the collected data after each session, which generated fewer and fewer new insights, if any at all.

B. Method

This technical user study was conducted in two parts. In the first part, a Think Aloud protocol was employed while participants used the authoring tool, with screen recording capturing their interactions. Participants were instructed to verbalize their thoughts and perceptions of the tool throughout the session, which allowed for the identification of both intuitive aspects of the system and potential issues that may arise unconsciously. During the process, the study leader asked clarifying questions to gain more insight into the participants' remarks and observations. The second part consisted of a semi-structured interview divided into six sections. All of the participants' responses were collected qualitatively.

The first section focused on **initial impressions**, where participants were asked two questions about their first impressions of the main menu and the visual programming interface. To aid their recall, images of the respective interfaces were presented on paper.

As **usability** is a dominant factor in Human-Computer Interaction (HCI) [36], the second section addressed this aspect based on the guidelines of the ISO 9241-11:2018 standard [37], which defines usability in terms of *effectiveness*, *efficiency*, and *satisfaction*. For *effectiveness*, participants were asked how well the tool's features supported the creation of the study. Additionally, the completion of the study process was collected via screen recording. For *efficiency*, participants were questioned about the perceived mental and physical effort required to design a study using the tool. For *satisfaction*,

Fig. 5: Example interaction created with the visual programming interface

participants were asked how enjoyable or frustrating they found the system.

The third section focused on **workflow factors**, drawing inspiration from Mayhew's task analysis methods [38]. Although Mayhew does not provide a universal method for interviewing users about workflows, his approach incorporates several key heuristics and strategies that can be tailored to effectively gather workflow information during an interview. Participants were first asked to describe the steps they followed to complete their research project using the tool. Follow-up questions asked how clear and logical they found the workflow, whether they always knew what the next step was, and whether any steps were confusing or difficult.

The fourth section dealt with **UI/UX factors**, with questions informed by Norman and Krug's principles [39], [40]. Here, participants were asked how easy it was to navigate the interface, whether any elements were unclear or failed to meet their expectations, how intuitive they found the system's navigation, and how consistent the design and functionality of the user interface appeared to be.

The fifth section focused on **simplicity**, asking participants how accessible they thought the VR authoring tool would be for individuals with little technical or programming knowledge who aim to create their own VR studies in HRI. Additionally, potential challenges or issues faced by this target group were discussed.

Finally, the sixth section covered **general feedback**, where participants were asked to highlight particularly positive aspects of the tool, suggest improvements, identify other research areas where a VR authoring tool could be beneficial, and express whether they would personally use or recommend the tool.

C. *Material*

Participants in the technical user study were tasked with setting up a study using the VR authoring tool. They were given the research question: *"What distance should a robot keep when approaching a person?"* The framework specified using the robot Pepper in a virtual library setting, with an automated interaction. The robot had to announce its movement, approach the participant, and stop when the participant pressed a key, indicating the optimal distance. Pepper would then verbally indicate it had stopped and say goodbye. These instructions were also provided on a printed sheet for reference.

D. *Procedure*

Participants were informed about data protection and the study's purpose, and consent was obtained for audio recording and data use. A screen recording captured their process of using the authoring tool on a desktop PC to configure the study, select the robot, and choose a measurement method. After setting up the interaction using the visual programming interface, the study concluded with a semi-structured interview.

V. RESULTS

A. *Simplicity of the VR Authoring Tool*

When asked how simple the VR authoring tool appears for researchers with little technical implementation expertise, all CS experts said that the tool is **likely very accessible**. P4 commented: *"I believe that if you work in the research* *field, you generally have some technical aptitude, and you're comfortable with computers and technical devices. So, it is quite intuitive."* Similarly, P6 said: *"Yes, very. I didn't find any aspect where my programming knowledge was particularly necessary."* P1 added that the simplicity of the tool could be improved by including barrier-free features such as text-to-speech.

When looking at potential challenges from a technical perspective, problems with **bugs** were mentioned. P4 noted: *"I think that if bugs occur, or if help is needed, it may not be immediately clear or obvious what to do"*. Additionally, the **workflow logic of the interaction blocks**, which is based on common programming conventions, was brought up as a potential challenge. P7 remarked: *"I'm not sure if the logic might be an issue—whether following the sequence could be a problem."*. It was also mentioned that a **lack of assistance** could pose challenges when using the tool, e.g. by P9: *"So I do think that there should be something like help again. The first thing that came to my mind was the Word paperclip, which existed at some point, but which I found rather annoying as I got to grips with Word. But perhaps it would be quite helpful at the beginning of trying out or using the software or the interface. So an introduction or training wouldn't be a bad idea"*.

B. General Perception of the VR Authoring Tool

Regarding the most positively mentioned aspects of the authoring tool, P1 highlighted its **ease of use**: *"Overall, the simplicity of handling it, the fact that it provides a very simple and quick way to implement such scenarios, especially considering my personal experience of how much time and effort goes into programming Pepper or VR systems. This is definitely the standout feature for me."* Similarly, P2 praised the tool for offering a **straightforward way to assemble workflows** from individual components.

Furthermore, P3 and P9 found the **display of the experimental sequence** on the left side of the visual programming interface helpful. P9 commented: *"What I particularly liked about the system was the layout of the menu on the first and subsequent pages. It made sense to me and felt very intuitive, having the menu on the left side as well as the sequential flow using arrows. That was very good."* P4 emphasized the **ability to test the created interaction**: *"The feature that stood out to me was the ability to try out what I had just created. Running through the workflow I had built to see what it looks like in action was very helpful."*

The **clarity in creating interactions** was also mentioned by P5 and P6. P5 said: *"Yes, the clarity was remarkable. Whether I wanted to create something speech-based or program a robot's movement, it was clear what I needed to do."* P7 focused on the **visualization aspect**: *"I really liked the visualization, seeing Pepper in the room and being able to stop the robot myself."*

When it came to suggestions for improvement, participants mentioned two key areas. First, several participants expressed a **desire for more information on specific features**, such as symbol icons. P4 suggested: *"What comes to mind would be an info box for each section, where you can see the documentation for that part without having to go back and read through everything, figuring out where you are. Just having a little icon for help on each page would be great."* Second, P8 asked for a **guided process for creating interactions**, echoing the same sentiment: *"It would be helpful to have an info field on each page, so you can access documentation specific to that part instead of going back and forth."*

Regarding whether participants would **want to use the tool themselves, all of them agreed**. P6 remarked: *"Yes, definitely. For future studies, it would certainly help streamline the setup."*

In addition, all participants would **recommend the tool** to HRI researchers. P1 stated: *"Absolutely. As I've mentioned before, I know how much of a pain this process can be. But I think this tool offers a great way to accomplish what you need in a performant and user-friendly manner. I would advise anyone without prior experience to give it a try, as it significantly simplifies the entry process."*

C. Usability of the VR Authoring Tool

Usability assessments were divided into three categories. For **effectiveness**, four participants indicated that they were able to complete the task effectively, thanks to **the tool's wide range of features and flexibility**. P1 commented, *"The fact that there was both the option to create custom blocks from scratch, with plenty of configuration options, as well as having certain presets that I could directly select, was helpful."* Another two participants stated that the tool's effectiveness was due to its **user-friendliness and ease of use**: *"I can do it myself as a layperson. I don't have to program a robot or modify hardware and install sensors to measure the distance"* (P3). However, three participants felt hindered in their effectiveness because there was **no option to access the tool's documentation** in the study setting. P4 mentioned: *"There was relatively little help. You had to figure things out for yourself. There wasn't a tool or tutorial to guide you through something simple."*

In terms of **efficiency**, eight participants reported that the system was efficient, with **minimal physical and mental effort required**. P3 stated: *"Minimal. I think once you get the hang of it, it's quite straightforward."* Similarly, P4 added: *"I didn't find it difficult. You could figure it out quite easily."*. In contrast to that, one participant stated, that the mental effort was higher to fulfill the task in the study because of the study environment itself and not because of usability problems regarding the tool (*"Well, I would say higher, because you are in the study and you want to be helpful and if you somehow don't know what exactly you have to do, then you get into a bit of a stress factor"*; P5).

Regarding **satisfaction**, 7 out of 9 participants said that using the system **was pleasant** and not frustrating. P7 remarked: *"I found it very pleasant. There was no point where I felt frustrated."* On the other hand, two participants were **not satisfied**. P3 expressed: *"It doesn't match the common* *structure I expect from such interfaces, like having the buttons closer together or seeing immediately how the elements relate to each other. For example, I wasn't sure how the timing between actions was connected after pressing the fall button."* P5 added, *"It became frustrating, especially when you didn't know what was happening during the robot interaction setup."*

Last but not least, with the help of the screen recordings, we checked whether the study objectives had been fully achieved by the participants. All of them except one were able to complete the task in the study using the authoring tool. In the case of the one person who did not succeed, the last interaction step with Pepper's verbal statement was missing. The participant explained the failure of the task itself with the subjective preference of how the elements are arranged in the tool (*"Possibly, but that is more of a subjective thing, that an arrangement would suit me better with another presentation"*; (P2)). Furthermore, we measured the times that the participants needed to complete this study. On average, they needed 9 minutes and 31 seconds to complete the task, with a standard deviation of 3 minutes and 38 seconds.

D. Workflow of the VR Authoring Tool

All participants found the workflow for creating the study in the evaluation with the VR authoring tool **to be logical and clear**. P7 commented: *"Overall, super clear. As mentioned, the different possibilities and pathways were very structured, logical, and easy to follow."* P4 added: *"It was very clear and logical. There were no problems understanding or implementing it."* When asked if they had a clear understanding of the next step in the creation process at all times, 5 participants agreed (*"It was always clear what the next step was";* P1). However, two participants mentioned that **due to arrangement of the elements**, they were not always sure what to do next. The arrangement of logic elements that could be added to the interaction blocks was presented in a single interface. As a result, participants expected to have to fill in all the logic elements, rather than choosing just one, as intended. P3 commented: *"That was the point where you helped me, where I had to press exactly on the logic implementation for it to be saved."* Additionally, there were **misunderstandings when adding interaction blocks**. P6 noted: *"For example, there was the step where you had to click on the interaction again to finalize it [when selecting the speech category]."* Furthermore, five out of nine participants **clicked the wrong item in the main menu** the first time. Intuitively, they clicked the first element ("Start Experiment"), although they were supposed to click the second element ("Create Study Design") to create the study. *"So, where exactly to click— I intuitively clicked the top left. The word 'Start' triggered me, and the top left also triggered me."* (P5).

E. UI/UX Design of the VR Authoring Tool

When asked how easily participants navigated the UI, two participants reported issues. P5 noted that certain elements were **more prominent than others**, and the UI, particularly in the logic selection, felt cluttered: *"So, the distribution was a bit strange. Some elements were just more prominent than what needed to be done. For instance, the interface or graphic of the library was much more prominent than the Modify-Robot button. Also, with the tiles, the green plus sign for adding was more noticeable than the tile itself because it was gray with white text. And then the tiles in the logic section, with the four outer options, felt a bit overloaded."* Similarly, P8 mentioned **not immediately finding their way around**, but only in the visual programming interface part: *"The first part here is easy because there are text and images, what you are expecting to have. In the next UI, it is not straightforward. It doesn't have introduction to what you are doing, what you start to do, and what you have to expect as a result, what you will do next. I mean there is no workflow until I ask you."* In contrast, the remaining seven participants reported **navigating the UI well** because of the clarity and familiarity of the instruction of the UI elements. P1 said: *"Neat and tidy. I found it very clear, especially for the beginning."* P9 added: *"I generally like that we have a menu on the left, which I believe is very familiar to many users, as you often find it on many websites, either on the left or at the top center."*

When asked how intuitive participants found the UI/UX, **all agreed that it was highly intuitive**. P1 praised the tool's flexibility: *"I liked that you could try things out without breaking anything. I could just look around, test things, and click on everything. I could click on any section without being forced to move forward immediately. It could have been the case that I said, Create Interaction, and then had to create a block right away or something like that."* P8 added that after **a short familiarization period, the UI/UX design seemed very intuitive**: *"The user interface itself was very intuitive. The camera guidance and control of Pepper, well, a bit simplified, but not bad. I would say after a short adjustment period, maybe 30 seconds where I tested everything, I was fully comfortable with the program and could carry out everything that was required."*

When asked how consistent the UI design was, **all participants agreed that it was consistent**. P9 mentioned the color palette used throughout the system: *"Very consistent, the colors look very simple. Along with the colors green, yellow, and red, I also remember blue and black. The color palette runs throughout the entire process. I didn't like the color of the robot, with black and yellow, but otherwise, the overall image was uniform."*

VI. Discussion

The VR authoring tool has been well received by CS experts. Although only the first functional iteration was used in this evaluation, the flexibility, range of functions and user-friendliness of the tool were emphasized, indicating that the needs and requirements were met. This is also reflected in the measured time for creating a VR study with the tool, which enabled participants to produce a functional version for a study in a relatively short period. CS experts rated the tool as suitable for researchers without programming knowledge. This indicates that the complex implementation processes for virtual robot interactions have been simplified to a near self-explanatory level. Such simplification could make the tool accessible to other research fields, serving as a valuable VR research instrument. This positive assessment by CS experts who are familiar with complex programming paradigms is particularly promising and indicates that the needs and requirements in terms of ease of use and the legitimization of the visual programming interface have been met. Participants recommended the tool for its ease and speed, appealing to both non-technical and CS researchers, suggesting broad interdisciplinary acceptance.

The majority found the tool's usability effective, efficient, and satisfactory, indicating it is on the right track with potential for broad use. The main concern was the lack of documentation, with participants requesting guides or a companion for support. This technical user study deliberately omitted such resources to capture raw impressions from CS experts. However, for the final release version, video tutorials, pre-built demo scenarios, and written documentation will be provided, as outlined in the requirements catalog in an earlier iteration [11].

The workflow received much praise as well. Participants particularly highlighted the logical sequence of configurations throughout the creation process, both in the main menu and the visual programming interface. This suggests that the guided sequence can be very supportive, allowing users to create a study with minimal mental and physical effort. Some participants raised comments about unclear elements in the workflow, such as the logic implementation overview, which was viewed as overly complex, or minor confusion regarding the tiles in the main menu. These are areas that can be improved in the next iteration to further enhance the tool's workflow.

Most participants found the UI/UX design easy to navigate due to its clear and consistent structure. Many reported that the interface was intuitive and posed no major challenges. However, some participants mentioned unclear element placement, such as the overly prominent scene featuring Pepper in the library, which distracted from the controls. One participant suggested a more guided approach, similar to the main menu's initial layout. While the original design aimed for an open-ended, sandbox-style interface to allow for experimentation, an optional guided version could be considered to enhance usability for some users, even if it might limit flexibility. In conclusion of this discussion it can be inferred that the current version of the tool has already gained positive feedback from technically adept researchers, affirming the usability of the concept. Since usability is a fundamental prerequisite for using such a tool, and this study has demonstrated that CS experts can effectively navigate it, we believe the tool is ready to be tested by non-technical researchers in the next phase. Nevertheless, results revealed that further adjustments to the UI/UX design and workflow could broaden the tool's applicability and better address the diverse needs of its users.

The evaluation and first iteration of the VR authoring tool have limitations. Visual irregularities, such as text overlapping interaction blocks and imperfect element arrangements, were noted by participants and will be corrected in the next iteration. Additionally, not all functions were included, as this would exceed the study's scope. A simple predefined research question was used to focus on usability, workflow, and UI/UX aspects. However, a more complex task might offer deeper insights into the tool's functionality. The Think Aloud method also presented challenges, as it can affect participant behavior [41] and not all participants verbalized their thoughts consistently. Not all needs and requirements raised in [11] could be tested in this technical user study. While CS experts praised the ease of use and simplified visual programming interface, the system's modularity could not be addressed due to the study's structured design. This aspect will be explored in a more open study with HRI researchers. Furthermore, the sample size of nine participants is limited. Involving more experts from a broader range of disciplines within CS could potentially provide more robust and comprehensive results.

VII. Conclusion & Future Work

A novel approach to utilizing VR in Human-Robot Interaction has been introduced through an authoring tool, designed to simplify the creation of VR studies in the HRI field and empower researchers without programming experience. The goal is to establish VR as a simplified research instrument in HRI. The system was evaluated by CS experts to gather feedback on usability, workflow, UI/UX design, and simplicity from a technical perspective, which is crucial for refining the tool before its deployment in the interdisciplinary field of HRI. Participants responded positively, highlighting its flexibility, functionality, and ease of use, though some inconsistencies in the UI/UX design remain and will be addressed in the next iteration. The next phase will focus on HRI researchers to assess how well the tool supports the intended user group, shifting the focus from technical feasibility to content-related aspects like robot interactions and measurement methods. We plan to use the revised version of the authoring tool, based on the insights and feedback from the CS experts, to provide non-technical researchers with an optimized and error-free version for the future study. Unlike the method described in this paper, participants will be encouraged to design studies that are more open-ended and aligned with their own research interests. The ultimate goal is to release the tool as open-source software for the community, with future enhancements including a digital companion to guide users, complementing written documentation, video tutorials, and demo scenarios to enhance guided usability.

Acknowledgements

This work was supported by the RuhrBots (www.ruhrbots.de) competence center (16SV8693) funded by the Federal Ministry of Education and Research Germany as well as internal PhD funding by the Ruhr West University of Applied Sciences. We would like to thank Carolin Straßmann, Karina Sturm and Jette Baltzer for their comments and contributions.

References

[1] E. Coronado, S. Itadera, and I. G. Ramirez-Alpizar, "Integrating virtual, mixed, and augmented reality to human–robot interaction applications using game engines: A brief review of accessible software tools and frameworks," *Applied Sciences*, vol. 13, no. 3, 2023, Article 1292.

[2] D. Gromer, O. Madeira, P. Gast, M. Nehfischer, M. Jost, M. Müller, A. Mühlberger, and P. Pauli, "Height simulation in a virtual reality cave system: Validity of fear responses and effects of an immersion manipulation," *Frontiers in Human Neuroscience*, vol. 12, 2018, Article 372.

[3] B. Renison, J. Ponsford, R. Testa, B. Richardson, and K. Brownfield, "The ecological and construct validity of a newly developed measure of executive function: the virtual library task," *Journal of the International Neuropsychological Society : JINS*, vol. 18, no. 3, pp. 440–450, 2012.

[4] J. Kisker, T. Gruber, and B. Schöne, "Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure," *Psychological Research*, vol. 85, no. 1, pp. 68–81, 2021.

[5] Y. Lei, Z. Su, and C. Cheng, "Virtual reality in human-robot interaction: Challenges and benefits," *Electronic Research Archive*, vol. 31, no. 5, pp. 2374–2408, 2023.

[6] T. D. Parsons, "Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences," *Frontiers in Human Neuroscience*, vol. 9, no. 660, 2015.

[7] I. A. Chicchi Giglioli, G. Pravettoni, D. L. Sutil Martín, E. Parra, and M. A. Raya, "A novel integrating virtual reality approach for the assessment of the attachment behavioral system," *Frontiers in Psychology*, vol. 8, 2017, Article 959.

[8] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq, "Studying human behavior with virtual reality: The unity experiment framework," *Behavior research methods*, vol. 52, no. 2, pp. 455–463, 2020.

[9] J. Grübel, R. Weibel, M. H. Jiang, C. Hölscher, D. A. Hackman, and V. R. Schinazi, "Eve: A framework for experiments in virtual environments," in *Spatial Cognition X*, ser. Lecture Notes in Computer Science, 2017, vol. 10523, pp. 159–176.

[10] M. Vasser, M. Kängsepp, M. Magomedkerimov, K. Kilvits, V. Stafinjak, T. Kivisik, R. Vicente, and J. Aru, "Vrex: an open-source toolbox for creating 3d virtual reality experiments," *BMC Psychology*, vol. 5, no. 4, 2017.

[11] A. Helgert, C. Straßmann, and S. C. Eimler, "Conceptualizing and designing a virtual reality authoring tool for human-robot interaction studies - learnings and guidance from expert interviews," *Mensch und Computer*, p. 331–341, 2024.

[12] L. I. Hanke, L. Vradelis, C. Boedecker, J. Griesinger, T. Demare, N. R. Lindemann, F. Huettl, V. Chheang, P. Saalfeld, N. Wachter, J. Wollstädter, M. Spranz, H. Lang, C. Hansen, and T. Huber, "Immersive virtual reality for interdisciplinary trauma management - initial evaluation of a training tool prototype," *BMC Medical Education*, vol. 24, no. 1, 2024, Article 769.

[13] L. Disch, S. Überreiter, and V. Pammer-Schindler, "Beyond textbooks: A study on supporting learning of molecule naming in a virtual reality environment," 2024, p. 310–322.

[14] A. Helgert, S. C. Eimler, and A. Arntz, "Stop catcalling - a virtual environment educating against street harassment," in *2021 International Conference on Advanced Learning Technologies (ICALT)*, 2021, pp. 419–421.

[15] S. B. i. Badia, P. A. Silva, D. Branco, A. Pinto, C. Carvalho, P. Menezes, J. Almeida, and A. Pilacinski, "Virtual reality for safe testing and development in collaborative robotics: Challenges and perspectives," *Electronics*, vol. 11, no. 1726, 2022.

[16] C. Straßmann, C. Eudenbach, A. Arntz, and S. C. Eimler, "Don't judge a book by its cover: Exploring discriminatory behavior in multi-user-robot interaction," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2024, pp. 1023–1027.

[17] L. Meng, J. Liu, W. Chai, J. Wang, and M. Q.-H. Meng, "Virtual reality based robot teleoperation via human-scene interaction," *Procedia Computer Science*, vol. 226, pp. 141–148, 2023.

[18] S. T. Mubarrat, A. Fernandes, and S. K. Chowdhury, "A physics-based virtual reality haptic system design and evaluation by simulating human-robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 54, no. 4, pp. 375–384, 2024.

[19] F. Rebelo, P. Noriega, E. Duarte, and M. Soares, "Using virtual reality to assess user experience," *Human Factors*, vol. 54, no. 6, pp. 964–982, 2012.

[20] C. M. Parsey and M. Schmitter-Edgecombe, "Applications of technology in neuropsychological assessment," *The Clinical Neuropsychologist*, vol. 27, no. 8, pp. 1328–1361, 2013.

[21] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. Mühlberger, "The impact of perception and presence on emotional reactions: a review of research in virtual reality," *Frontiers in Psychology*, vol. 6, no. 26, 2015.

[22] A. Gorini, E. Griez, A. Petrova, and G. Riva, "Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders," *Annals of General Psychiatry*, vol. 9, no. 30, 2010.

[23] A. Gorini, C. S. Capideville, G. de Leo, F. Mantovani, and G. Riva, "The role of immersion and narrative in mediated presence: the virtual hospital experience," *Cyberpsychology, behavior and social networking*, vol. 14, no. 3, pp. 99–105, 2011.

[24] Z. Campbell, K. K. Zakzanis, D. Jovanovski, S. Joordens, R. Mraz, and S. J. Graham, "Utilizing virtual reality to improve the ecological validity of clinical neuropsychology: an fmri case study elucidating the neural basis of planning by comparing the tower of london with a three-dimensional navigation task," *Applied neuropsychology*, vol. 16, no. 4, pp. 295–306, 2009.

[25] H. E. Yaremych and S. Persky, "Tracing physical behavior in virtual reality: A narrative review of applications to social psychology," *Journal of experimental social psychology*, vol. 85, 2019, Article 103845.

[26] C. J. Bohil, B. Alicea, and F. A. Biocca, "Virtual reality in neuroscience research and therapy," *Nature Reviews Neuroscience*, vol. 12, no. 12, pp. 752–762, 2011.

[27] M. Fusaro, G. Tieri, and S. Aglioti, "Seeing pain and pleasure on self and others: behavioural and psychophysiological reactivity in immersive virtual reality," *Journal of Neurophysiology*, vol. 116, pp. 2656–2662, 2016.

[28] A. Arntz, A. Helgert, C. Straßmann, and S. C. Eimler, "Enhancing human-robot interaction research by using a virtual reality lab approach," in *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, 2024, pp. 340–344.

[29] N. Martelaro, "Wizard-of-oz interfaces as a step towards autonomous hri," in *AAAI Spring Symposia*, 2016.

[30] V. C. Segura and S. D. Barbosa, "Uiskei++: Multi-device wizard of oz prototyping," in *Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, ser. EICS '13, 2013, p. 171–174.

[31] T. Pellegrini, V. Hedayati, and A. Costa, "El-woz: a client-server wizard-of-oz open-source interface," 2014.

[32] A. Helgert, C. Straßmann, and S. C. Eimler, "Unlocking potentials of virtual reality as a research tool in human-robot interaction: A wizard-of-oz approach," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, p. 535–539.

[33] ScratchFoundation, "Scratch: Programming for all." [Online]. Available: https://scratch.mit.edu/

[34] GoogleDevelopers, "Blockly: A visual programming language." [Online]. Available: https://developers.google.com/blockly

[35] C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Šabanović, *Human-Robot Interaction: An Introduction*. Cambridge University Press, 2020.

[36] T. Gross, "Interaction research and design across times in hci," in *Proceedings of the European Conference on Cognitive Ergonomics 2024*, ser. ECCE '24, no. 14, 2024, pp. 1–7.

[37] *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*, International Organization for Standardization Std. ISO 9241-11:2018, 2018. [Online]. Available: https://www.iso.org/standard/63500.html

[38] D. J. Mayhew, *The Usability Engineering Lifecycle: A Practitioner's Handbook for User Interface Design*. Morgan Kaufmann, 1999.

[39] D. A. Norman, *The Design of Everyday Things*. Cambridge, MA: The MIT Press, 2013.

[40] S. Krug, *Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability*. New Riders, 2014.

[41] J. Nielsen, *Usability Engineering*. San Diego, CA: Academic Press, 1993.