A VR based Mobile Usability Lab to study multi modal Human Robot Communication

Stefan Friesen, Tabea Runzheimer, Rainer Blum, Jan-Torsten Milde*
* Fulda University of Applied Science, Fulda, Germaniy
milde@hs-fulda.de

***Abstract*—In this paper we describe the development of a VR-based mobile usability lab to study multi modal human robot interaction. Our work is part of a 3 years multi party project, that targets the development of the CityBot, an autonomous modular mobile robot vehicle. The human robot interaction with the CityBot is planned to be multi modal: speech and gestures are going to be used for instructing and guiding the robot. In order to effectivly perform usability experiments and collect speech and gestural data, we developed a mobile VR simulation system of the CityBot in its environment.**

Keywords*—*VR technology, multi modal, human robot interaction, mobile autonomous robot

I. Introduction

We describe the development of a VR-based mobile usability lab to study multi modal human robot interaction. Our work is part of the 3 years multi party project *Campus FreeCity* (see [EDA22]), that targets the development of the CityBot, an autonomous modular mobile robot vehicle. A fleet of CityBots is going to be used for people transport, logistics and other tasks.[1] A physical research prototype is already available (see figure 1).

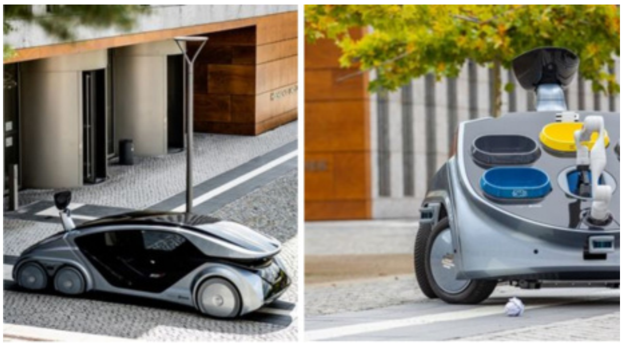

Figure 1: Two configurations of the CityBot physical research prototype on campus of the Fulda University.

The human robot interaction with the CityBot is planned to be multi modal: speech and gestures are going to be used for instructing and guiding the robot. For this, the robot is equipped with the so called *Avatar*, a movable head with mounted cameras, microphones and a display for visual feedback.

II. VR for usability studies

When developing autonomous mobile robots the validation of suitable HMI concepts poses considerable problems. *Building physical robots* is a complex (and costly) task. In this project it is going to take at least 2 years to produce a fully functional physical prototype of the CityBot. This obviously would retard the necessary usability experiments. An equally significant problem is *user safety*. Interacting with the robot could be harmful, especially when the underlying control system is still under development. Using VR environments constitutes an sound alternative to real-life tests. In a VR simulation system central parameters of HMI can be systematically evaluated, while it is much simpler to control the parameter space of the experiments. VR technology has therefore become a standard tool for HMI evaluation in the field of autonomous robots (see [SCNTF19]).

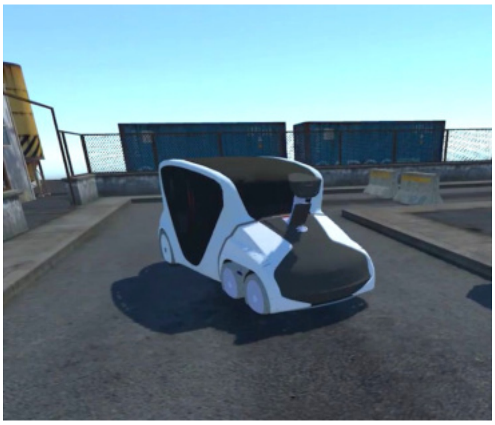

Figure 2: The VR Lab: a scenario consisting of the CityBot base module (tractor) with the small taxi module attached.

In order to effectively perform usability experiments and collect speech and gestural data, we developed a mobile VR simulation system of the CityBot and its environment. The system runs on Meta Quest 2 (see [Met22]), an inexpensive VR head set, which is independent of a rendering computer. The Quest 2 supports speech recording, hand tracking and head tracking. Our VR simulation system can be controlled via OSC messages (Open Sound Control, see [Ope22]). This makes it possible to easily setup Wizard of Oz experiments (see [DJA93], [SC93]), and present movements and reactions of the robot to the participants.

[1] Our special thanks go to the Federal Ministry for Digital and Transport (BMDV) of Germany for financial support within the framework of the joint project "Campus FreeCity". The research is supported under grant *45KI15H081: Hochschule Fulda.*

The immersion of the system is very effective. Participants are able to move freely within a specific region of the setup, thus e.g. circumnavigate the CityBot.

By using the highly portable VR technology it becomes possible to conduct experiments outside the indoor usability laboratory of the university. Experiments can be taken to any place, where test participants are able to move around safely, both in the virtual world and the real world.[2]

A large set of multi modal data is recorded with this mobile VR usability lab (*VR lab*). From within the VR head set the lab is transmitting the current field of view of the test participant, the head position, the head direction, the hand positions, the hand directions and the audio signal (see figure 4). Transmission takes place in real time with a small, but definite latency. As the Meta Quest 2 is not yet able to perform body tracking, we extended the system with an external stereoscopic (depth) camera, the ZED 2 (see [Ste22]).

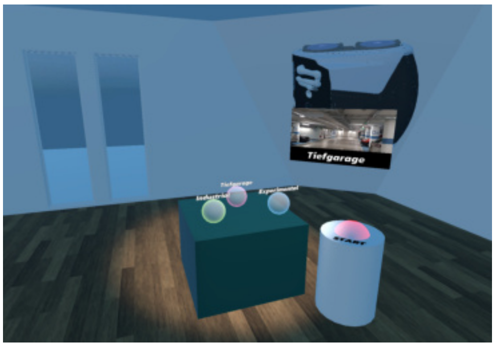

Figure 3: The lobby of the VR Lab. Different scenarios can be selected by the user.

A. Collecting multi modal data

In order to develop a viable and robust interaction concept for the CityBot, collecting relevant multi modal data is indispensable. The creation of a sufficiently specific multi modal corpus, that captures all relevant aspects of a real human robot interaction is an important first development step. This corpus can then be used to

- define the verbal/gestural vocabulary
- identify syntax, semantics and pragmatics of the utterances/gestures used
- identify the language accompanying gestures
- describe communication failures
- define possible (multi modal) reactions of the system
- get insight into the level of understanding of the test participants with respect to current task or situation
- get insight into the emotional state of the test participants with respect to current task or situation

The mobile lab thus provides the base for upcoming usability experiments. These are by no way restricted to human robot interactions. The settings can be easily adopted to the specific needs of the usability experiment at hand. Switching between scenarios is implemented in the form of an interactive lobby (see figure 3).

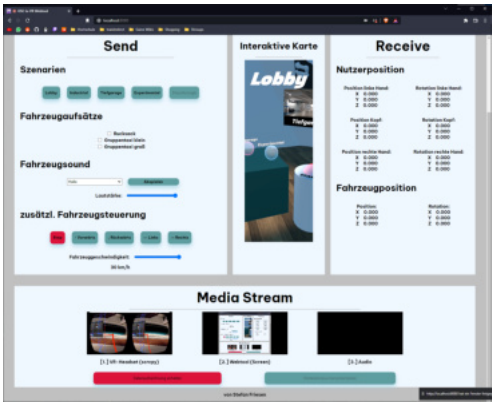

Figure 4: The OSC control tool of the VR Lab. The VR Lab is transmitting the current field of view of the test participant, the head position, the head direction, the hand positions, the hand directions and the audio signal in real time.

III. Open task experiments

A set of experiments have been designed to investigate the multi modal interaction with the CityBot. Modalities (here speech and gestures) are usually temporally intertwined, relate to one another, and can be synchronized at certain points in time in order to support communication (see [McN98]).

While it would be possible to predefine the syntax, vocabulary and semantics of both speech and gestures, our approach is to conduct experiments with an open task specifications. The participants are faced with a situation and are asked to perform a certain task. As they interact with the system according to their own conception, the robot will be controlled by members of our team, thus displaying simulated "intelligent" behavior to the participant. For the given use cases four basic types of situations can be identified:

- situation 0: stopping the robot
- situation 1: directing the robot
- situation 2: specifying (longer term) actions and goals
- situation 3: requesting information

The reactions of the robot are scripted, making it possible to spontaneously reproduce a specific behavior of the robot. Communication success and communication failure will be induced and while communications failure will most certainly put stress or frustration onto the participants, we need to control the level of emotional engagement. Language and gestures will be uttered in varying situations and hence have to be interpreted within these situations. The open task experiments and the corpus of multi modal data are designed to find answers to the following questions:

[2] Free physical space of approximately 16 m^2 is sufficient.

A. What gestures are produced ?

In order to find the gestures produced, a fine grained analysis of the video recordings, the skeletal movement data, the hand tracking data and the head tracking data needs to be performed. Gestures can be very subtle, very quick and show a wide variation range.

B. What meaning do these gestures convey ?

The interpretation of gestures is largely dependent on the current situation. Within the experiments we therefore need to fix the situation parameters. The same experimental setup is going to be repeated a number of times with each participants. This will (may) lead to habituation effects, which then might effect the expression of the gestures used. If similar gestures can be found across experiment repetitions and participants, they may carry a comparable semantic.

C. Do the gestures have structure and can they be generalized ?

Once the overall gestures in the corpus and the underlying structural elements are identified, we will try to find generalizations and hope to be able to develop a formal gesture syntax.

D. What utterances are produced?

Utterances (speech) will be recorded. The audio recording is synchronized to gesture data. All utterances will be transcribed (manually), syntax structures, intent of the utterance and phonetic features will be annotated (markup). If needed, more descriptive layers can later be added. The annotation will follow formal standards available (for an overview see [AG13]).

E. What kind of prosodic features can be detected in the utterances ?

It is assumed, that the syntactic structure of the utterances could be quite simple. Specifically when directing the robot, we expect the language to consists of short one or two word sentences ("Halt!", "Nach rechts!", "Da hin", "Weiter")[3]. Here the prosodic features (loudness, speed, stressed or unstressed intonation, pitch etc.) of the utterances could convey much information.

F. How do utterances and gestures relate?

As the gestures and utterances have been annotated, it becomes possible to inspect their temporal relation, and, based on these findings, find interpretations.

G. Is there a "natural" way of interacting with the robot?

Once a corpus of sufficient size has been collected, we might even be able to identify more general patterns of human robot interactions (of course limited to the given scenarios).

H. What kind of feedback (visual, auditory, mimic, gaze, speech, head and body movement) is needed to convey the inner state/current action of the robot?

Finally, we need to define the communicative means of the robot. As the system is, in a broader sense, anthropomorphic, it possesses a large set of modalities to convey information.

IV. Use Cases

During the final phase of the project (lasting 9 months) a large case study with the physical research prototype of the CityBot will be conducted. The study is based on a set of use cases, which are defined for the *LivingLab* scenario.

The LivingLab is providing an environment for realistic live testing. This includes all aspects of system usage, multi modal human robot interaction, system control, identification of system errors and system failure, analysis of system errors and fixing system errors.

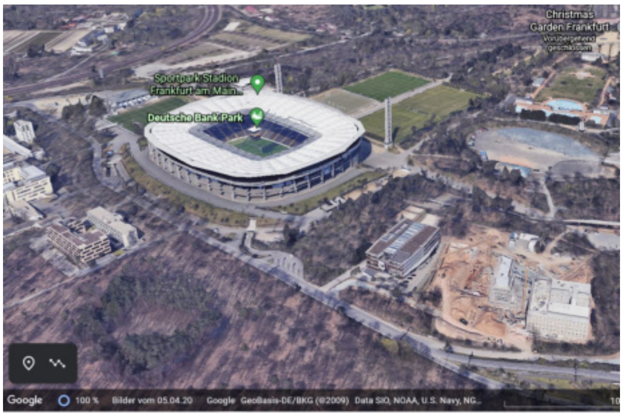
Christmas Garden Frankfurt
Sportpark Stadion Frankfurt am Main
Deutsche Bank Park
Google

Figure 5: Google Earth: the Deutsche Bank Park consists of the stadium, a number of training areas, some office buildings and forest areas. (© Google, see image)

The LivingLab is implemented on the terrain of the *Deutsche Bank Park* in Frankfurt (see figure 5). The Deutsche Bank Park are the club grounds of the German national league team Eintracht Frankfurt. The site's area is of approximately 420.000 square metres in size. In the center of the Deutsche Bank Park you find the soccer stadium with a capacity of more then 50.000 visitors. In addition 5 training fields are placed in near vicinity of the stadium. Two larger office buildings are used by the club's administration. They also serve as the so called *ProfiCamp*, providing optimal care and support for the members of the soccer team.

Furthermore a number of supporting buildings and areas exist: supply and disposal areas, storage rooms for training materials, parking spaces and underground car parks (see figure 6). Access areas are used by employees and visitors alike. It is important to understand, that the LivingLab is not at all a closed area. A large portion of the Deutsche Bank Park consists of forests which have to be publicly accessible. These forests are subject to strict environmental rules, prohibiting extensive sealing of the surface. Indeed

[3] All experiments will be conducted in German. Utterances denote: Stop, Turn right/Move right, Put it there/Go there, Continue.

many routes through the Park are unpaved trails within woods.

CityBots are intended to move freely across the area of the Deutsche Bank Park. They are taking on transportation tasks, which include:

1) transport of mobility inhibited persons (to/from the stadium)
2) transport of grass clippings and garbage
3) transport of the players to and from the training area
4) transport of material to and from the training area
5) transport of food and drinks to numerous target points (beer tents, shops)

As the topology of the Deutsche Bank Park is fixed and defined, all routes that have to be taken by the CityBots are also known by the control system. Starting- and ending positions for each use case and transportation tasks are defined.

A static starting position within many scenarios is the underground parking space (see figure 6). All CityBots are parked here. The parking lot provides the charging stations. It is also used as a the technical maintenance and support area. As such, most routes of the CityBot start from the underground parking space. Eventually the robot will return to this home position, so navigation from any positions on its routes to the underground parking space must be possible.

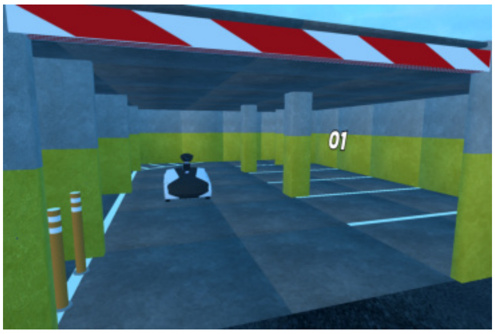

Figure 6: The VR simulation of the underground car park.

Within the underground parking space, a number of specific uses cases are defined:

1) charging the base module (tractor)
2) park in and park out of the parking bays
3) switch modules on the base module (taxi, bag back, trough etc.)
4) generic technical maintenance
5) cleaning of the system

A. Pre-test: stopping the CityBot

Pre-tests are an important step when defining the setup of a usability experiment. They ensure higher quality standards and allow us to evaluate the reliability and validity of the VR experiments prior to their final execution.

A first set of pre-tests have been conducted for *situation 0* (stopping the moving robot). We are using the underground parking lot scenario. A visual marker is placed on the floor of the parking lot. Test participants are asked to move to this position. The CityBot is placed in 20 m distance.

The test participants were then instructed to stop the moving robot when they think, an *acceptable* distance between them an the robot has been reached. No further instruction on how to stop the robot were given, but they knew, that the robot was able to understand language and gestures.

The preliminary results of this first set of pre-tests are promising: the participants behave in an natural way and already a number of insights could be drawn with respect to the upcoming design of the yet to develop interaction system.

V. Conclusions

We developed a mobile VR-based usability lab to inspect the the multi modal human robot interaction with the CityBot, a mobile autonomous robot vehicle. The VR lab allows to easily set up experimental scenarios, control the simulated robot via OSC and transmit the relevant multi modal data to a host computer in real time.

A first set of pre-tests has been successfully conducted, showing that the test participants immerse in the situation and react in a "natural" way.

References

[AG13] Ágnes Abuczki and Esfandiari Baiat Ghazaleh. An overview of multimodal corpora, annotation tools and schemes. *Argumentum*, 9(1):86–98, 2013.

[DJA93] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4):258–266, 1993.

[EDA22] EDAG. The CityBot Homepage. https://www.edag-citybot.de/en/, 2022. [Online; accessed 1-February-2022].

[McN98] David McNeill. Speech and gesture integration. *New Directions for Child Development*, pages 11–28, 1998.

[Met22] Meta/Oculus. The Meta Quest2 VR Headset. https://www.oculus.com/quest-2/, 2022. [Online; accessed 1-February-2022].

[Ope22] OpenSoundControl.org. The OpenSoundControl (OSC) specification. https://ccrma.stanford.edu/groups/osc/index.html, 2022. [Online; accessed 1-February-2022].

[SC93] Daniel Salber and Joëlle Coutaz. Applying the wizard of oz technique to the study of multimodal systems. In *International Conference on Human-Computer Interaction*, pages 219–230. Springer, 1993.

[SCNTF19] Sebastian Stadler, Henriette Cornet, Tatiana Novaes Theoto, and Fritz Frenkler. A tool, not a toy: using virtual reality to evaluate the communication between autonomous vehicles and pedestrians. In *Augmented reality and virtual reality*, pages 203–216. Springer, 2019.

[Ste22] Stereolabs. The ZED-2 stereoscopic robot camera. https://www.stereolabs.com/zed-2/, 2022. [Online; accessed 1-February-2022].