



HAL
open science

An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual Reality

Florent Alain Sauveur Robert, Hui-Yin Wu, Lucile Sassatelli, Stephen
Ramanoel, Auriane Gros, Marco Winckler

► **To cite this version:**

Florent Alain Sauveur Robert, Hui-Yin Wu, Lucile Sassatelli, Stephen Ramanoel, Auriane Gros, et al..
An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual
Reality. 2023 IMX - ACM International Conference on Interactive Media Experiences, Jun 2023,
Nantes, France. <https://dl.acm.org/conference/imx>, 10.1145/3573381.3596150 . hal-04102737v2

HAL Id: hal-04102737

<https://hal.science/hal-04102737v2>

Submitted on 16 Jun 2023

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.

Copyright

An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual Reality

Florent Robert
florent.robert@inria.fr
Centre Inria d'Université Côte d'Azur
Sophia-Antipolis, France

Hui-Yin Wu
hui-yin.wu@inria.fr
Centre Inria d'Université Côte d'Azur
Sophia-Antipolis, France

Lucile Sassatelli
lucile.sassatelli@univ-cotedazur.fr
Université Côte d'Azur, CNRS, I3S
Institut Universitaire de France
Sophia-Antipolis, France

Stephen Ramanoel
stephen.ramanoel@univ-cotedazur.fr
Université Côte d'Azur, LAMHES
Sorbonne Université, INSERM, CNRS,
Institut de la Vision
Nice, France

Auriane Gros
auriane.gros@univ-cotedazur.fr
Université Côte d'Azur, CHU de Nice,
CoBTeK
Nice, France

Marco Winckler
marco.winckler@univ-cotedazur.fr
CNRS, I3S
Centre Inria d'Université Côte d'Azur
Sophia-Antipolis, France

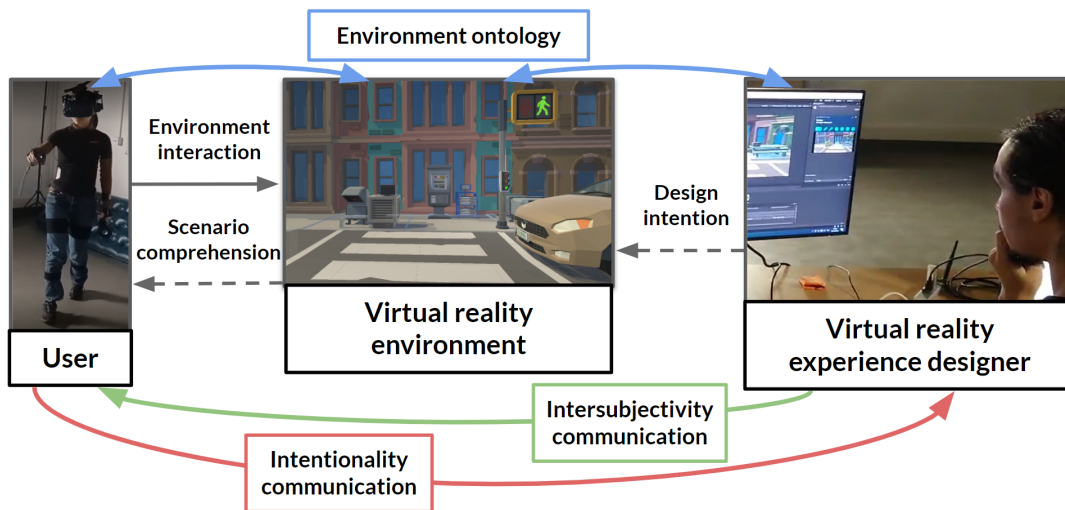


Figure 1: Our virtual reality framework based on Dourish's [15] theory of embodiment. Dourish describes three elements to address when bridging gaps of perception between user, system, and designer during user experience analysis, including (1) an environment ontology, (2) the intersubjectivity communication and (3) intentionality communication.

ABSTRACT

Virtual Reality (VR) technology enables “embodied interactions” in realistic environments where users can freely move and interact, with deep physical and emotional states. However, a comprehensive understanding of the embodied user experience is currently limited by the extent to which one can make relevant observations, and the accuracy at which observations can be interpreted.

Paul Dourish proposed a way forward through the characterisation of embodied interactions in three senses: ontology, intersubjectivity, and intentionality. In a joint effort between computer and neuro-scientists, we built a framework to design studies that investigate multimodal embodied experiences in VR, and apply it to study the impact of simulated low-vision on user navigation. Our methodology involves the design of 3D scenarios annotated with an ontology, modelling intersubjective tasks, and correlating multimodal metrics such as gaze and physiology to derive intentions. We show how this framework enables a more fine-grained understanding of embodied interactions in behavioural research.

CCS CONCEPTS

• **Human-centered computing** → **Systems and tools for interaction design; User studies; Computing methodologies** → **Virtual reality.**

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
IMX '23, June 12–15, 2023, Nantes, France
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0028-6/23/06...\$15.00
<https://doi.org/10.1145/3573381.3596150>

KEYWORDS

Embodied experiences, 3D environments, immersion, scene ontology, task modeling, interaction, navigation, user experience analysis

ACM Reference Format:

Florent Robert, Hui-Yin Wu, Lucile Sassatelli, Stephen Ramanoel, Auriane Gros, and Marco Winckler. 2023. An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual Reality. In *ACM International Conference on Interactive Media Experiences (IMX '23)*, June 12–15, 2023, Nantes, France. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3573381.3596150>

1 INTRODUCTION

Virtual Reality (VR) offers exciting opportunities for designing embodied experiences – experiences that offer users a high-level of immersion, with life-like scene events, rich interactions, freedom of movement, and the feeling of living and breathing in the 3D world. Inspiring works have shown how embodiment impacts racial [4] and gender [29, 38] bias and how immersive journalism invokes empathy [11]. These perspectives open exciting avenues of research in domains of education, training, and rehabilitation, where user behaviour can be observed in realistic situations in order to provide personalised content and protocols.

However, building a VR experience enabling complex interactions while allowing a proper embodied experience analysis includes multiple challenges and limitations. Analysing embodied interactions to interpret and understand the user's experience and intentions is complicated: on the system level, there can be a mismatch between the intentions of the user and the designers' expectations when the experience is first created [16]; on the individual level, there are non-visible changes in our visual, emotional, motion processing and related sensations that reflect one's intentions [31]. Creating an environment allowing the fluid and clear communication of intentions between the designer and the user is very important when we want to design personalised applications, and thus a grand challenge for VR media.

In order to reduce the gaps in perception between the designers, the user, and the system, previous work [46] identified three major components to develop a framework for embodied experiences, shown in Figure 1, that was inspired by Dourish's work on embodiment theory [15]:

- **Ontology:** building semantically understandable 3D environments and scenarios in which the user's interactions (e.g., navigating to a location, picking up an object) can be understood without being precisely explained.
- **Intersubjectivity:** designing real-time visualisation and control systems such that designers can communicate goals and constraints of the scenario to the users (e.g., for the task "open the door", indicate that they need to first find a key).
- **Intentionality:** designing computational methods for the analysis and identification of users' actions and their purpose of enacting an effect on the world (e.g., taking a key in order to open a door enables navigation to a previously inaccessible space).

Based on Dourish's theory, we used the open GUST-3D framework [46] to develop an experimental paradigm in a joint effort with researchers in computer and cognitive sciences, and designed

tools that can give better insight into user embodied interactions in VR environments. The chosen study investigates the impact of low-vision conditions on life-sized road crossing scenes. Users were given (1) various tasks in succession (process visual indices, obtain objects, navigate to points of interest, interact with entities), (2) under either normal or simulated low-vision conditions (i.e. with a virtual scotoma, a region in the centre of the visual field where visual information is blocked out), (3) and with real walking or simulated walking using a headset gamepad. The ultimate goal of the study is to investigate the potential of VR for rehabilitation and training for patients with low-vision and other motor-sensory impairments.

This work focuses not on the results of the study itself, but the implementation of the open framework that enabled the study to be realised: from the design of the interactive task model, to setup and running of the study, to preliminary analyses that investigate the complexity of embodied experiences. The framework presents three main contributions:

- the definition and encoding of a task model at levels of ontology, intersubjectivity, and intentionality,
- realisation of a technical platform with multimodal sensors and freedom of navigation, and
- a study and preliminary analysis of 16 participants under this framework to show its potential to support global and fine-grained analyses.

In the following sections, we first present in section 2 the related work on user experience analysis in VR, and how this work is positioned in respect to the existing. Then we describe in section 3 more precisely the technical setup for embodied experience analysis, the study design, and how our methodology allows us to respond to the three senses of embodiment. Then in section 4 we present the study design and early results in section 5 showing both aggregated results and a fine-grained analysis of user behaviour. Finally, we discuss in section 6 the limitations and present the next steps of this work and the future analyses we are aiming for.

2 RELATED WORK

Virtual Reality triggers stronger emotions and sense of presence as compared to traditional 2D media, as the user conduct embodied interaction to exchange with the system, stimulating the brain in a way comparable to real life interactions as shown in Alcañiz et al. [1]. Multimodal behavioural indices (e.g., motion, attention, emotion) can thus enrich and better characterise the multimodal embodied user experience that VR systems offer, as compared to traditional media.

In order to address these elements, we base our study of existing work on Dourish's theory [15], the most dominant and influential theory in the domain of embodied interactions applied to research in HCI, explaining how the gap of comprehension between user and designer can be created on three levels: the understanding of the environment (ontology), the understanding of the task to perform (intersubjectivity), and the communication of the experience to the designer (intentionality). These three elements are needed for a complete embodied experience analysis. The model proposed by Dourish has the strong advantage of being simple but comprehensive in characterising all the interactions and gaps

of perception that take place between the 3 actors (user, designer, environment), which is well aligned with common software architectures in human-computer interactions.

In this section we review the way VR technologies have been adopted for multimodal user understanding in interactive scenarios. With regards to the three axes of embodied experiences, we review: (1) systems for the study of user behaviour in VR with annotated contexts, (2) a taxonomy of task modelling and interaction for VR, and (3) behaviour understanding in cognitive sciences using VR setups.

2.1 Annotated contexts for user understanding

These past years, with the popularisation of VR technologies, there has been an explosion of research into user understanding in specific contexts. Its potential for applications in contexts that can be considered as highly dangerous or expertise intensive are most notable, such as for understanding driving habits [28], sports learning [45], and firefighter training [9]. With rich representations and annotations of the 3D environments, we can envision much more personalised training and feedback protocols.

There are few works that propose ontology-based methods for context representation in interactive 3D applications. One notable example is that of Bouville et al. [7] who built the system #FIVE on modelling interactive VR environments through the annotation of elements composing the environment with properties, though with a pre-defined ontology. Kim et al. [25] used deep learning techniques to construct 3D scene graphs from images to create a simulated environment for robot interactions with real environments. The vocabulary in the scene graph is primarily composed of objects and their visual properties, without knowledge about their interactive capabilities. The recent workflow proposed by Wu et al. [46] integrates a customizable ontology component in order to define a set of vocabulary representing objects, a limited number of interactive possibilities, and navigation regions.

The design of VR embodied experiences can vary due to the design goals, target user experience, as well as social, cultural and language contexts. A shift from a universal vocabulary to dynamic vocabulary, an ontology, that can adapt to the situation of usage is thus key to creating personalised experiences for a larger variety of contexts.

2.2 Interaction modalities and task models

The Unity game engine is one of the most popular applications used to help researchers in the creation of experiences in VR, with works such as Ugwitz et al. [40] allowing the trigger and the record of various interactive events in the environment, or Villenave et al. [43], allowing the record of multimodal data on the user experience, allowing a visualisation of the data combined for a more comprehensive view of the user experience.

Few works involve long range (greater than in a 4m x 3m space) navigation tasks with real walking, mainly due to the limits of current VR technologies. Boldt et al. [6] mention room-scale navigation but cite the HTC Vive Pro's 12m² space limitation and do not precise the scale of their study design. They designed a task involving walking to a target and interacting with it, with the goal to investigate the effect of haptic and auditory feedback on wall

perception. Mousas et al. [34] conducted a path following task of 150 meters with a VR backpack, and analysed physiological and pose data to investigate how mismatching virtual and real environments can strongly affect user behaviour. This paper puts to light the fact that both virtual and real environment matter in VR experiences, adding a layer of complexity on the design of a proper VR experiences in which real walking is included. VRoamer [8] also chooses a VR backpack to allow navigation within a 40m x 40m space, though the work focuses on enabling free navigation and is not task oriented. Finally, studies on perception during navigation in VR with EEG [12] also used a VR backpack.

In our work, we present a setup allowing real walking in a contextual environment in a 10m x 4m space by using a wireless module for VR headset. While technical requirements, documentation, and error management of the device was experimental, this turned out to be an efficient solution for natural navigation and sense of presence. To our knowledge, the usage of this technology in a 6DoF movement study has never been done before.

2.3 User cognition understanding

The question of the relations between VR content, user interactions, and user perception is gaining interest in the area of cognitive neuroscience. We see this from works using cognitive modelling techniques to understand the general relation between emotion and agency [22], or properties of the design of content such as visual complexity [18] and information placement [39].

In parallel, we also see a general rise in the use of methodologies of perception, emotion, and behaviour analysis from neurocognition for the study of the user experience in VR. Alcañiz et al. [2] apply transcranial Doppler sonography to investigate how VR experiences stimulate the brain, finding them comparable to real life experiences. Dickinson et al. [14] investigated diegetic interfaces (interface integrated and adapted to the virtual environment) in VR explain that while these type of interfaces show good efficiency in traditional 2D media, in VR, such interfaces did not bring immediate benefits, but did however increase the workload and completion time of users. These works show how the modalities of interactions in VR differ greatly from 2D media, and may require a whole new set of interactive metaphors, and ways to observe and analyse the user experience. Many works such as Dewez et al. [13], Aseeri et al. [3], Peck et al. [35] and Ricca et al. [37] show how embodiment, through body or hands visualisation, is an important aspect of VR interaction, affecting how users will interact with the environment and interact with other users.

Many studies focus on experience analysis of users in VR, such as the of works Xue et al. [47] and Guimard et al. [17], showing how multiple videos triggering different levels of arousal and valence in the user in order to observe and analyse their physiological responses. These works allowed only 3DoF (3 Degrees of Freedom) head movement, such that users had to stay in-place, reducing the amount of constraints to take into account for the proper analysis of the experience.

Physiological analysis of VR experiences in 3D environments are often used for multiple purposes such as Keighrey et al. [23] measuring the QoE (Quality of experience) based on physiological

data and interactions, or Bender et al. [5] quantifying with physiological data the gratification people feel while carrying out violent actions in virtual scenes, such as killing zombies in a VR game. In real-life applications, physiological information have been used in clinical settings to analyse the physiological and mental state of the user for medical purposes for anxiety prediction [19, 33] as well as stress and cognitive load assessment [42]. Other applications uses physiological data for driver stress detection [32], and emotion evaluation with smart clothing [36].

The aforementioned review highlights the need of a framework for the comprehensive understanding of the embodied experience in VR in various stages of an application scenario for behavioural studies: from the design of the virtual scenario, to the data recording, and finally in the analyses in order to address the complex interactions between scene context, task design, and user's experience. Most systems and usages we surveyed focus on a subset of these aspects in a well-defined situation of usage. In our work, based on Dourish's [15] three senses of embodiment, we built a methodology aiming to address a wide range of aspects that could create a gap of perception between a designer and a user, in the environment creation and descriptive ontology, in providing correct user guidance in the scenario, and in the collection of multimodal metrics and subsequent analysis of the user's experience. With a well-controlled environment naturally offered by VR, there is the challenge of providing sufficient control of the experience and the elements affecting it for content designers from non-technical backgrounds, and provide cognitive science researchers with the tools to understand the system and cater it to their research hypotheses, as well as avoid unwanted external factors that may influence the final results. This is the challenge we must address in this work.

3 METHODS AND MATERIALS

In a joint effort of computer science, neuroscience, and clinical practitioners, our aim for this study was to investigate the impact that a simulated low vision condition had on user navigation in complex environments, from the analysis of attention, emotion, and behaviour. The complex environment of choice was road crossing scenes: a common daily situation where the difficulty to access and process visual information (e.g., traffic lights, approaching cars) in a timely fashion can lead to serious consequences on a person's safety and well-being. Analysing the impact of low vision and navigation in such situations imposed a number of technical constraints that needed to be fulfilled:

- Viably capture gaze, physiology, and motion data without the sensors interfering with the tasks.
- Free natural walking in a large space – at least the length of a standard two lane road pedestrian crossing.
- Logging of user interactions within a scene context – scene annotated with an ontology representing object, interaction, and navigation properties.

The rest of the section, we present and discuss the design of such a study and the rationale for our design choices.

3.1 Multimodal sensors

The metrics chosen shown in Figure 2 were selected in order to be able to rebuild the whole experience by synchronising and correlating the data relevant to the embodied experience (i.e., where was the user, what the user was interacting with, in what state was the environment, in what emotional state was the user, ...). Table 1 summarises the modalities of data recorded : system logs, physiological data, motion capture data, and gaze and head motion.

The framework was made in order to incorporate a large variety of metrics, and be as flexible as possible in order to adapt to specific other potential study conditions. We surveyed a large range of equipment and their usage in existing work. On the technical requirements side, an important motivation for the choice of metrics was the presence of (or possibility to add) a Unix Timestamp, and the minimisation of latency, in order to facilitate the synchronisation of all the data together at any time of the experiment, necessary to the analysis and viability of the results we wish to present.

Eye tracking headset. Eye tracking is a strong prerequisite for this study, in order to place a virtual scotoma in real time based on the gaze position for the simulated low-vision condition. However, eye tracking data itself is also insightful in an embodied experience analysis, allowing the observation of the user's attention during a scenario, as done in Xue et al. [47] and Guimard et al. [17]. We surveyed the HTC Vive Pro Eye ¹, the HP Omnicept Reverb 2 ², and the Varjo VR2 ³ headset, all of which included more or less equal eye tracking capabilities. The weight of the Varjo VR2 excluded it from our choices. In the end we chose the HTC Vive Pro Eye, which includes by default eye and head tracking functionality. With this headset, the SteamVR application was used for the VR environment configuration. The main factor influencing headset choice was the navigation constraints, which will be detailed in Section 3.2. One issue that is present for most eye tracking devices in VR headsets is latency in the recorded data. The head data on the HTC Vive Pro Eye does not have this delay, which we use as a baseline to align with the eye tracking data before synchronising with the data captured with other equipment.

Motion capture. Motion capture provides very rich information about embodied experiences, as shown in Mousas et al. [34] work, using metrics such as step length of participants to evaluate how confident they feel walking in a VR environment. We surveyed two inertia-based motion capture systems: the Rokoko mocap costume ⁴ with 19 sensors and the MVN Awinda (jacket + 17 sensors) ⁵ for body motion tracking. The MVN Awinda system was chosen in the end due to its resilience towards magnetic interference as well as the precision of the captured data. The Xsens MVN 2022 software was used to calibrate and record the data, and the Xsens MVN motion cloud was used to convert the records in formats *.mvnx* and *.bvh* for later analysis, the *.mvnx* format being the only one containing Unix Timestamp.

¹<https://www.vive.com/eu/product/vive-pro-eye/overview/>

²<https://www.hp.com/us-en/vr/reverb-g2-vr-headset-omnicept-edition.html>

³<https://varjo.com/>

⁴<https://www.rokoko.com/products/smartsuit-pro>

⁵<https://www.movella.com/products/motion-capture/xsens-mvn-awinda>

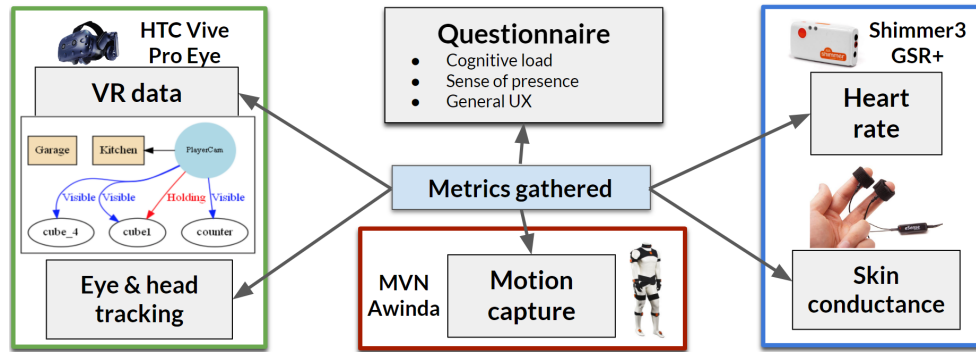


Figure 2: The recorded data includes sensors in the headset and system – gaze and head tracking and system logs –, motion capture data from the XSens MVN Awinda starter, and skin conductance and heart rate from Shimmer3 GRS+ sensors.

Physiology sensors. We decided to use sensors including skin conductance and heart rate, metrics popularly used by Memar et al. [32], and Pidgeon et al. [36] for the evaluation of the level of user emotion arousal. We surveyed the Empatica E4⁶ and the GSR Shimmer3+⁷ which were at similar price ranges and included sensors for skin conductance and heart rate. The GSR Shimmer solution was chosen in the end for its higher data rate (15Hz compared to 4Hz). The Consensus software was used for the configuration and the data record done on the Shimmer’s SD card.

Sensors not interfering with the VR embodied interaction was another important criterion in the design of this setup, as it would affect both user behaviour, and introduce noise into the data. We did multiple test to find the ideal setup and how the user would use it, notably for physiological sensors and controllers, both relying on hands for interaction and data record. Using the non-dominant hand for physiological sensors and the dominant hand for simulated walk on controller was the most efficient solution we found.

Table 1: The collected data modalities, the logging frequencies, and a brief description of the data modalities.

Type	Freq.	Description
System scene log	10 Hz	objects position, objects state, objects interactive properties
System user log	10 Hz	position, interactions, current task, object visibility to the user
Physiological sensors	15 Hz	Heart Rate (HR), electrodermal activity (EDA, skin conductance)
Motion capture	60 Hz	17 sensors for head (1), torso (4: shoulders, hip, and stern), arms and legs (8: upper and lower limb), and feet and hands (4)
Gaze and head tracking	120 Hz	For left, right, and cyclopean eye (combined gaze vector of both eyes): gaze vector (x,y,z) , pupil size, eye openness percentage, and data validity mask

3.2 Navigating in a large space

In order to analyse the impact of low vision on navigation in large spaces, such as for our road crossing task, we first have to define the size of the environment. In standard road crossings, the minimum required width of a car lane is 3.5m, and the minimum width of the pedestrian crosswalk was 2.5m with the standard being 4 – 6m. We included one meter of pedestrian crossing on each side of a two lane road, and some margin on the sides of the crosswalk to ensure safety. In total, the required navigation space for this study is 10m x 4m, delimited in Figure 3.

To find a lightweight solution to navigate in a large space represented the most challenging part of the setup. Existing headsets that have eye tracking capabilities which we were able to survey were all tethered headsets. We therefore had only three options: buy a very long headset cord, use a VR backpack along with other sensors, or find ways to untether the headsets. The first two were discarded after multiple rounds of discussion: the longest available cord for all three headsets would only just allow the user to reach the borders of the space, resulting in tension at the back of the head. With all the required sensors for the study, the VR backpack would add a significant load on the user and impact analysis on the pose.

In the end, we went with the wireless module⁸ for the HTC Vive Pro Eye. However, the module was extremely experimental, mostly relying on community forums for information. There was, for example, no information on the range of the module sensor, nor whether eye tracking data could be collected with the module. We ran multiple tests and were able to verify a number of capabilities of the module shown in Table 2, including the range of navigation we desired, as well as compatibility with the eye tracking data.

To our knowledge, we are the first to use the wireless module in a study requiring 6DoF movement, with few very recent and notable instances of its usage to facilitate placement of EEG electrodes [21, 26]. Nevertheless, it ticked all of the boxes for weight, data collection, and freedom of movement. By removing the cable, pilot testers felt that the movement was much more natural, not restrained by the cable, increasing their feeling of presence. Compared to other existing solutions, we thus decided that we would work around the limitations.

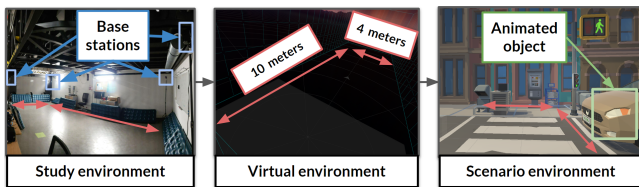
⁶<https://www.empatica.com/en-eu/research/e4/>

⁷<https://shimmersensing.com/product/shimmer3-gsr-unit/>

⁸<https://www.vive.com/eu/accessory/wireless-adapter/>

Table 2: Technical capacity of the HTC Vive Pro wireless module based on our homemade tests

Technical requirements	We tested two computers, one with a 2080 GTX and i9 CPU, and the second with a 3080 RTX and Xeon CPU, both with the required PCIe slot, but only the latter was able to smoothly launch the scenes.
Navigation space	The camera needs to be able to visibly “see” the headset at all times. We used four base stations, one at each corner of the space, and were able to calibrate the space of 4m x 10m.
Data	Compatible with Steam VR to collect eye tracking data and 6DoF headset position and rotation data.
Battery	The battery allows roughly two hours of usage
Comfort	The module was relatively light, but continuous usage for longer than an hour could result in overheating and transmission errors, causing significant lag and screen freezes

**Figure 3: The environment used for the experience measures 4 by 10 meters in navigation zone is delimited by four base stations, one at each corner, and aligned with the virtual environment. Mattresses surround the zone for security.**

3.3 System logging

We conceived a study with different types of tasks, in two movement conditions and two vision conditions. We then needed a way to annotate the scene with object types and interaction possibilities, then record fine-grained interactions in the scene context.

Our task management and logging system is built on the open source workflow⁹ of GUST-3D [46], allowing the conception of a vocabulary to describe 3D environments and their interactive properties, as well as the definition, with this vocabulary, of tasks for the user to carry out. The vocabulary ultimately becomes an ontology in the form of a scene graph. This ontology is then used for the creation of scenarios and the guidance of users through it. The system will automatically record all the states and interactions about the environment (i.e., object the user is currently holding, current state of the traffic light, position of the cars) and can be used post-study to provide visualisations to the designer. This workflow appears to be very effective to offer a proper control of the complete experience, from the environment design to the experience analysis, in coherence with Dourish’s theory.

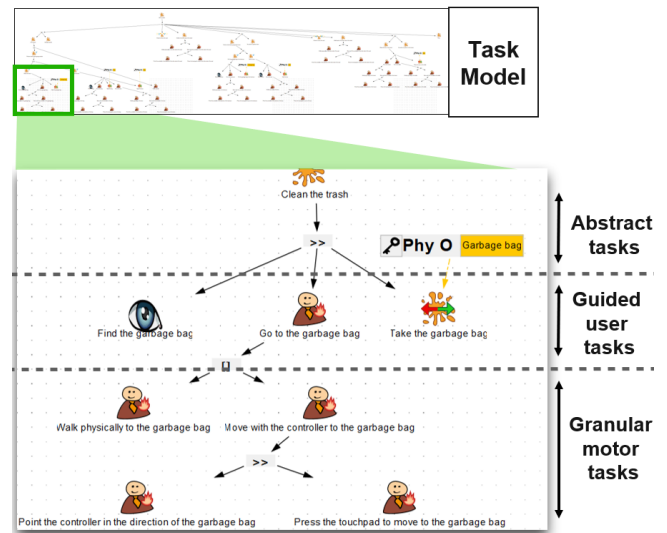
4 STUDY

We conceived a framework for the analysis of embodied experiences in VR with multimodal data. Using this framework, we design a study to analyse user experience in VR, with different types of

tasks (observation, movement, grasping) in two different movement conditions (walking physically, walking with a joystick) and two different vision conditions (normal vision, simulated low-vision), for a total of four conditions. The study was reviewed and approved by the Université Côte d’Azur’s ethics review board (CER).

This section details the framework, using a task model for subsequent design of the conditions and tasks in the study, and elaborates on the questionnaires that complement the physiological and motion data recorded.

4.1 Task model

**Figure 4: Modelling tasks supported by the system for the study in a task model. The task model defines the precise the order of tasks, and the subdivision of the tasks in various levels of granularity from abstract to low-level motor tasks. In the study, instructions are given at the level of “Guided User Tasks”, and the lower granular motor tasks act as a cursor for intentionality analysis.**

The framework is based on a task model shown in Figure 4, created to define all scenario tasks, and visualise all objects and interactions the system should be able to support in the study. To allow fine-grained analysis, knowing precisely what task is done and in what order allows us to synchronise the recorded data, to match the physiological state of the user with the task they did. The tasks were described from the high abstract level of the task such as “go outside” to the motor level of the task such as “approach the key” or “press the game pad trigger to grab the key”.

Choosing at which level to give to the user instructions depends on the type of study and desired subsequent analysis. In our case, this study has the purpose to analyse user behaviour and intention, establishing intersubjectivity with the user, while still leaving a certain amount of freedom such that they may carry out the tasks somewhat differently, allowing a richer analysis of intentionality later on. The intermediate “Guided user task” level in Figure 4 was therefore selected and transcribed as audio instructions played to

⁹<https://project.inria.fr/creative3d/gust-3d/>

the user. Tasks below this level are therefore considered as information, the “intentions” that we want to deduct and analyse from our multimodal data.

4.2 Task design

Based on the task model, scenarios were created with the suitable interactions. The task design presented two main constraints:

- The tasks themselves needed to involve both navigation and interaction according to the task model, in somewhat realistic situations, and with a slight amount of changes between each scene, to avoid learning effects from repeating the exact same scene.
- To limit fatigue, the study should fit in two hours, including the time taken to setup, and no more than one hour in the headset.

In the end, we considered that for each condition, we could perform six scenarios of one to two minutes long, each with precise properties to trigger certain types of user behaviour.

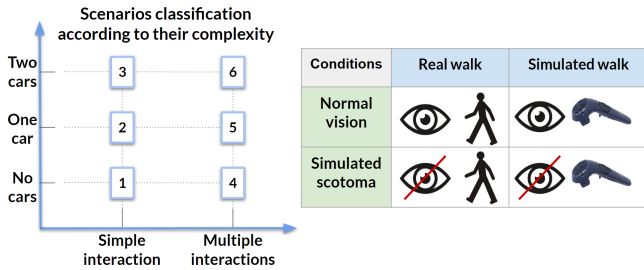


Figure 5: (Right) Our study was comprised of four conditions as a combination of real or simulated walking, and with or without a simulated scotoma as a low vision condition. (Left) Each condition involved six scenarios with varying levels of interaction complexity and cognitive load (number of cars)

In order to build a dataset composed of a large range of user behaviours, six scenarios were designed around two axis of metrics we wanted to observe on the user experience shown in Figure 5 :

- **Cognitive load axis** affected by changing the amount of road lanes and cars driving in the VR environment, ranging from 2 lanes with a car on each, to 1 lane with no car at all, as shown in Figure 6.
- **Interaction complexity axis** affected by changing the amount and the type of interactions asked to the player during the scenario, ranging from only one task asking to pick up one object, to multiple tasks of object interaction, object pick up, object placement, traffic light observation. Figure 7 show a scenario with high interaction complexity.

The six scenarios were performed in four conditions: two movement conditions and two visual conditions, for a total of 24 scenarios, as shown in Figure 8:

- **Simulated walking** the user moves forward or backward using the touchpad of the controller. The direction of movement is determined by the direction of the gamepad, leaving the user’s head free to explore the environment. The user could turn on the spot, but physical walking wasn’t allowed.



Figure 6: Top-down view of the two different type of environments

used in the scenarios. **One lane road** is used for scenario with 1 or less cars, **Two lane road** is used for 2 cars scenarios.

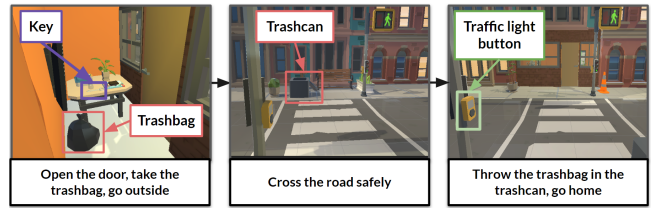


Figure 7: Participant point of view of the scenario #4 from the study featuring multiple interactions (picking up the trash bag and pushing the traffic button) and crossing a single lane street with no cars.

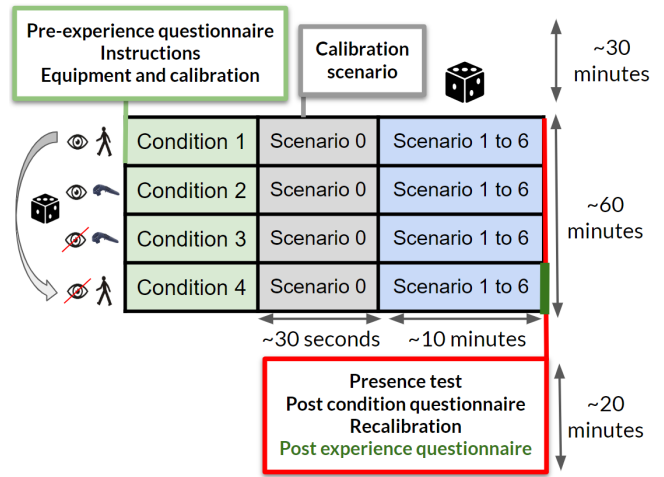


Figure 8: The study protocol consists of three main stages: (1) pre-experience preparation including signing the informed consent, a questionnaire and equipping the headset and sensors (2) the study involving seven scenarios per condition (the first of which is a calibration scenario), two perspective taking tasks in the middle and at the end of a condition, and a post-condition questionnaire after each condition, and (3) a post-study questionnaire and removal of equipment. The condition and scenario sequence were pseudo randomised. The entire study lasted roughly two hours.

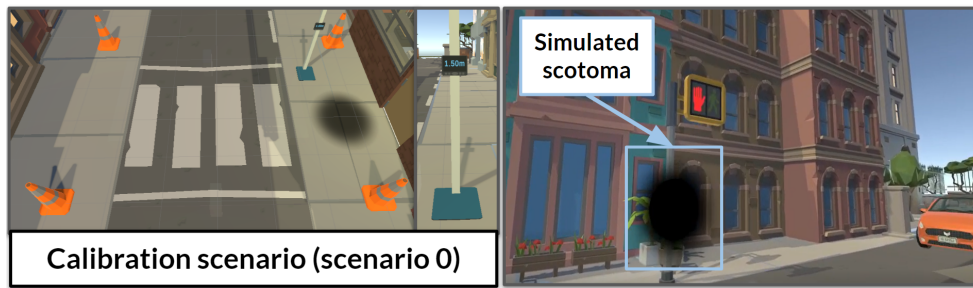


Figure 9: (Left) Top-down view of the calibration scenario environment, with no cars nor objects to interact with. In this scene, the participant tests the current walking and visual condition. They are then asked to stand up straight in front of the meter (right side of the figure) to calibrate their height in order to avoid feelings of loss of balance. (Right) Participant view of the simulated scotoma – a region at the centre of the visual field with no visual information – following the gaze of the participant. The scotoma represents 10° in diameter of the foveal field of view, roughly equivalent to the max distance between the index and middle finger with the arm fully stretched.

- **Real walking** the user walks physically and naturally in the 10 meters by 4 meters tracked space.
- **Normal vision** no vision change.
- **Simulated low-vision** using eye tracking, a black circle with a diameter of 10° of the foveal field is placed in the centre of the vision of the user, as shown in Figure 9.

4.3 Questionnaire

The questionnaire was split into three parts: pre- and post-study questionnaires, and post-condition questionnaire that were identical following each condition. The questions came globally from existing studies of presence (Witmer & Singer[44], emotions (SAM[27]), simulator sickness (SSQ[24], and technology acceptability (UTAUT2[41] with the validated French version[20]). The full list of questions included in the questionnaires is listed in Table 3.

4.4 Study protocol

With all the above elements, the experimental protocol is presented in Figure 8. The participants upon recruitment were sent a message with their booked time slot, guidelines to wear fitting or light attire, as well as the informed consent to permit ample reading time. The study lasted approximately two hours long, and was conducted in either English or French at the preference of the participant. 20 euros compensation was given in the form of a check at the end of the study. At the study time, the participants were first invited to sign the informed consent, answer the pre-study questionnaire, and fitted with the equipment. Participants were briefed on the risks of nausea, fatigue, and motion sickness, and were encouraged to ask for a pause or request ending the study if they felt discomfort, which would not impact their compensation. Snacks and drinks were made available to the participant and offered by the experimenters between conditions and at the end.

During the study, two experimenters were always present to help arrange the equipment, answer questions, and provide the participant with guidance in using the equipment. When the participant is walking with the headset on, one of the experimenters is always focused on the participants to notice any loose equipment,

check for risk of collision or falling. Inflatable mattresses surround the navigation zone (Figure 3) to prevent any collision with walls or equipment.

At the beginning of each condition, a pilot scenario (numbered 0) was presented to the participant to help them discover the environment, familiarise with the interactive and navigational modalities in order to lower the learning curve. This scenario is also used to calibrate the headset height, as shown in Figure 9. The calibration was designed after numerous pilot tests that showed a miscalculation of headset height using the Vive's integrated sensors, and also encouraged users to maintain a more upright pose to avoid instability.

Following the pilot scenario, each participant then completed six scenarios under the four conditions, for a total of 24 scenarios. The sequence of the conditions and the sequence of scenarios under each condition were pseudo-randomised using the Latin Square attribution to avoid the effect of repetitive learning and fatigue on specific conditions.

At the end of every three scenarios, participants also performed a spatial perspective taking test [10], to quantify the level of presence the user experiences in the environment. During this test, the virtual environment is hidden, and the participant is asked to point with their arm in the direction of a target designated by the audio instruction, usually a salient object with which they interacted during the scenario such as a trash can or the initial location of the key or garbage bag. A strong deviation between the participant's arm and the correct direction would indicate a higher level of spatial disorientation, and likely a reduced level of presence.

The post-condition questionnaire was presented to the participant after each condition, which also served as a pause period, to re-calibrate motion tracking, eye tracking, physiological sensors, and give the participants the opportunity to ask any questions and declare sensations of fatigue or nausea. At the end of all conditions, the equipment was removed and the final post-study questionnaire was presented to the participant.

Table 3: The questionnaire we used, split into pre-experience, post-experience, and post-condition (following each condition). The participants were allowed to skip questions if they did not wish to respond.

Section	Question	Description / Options
Pre-experience	Questionnaire information	User ID, study language
	Demographics	Gender and age group
	Previous experience with VR	Never, 1-2 times, sometimes, frequently, developer
	Do you play video games	never, rarely, sometimes, frequently, always
	Do you experience motion sickness while playing video games or on transportation?	never, rarely, sometimes, frequently, always
	What situations cause motion sickness?	open question
Post-condition (x4 conditions)	UTAUT2	the UTAUT2 questionnaire[41] and validated French version [20]
	Condition	real/simulated walk; with/without scotoma
	Set of questions on investment	Scale of 1 (not at all) to 10 (very much) on: physical load, mental load, time pressure, success in carrying out tasks, required level of attention to maintain performance, and level of anxiety/discouragement/irritation/annoyance/stress
	Intensity of emotions	Scale of 1 (not at all intense / positive) to 5 (very intense / positive)
	Positivity of emotions	
	Set of questions concerning mode of navigation	Scale of 1 (not at all) to 5 (very strongly): eye strain, physical strain on shoulders/back/legs, nausea, general discomfort, headache, difficulty concentrating, feeling of really walking in the 3D environment
	Set of questions concerning task design	Scale of 1 (not at all) to 5 (very strongly): the task was interesting, the task was repetitive, the mocap equipment interfered with the task, the physiological sensors interfered with the task, the scenes were realistic, feeling of really interacting with the objects
Did you have difficulties understanding or accomplishing the tasks?	open question	
Post-experience	Did you have any difficulties in general?	open question
	Set of questions on presence	Scale of 1 (not at all) to 7 (very strongly): feeling of really being present in the virtual environment, thoughts of being present in the virtual environment, considered the virtual environment to be closer to somewhere you visited before (as compared to an image seen), dominant feeling of being in the virtual scene (as compared to elsewhere), recollections are similar in structure to an actual memory
	Set of questions on emotion	Scale of 1 (not at all) to 47 (very strongly): interested, anxious, excited, annoyed, irritated, enthusiastic, alert, inspired, attentive
	UTAUT2	same as in pre-study questionnaire
	Any additional comments or suggestions	open question

5 RESULTS

To achieve this study and result analysis, we follow the methodology in Figure 10. The conception of this study was based on a task model which allows us to efficiently represent the generated scenes, list grids for configuration and scene parameters. Following a pilot study, we converged with partners on a merged grid of final parameters, designing observations forms for the formal study. After the study, behavioural data was then mapped to contextual data for analysis using the scene annotations and user interaction logs generated based on the task model.

We exemplify the potential of this methodology with the collected data in this section by presenting a selection of correlational analysis for UX understanding. We then elaborate the potential for fine-grained analysis of user intentions in scene context and discuss the limitations so far.

We use a number of abbreviations as follows: The two walking conditions are abbreviated as real walking (RW) and simulated walking (SW). The two visual conditions are abbreviated as normal vision (NV) and simulated low vision (LV). The six scenarios are

abbreviated S1-S6 in the order presented in Fig. 5. The EDA (electrodermal activity) signal is the raw measurement of skin conductance in micro-Siemens (uS).

5.1 Demographics and preliminary questionnaire results

In this first stage of studies, we published an open call to the local university and laboratories. Participants needed to have normal or corrected to normal binocular vision, no motor sensory difficulties, but no other restrictions. As a result, a total of 16 people were recruited (14 men and 2 women) between ages 18 and 34. Three participants used VR for the first time during this study, five only used it once or twice before. In the end, all recruited participants were able to complete the entire study.

From the questionnaires results, we made two observations on the study design:

Relation between fatigue and study duration. Despite the study length, participants did not report an increase in the level of fatigue over time, nor any levels of nausea where they required a pause or skipping of conditions. This could be mainly a result of our

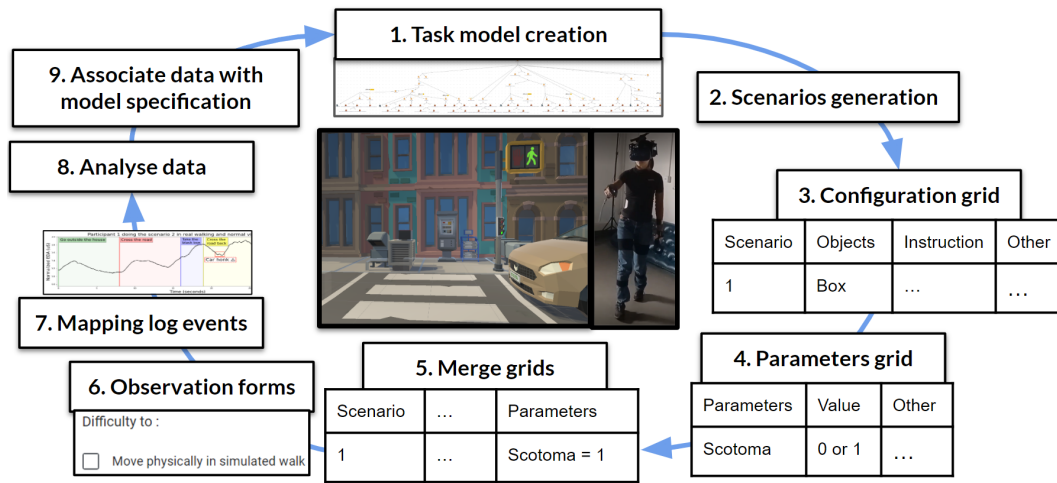


Figure 10: Our methodology to configure and analyse the study involving multiple steps that are first and foremost based on a detailed task model. From the model, we then created annotated scenes, and compiled all the study configuration and parameters. Following pilot studies, we then converged with partners on a final merged parameter grid. After the study, we are then able to generate a mapping between the behavioural data, interaction logs, and the scene context based on the task model, facilitating fine-grained analysis on the embodied experience.

demographics, mainly young (18-35 age range) and mostly men who have shown to be less susceptible to motion sickness.

Relation between condition and cognitive load. On the self-reported scale of 1 (low) to 5 (high) on cognitive load, we observed that the RW+NV condition was clearly the easiest with an average of 2.0, followed by 2.5 for real walking+low vision and 2.8 for SW+NV. Naturally, the SW+LV condition was ranked with the highest cognitive load, with an average of 4.0 on the score.

5.2 Preliminary analysis of embodied interactions

Our large variety of data was synchronised by their Unix Timestamp and processed in a Jupyter notebook. The Python toolkit Neurokit [30] was used to process EDA data, allowing the extraction of the phasic and tonic components. Every person has a very different EDA baseline level, for this reason, we used the normalised EDA, calculated for every person individually by deducting to their EDA data the average EDA of their complete experience. Well known Python libraries were used for data processing, such as Pandas to organise dataframes, NumPy for mathematical functions, and Matplotlib for data visualisation. Combining multimodal data (physiological, motion, gaze, questionnaire, and scenario interaction data) through a notebook allows a clear perception and analysis of the user embodied experience.

Learning curve. As we can see from Figure 11, participants are not faster in carrying out the tasks in the third and fourth conditions they encountered, as compared to the first and second one. Participants are however slower on the last scenarios of the last conditions, which is unexpected. A potential explanation is that, unlike the answers given in the questionnaire, at this moment of the study, participants experienced some fatigue.

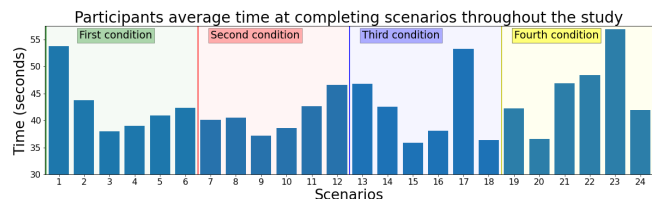


Figure 11: The evolution of the average time spent by all participants over the study across conditions and scenarios in the order executed following the Latin Square assignment. It shows that the participants usually spent more time on the first scenario of each condition than the scenarios that immediately followed, except for the last condition.

Relation between scenario and skin conductance. In the top half of Figure 12, we can see results on the EDA in relation to the scenarios. Based on the original design, shown in Figure 5, the three most emotionally intensive scenarios are supposed to be the third, fifth and sixth. As displayed by the global data, the sixth scenario indeed resulted in higher arousal (0.064 uS), but the second and third highest arousal was observed in S1 (0.106 uS) and S4 (0.232 uS), both scenario with no cars at all. This echoes a comment two participants had, explaining the fact that when no cars are visible, they were not sure if a car would suddenly appear. In a scenario with a car, they always see where is the car, and at what speed it is moving, making it easier for them to plan their action.

Impact of low vision - first view. As shown in the top half of Figure 12, the LV condition had an impact on both the time to do the conditions, and the EDA measured. The impact on speed is small, taking on average less than 1 second more to finish regardless of walking condition – RW (43.3 to 44.1 seconds) and SW (46.8 to 47.2

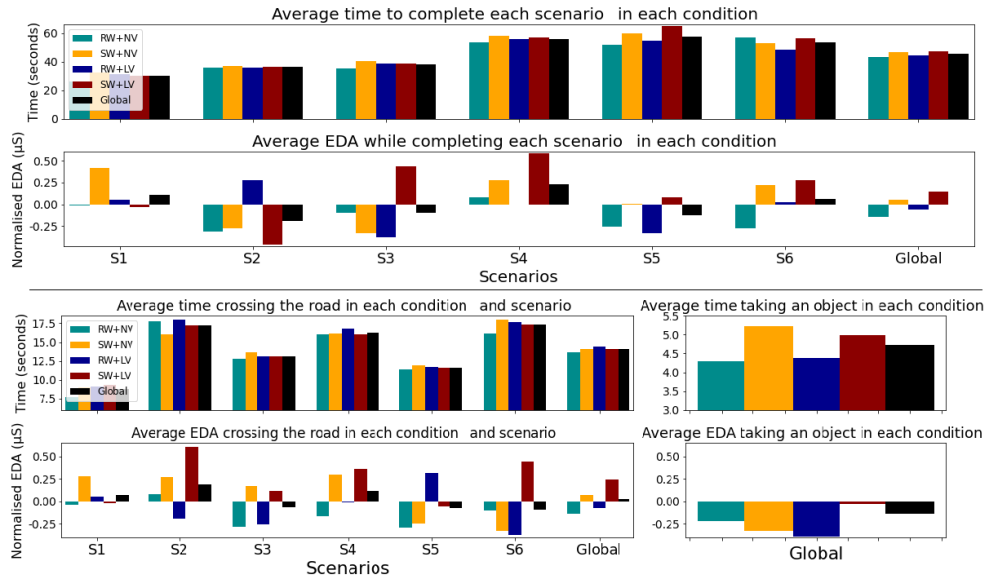


Figure 12: (Top) Average time participants takes to complete scenarios for every condition and the electrodermal activity (EDA) measured during these tasks. (Bottom Left) Average time participants take to cross a road for every condition and the EDA measured during these tasks, (Bottom Right) compared to the average EDA and time to take an object.

seconds). The impact is however more noticeable on EDA, rising from -0.145 to -0.058 μS in RW, and from 0.052 to 0.146 μS in SW. In accordance with this results, participants in the questionnaire said that SW+LV was the most mentally demanding task, and often commented that they had difficulty to handle the joystick while monitoring the road traffic, explaining why this condition is the one resulting in the highest arousal.

Fine-grained task analysis. In the bottom left side of Figure 12, we can see the time to perform crossing a road task. As shown in the global results, the condition generating the highest EDA is the condition SW+LV. It is interesting to note that between the least arousal-generating condition (RW+NV), and the most arousal generating one (SW+LV) in Figure 12 top half (complete scenario) compared to bottom left side (road crossing task), the difference in EDA rises from 0.291 μS to 0.377 μS . This shows that the SW+LV condition has a stronger impact on participant arousal when they have to cross the road than in the rest of the scenario.

In the bottom half of Figure 12, we can see in more detail the impact between a task asking the user to take an object, and the task to cross the road. We can notice that while the road crossing scenario generate much more EDA overall, both have a big increase when going from SW+NV to SW+LV, reported by most participants as the complicated condition, as they have to handle both the walk with the controller and the black dot in the center of the vision, asking for more effort.

Fine-grained emotion analysis. In Figure 13 we can see the evolution of the EDA of one participant across the different tasks of S2, including one car and simple interactions. Most notably, we observed a momentary sharp rising of EDA when they were honked at by a car while jaywalking during a red light.

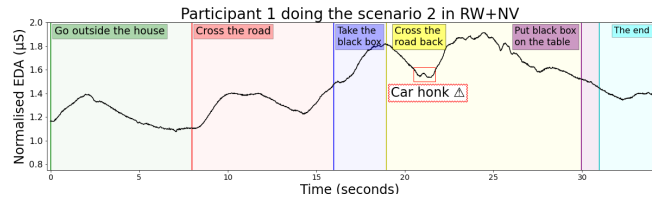


Figure 13: The EDA of participant 1 during the scenario 2 in RW+NV condition, honked by a car while jaywalking during a red traffic light.

5.3 Potential in-depth analysis on embodied experiences

The comprehensiveness of the metrics that were captured in relation to our formalised task model affords many rich directions of analysis from a neuroscience and cognitive science point of view. We noted in particular multiple types of analyses that are yet to be explored:

Contextual mapping. With the use of granular user interaction logs and the task model, we can correlate behavioural metrics including gaze, emotion, and motion to the actual tasks and interactions the user is carrying out, or the stimuli the user is reacting to, such as the example shown in Figure 13. This figure illustrates the workflow’s capacity to allow designers to observe the evolution of a single participant’s emotional intensity over the time of one scenario (5 tasks). It enables a multimodal visualisation of the user experience, combining information regarding the physiological state of the user, and their interactions in the scene.

Cross modality analysis. Previous analysis of user behaviour has often focused on single modalities, with an exception of few recent

works, notably Guimard et al.[17] who investigate the correlation between emotion, gaze, and content. Multimodal analysis will allow us to both better understand the interactions between modalities of user experiences, while appreciating its complexity.

Impact of low vision. Thanks to the possibility of conducting real walking in large spaces, as well as the improvement of eye tracking techniques deployed in VR headsets, we can simulate accessibility constraints such as low vision in order to better understand the impact on patients' everyday lives.

Characterising presence. The question of how to create engaging media content for each individual can help improve the way VR can be used both for entertainment as well as applications in education, rehabilitation, and training. Through correlating questionnaires to emotional responses, we hope to be able to identify factors that can influence the feeling of presence in interactive VR environments.

5.4 Limitations

In terms of experimental setup, we noted difficulties correctly capturing the heart rate of participants due to the sensitivity of the equipment. The sensors were placed at the base of the finger, but would easily be disrupted from slight motions or pressing. Another limiting factor were the demographics of the study participants, mostly being from the similar age and gender groups. Diversifying the recruitment of participants will be a top priority for the follow-up studies.

The focus of the paper is on the framework and its deployment in an actual study, and how the technical platform was established incarnating Dourish's theory. The preliminary analyses we present here are therefore limited, and will be the focus of the next phase of work to investigate the hypotheses posed on the impacts of low vision, the sense of presence, and the potential of VR as a viable tool for training in a clinical setting. Questions on how we can analyse granular data collected from sensors and self-reported metrics in questionnaires will be key to characterising a comprehensive embodied experience, and moving towards a better understanding of user intentions in VR.

6 CONCLUSION

In this work we have established a framework for the design, capture, and analysis for embodied interactions. We validated the feasibility of the framework in a study with the goal to understand the impact of low vision conditions in complex real walking situations, and in the long term to investigate the usage of VR for training and rehabilitation in clinical settings. Our main contribution is the design and encoding of the task model, the implementation of the technical platform to capture of multimodal behavioural indices, and a preliminary presentation of the results to show potential future analyses.

In the continuation of this work, we will evolve our technical platform based on the limitations we observed and conduct a larger follow-up study. The most important next step of work involves the statistical analysis and modelling of the data, in order to investigate hypotheses in cognitive science on the impact of low vision conditions, and from the perspective of human-computer interactions, establish a model to interpret the data in relation to user

intentions. It is our hope and belief that this framework will bring us to yet-unexplored dimensions of embodied experiences in VR.

ACKNOWLEDGMENTS

This work has been partially supported by the French National Research Agency through the ANR CREATTIVE3D project ANR-21-CE33-0001 and UCA^{JEDI} Investissements d'Avenir ANR-15-IDEX-0001 (IDEX reference center for extended reality XR²C²).

We would also like to thank Johanna Delachambre, Clément Quéré, Franz Franco Gallo, and Kateryna Pirkovets for aiding with the user studies, and Marie-Cécile Lafont and Nadia Belfegas for their administrative support in recruitment, equipment purchases, and compensating user studies.

REFERENCES

- [1] Mariano Alcañiz, Beatriz Rey, Jose Tembl, and Vera Parkhulik. 2009. A neuroscience approach to virtual reality experience using transcranial Doppler monitoring. *Presence: Teleoperators and Virtual Environments* 18, 2 (2009), 97–111.
- [2] Mariano Alcañiz, Beatriz Rey, Jose Tembl, and Vera Parkhulik. 2009. A neuroscience approach to virtual reality experience using transcranial Doppler monitoring. *Presence: Teleoperators and Virtual Environments* 18, 2 (2009), 97–111.
- [3] Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (May 2021), 2608–2617. <https://doi.org/10.1109/TVCG.2021.3067783> Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- [4] Domna Banakou, Parasuram D Hanumanthu, and Mel Slater. 2016. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience* 10 (2016), 601.
- [5] Stuart M. Bender and Billy Sung. 2021. Fright, attention, and joy while killing zombies in Virtual Reality: A psychophysiological analysis of VR user experience. *Psychology & Marketing* 38, 6 (2021), 937–947. https://doi.org/10.1002/mar.21444_eprint; <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mar.21444>.
- [6] Mette Boldt, Michael Bonfert, Inga Lehne, Melina Cahnbley, Kim Korsching, Ioannis Bikas, Stefan Finke, Martin Hanci, Valentin Kraft, Boxuan Liu, Tram Nguyen, Alina Panova, Ramneek Singh, Alexander Steenbergen, Rainer Malaka, and Jan Smeddinck. 2018. You Shall Not Pass: Non-Intrusive Feedback for Virtual Walls in VR Environments with Room-Scale Mapping. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Tuebingen/Reutlingen, Germany, 143–150. <https://doi.org/10.1109/VR.2018.8446177>
- [7] Rozenn Bouville, Valerie Gouranton, Thomas Boggini, Florian Nouviale, and Bruno Armaldi. 2015. #FIVE : High-level components for developing collaborative and interactive virtual environments. In *2015 IEEE 8th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)*. IEEE, Arles, France, 33–40. <https://doi.org/10.1109/SEARIS.2015.7854099>
- [8] Lung-Pan Cheng, Eyal Ofek, Christian Holz, and Andrew D. Wilson. 2019. VRoamer: Generating On-The-Fly VR Experiences While Walking inside Large, Unknown Real-World Building Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 359–366. <https://doi.org/10.1109/VR.2019.8798074>
- [9] Rory M.S. Clifford, Sungchul Jung, Simon Hoermann, Mark Billinghurst, and Robert W. Lindeman. 2019. Creating a Stressful Decision Making Environment for Aerial Firefighter Training in Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 181–189. <https://doi.org/10.1109/VR.2019.8797889> ISSN: 2642-5254.
- [10] Rossana De Beni, Francesca Pazzaglia, and Simona Gardini. 2006. The role of mental rotation and age in spatial perspective-taking tasks: when age does not impair perspective-taking performance. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 20, 6 (2006), 807–821.
- [11] Nonny De la Peña, Peggy Weil, Joan Llobera, Bernhard Spanlang, Doron Friedman, Maria V Sanchez-Vives, and Mel Slater. 2010. Immersive journalism: Immersive virtual reality for the first-person experience of news. *Presence* 19, 4 (2010), 291–301.
- [12] Alexandre Delaux, Jean-Baptiste de Saint Aubert, Stephen Ramanoël, Marcia Bécu, Lukas Gehrke, Marius Klug, Ricardo Chavarriaga, José-Alain Sahel, Klaus Gramann, and Angelo Arleo. 2021. Mobile brain/body imaging of landmark-based navigation with high-density EEG. *European Journal of Neuroscience* 54, 12 (2021), 8256–8282.
- [13] Diane Dewez, Ludovic Hoyet, Anatole Lécuyer, and Ferran Argelaguet Sanz. 2021. Towards “Avatar-Friendly” 3D Manipulation Techniques: Bridging the Gap Between Sense of Embodiment and Interaction in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing*

- Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3411764.3445379>
- [14] Patrick Dickinson, Andrew Cardwell, Adrian Parke, Kathrin Gerling, and John Murray. 2021. Diegetic Tool Management in a Virtual Reality Training Simulation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, Lisboa, Portugal, 131–139. <https://doi.org/10.1109/VR50410.2021.00034> ISSN: 2642-5254.
- [15] Paul Dourish. 2004. *Where the action is: the foundations of embodied interaction*. MIT press, USA.
- [16] Pawel Grabarczyk and Marek Pokropski. 2016. Perception of Affordances and Experience of Presence in Virtual Reality. *Avant. The Journal of the Philosophical-Interdisciplinary Vanguard* 7, 2 (2016), 0.
- [17] Quentin Guimard, Florent Robert, Camille Bauge, Aldric Ducreux, Lucile Sassatelli, Hui-Yin Wu, Marco Winckler, and Auriane Gros. 2022. PEM360: A dataset of 360 videos with continuous Physiological measurements, subjective Emotional ratings and Motion traces. In *Proceedings of the 13th ACM Multimedia Systems Conference*. ACM, Athlone, Ireland, 252–258.
- [18] Joshua Peter Handali, Johannes Schneider, Michael Gau, Valentin Holzwarth, and Jan vom Brocke. 2021. Visual Complexity and Scene Recognition: How Low Can You Go?. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, Lisboa, Portugal, 286–295. <https://doi.org/10.1109/VR50410.2021.00051> ISSN: 2642-5254.
- [19] Dan Hawes and Ali Arya. 2021. VR-based Student Priming to Reduce Anxiety and Increase Cognitive Bandwidth. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, Lisboa, Portugal, 245–254. <https://doi.org/10.1109/VR50410.2021.00046> ISSN: 2642-5254.
- [20] Meggy Hayotte, Pierre Th rouanne, Laura Gray, Karine Corri n, and Fabienne d'Arripe Longueville. 2020. The French eHealth acceptability scale using the unified theory of acceptance and use of technology 2 model: instrument validation study. *Journal of medical Internet research* 22, 4 (2020), e16520.
- [21] Dongjin Huang, Xianglong Wang, Jinhua Liu, Jinyao Li, and Wen Tang. 2022. Virtual reality safety training using deep EEG-net and physiology data. *The visual computer* 38, 4 (2022), 1195–1207.
- [22] Crescent Jicol, Chun Hin Wan, Benjamin Doling, Caitlin H Illingworth, Jinha Yoon, Charlotte Headey, Christof Lutteroth, Michael J Proulx, Karin Petrini, and Eamonn O'Neill. 2021. Effects of Emotion and Agency on Presence in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–13. <https://doi.org/10.1145/3411764.3445588>
- [23] Conor Keighrey, Ronan Flynn, Siobhan Murray, Sean Brennan, and Niall Murray. 2017. Comparing User QoE via Physiological and Interaction Measurements of Immersive AR and VR Speech and Language Therapy Applications. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017 - Thematic Workshops '17*. ACM Press, Mountain View, California, USA, 485–492. <https://doi.org/10.1145/3126686.3126747>
- [24] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.
- [25] Ue-Hwan Kim, Jin-Man Park, Taek-jin Song, and Jong-Hwan Kim. 2020. 3-D Scene Graph: A Sparse and Semantic Representation of Physical Environments for Intelligent Agents. *IEEE Transactions on Cybernetics* 50, 12 (2020), 4921–4933. <https://doi.org/10.1109/TCYB.2019.2931042>
- [26] Meghan Kumar, Connor Delaney, and Dean Krusienski. 2022. Estimation of Affective States in Virtual Reality Environments using EEG. In *Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments*. ACM, New York, NY, United States, 396–401.
- [27] Peter Lang. 1980. Behavioral treatment and bio-behavioral assessment: Computer applications. *Technology in mental health care delivery systems* 0, 0 (1980), 119–137.
- [28] Yining Lang, Liang Wei, Fang Xu, Yibiao Zhao, and Lap-Fai Yu. 2018. Synthesizing Personalized Training Programs for Improving Driving Habits via Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, 297–304. <https://doi.org/10.1109/VR.2018.8448290>
- [29] Sarah Lopez, Yi Yang, Kevin Beltran, Soo Jung Kim, Jennifer Cruz Hernandez, Chelsy Simran, Bingkun Yang, and Beste F Yuksel. 2019. Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In *Proceedings of the 2019 CHI Conference on human factors in computing systems*. ACM, New York, NY, United States, 1–12.
- [30] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, Fran ois Lespinasse, Hung Pham, Christopher Sch lzel, and S. H. Annabel Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods* 53, 4 (feb 2021), 1689–1696. <https://doi.org/10.3758/s13428-020-01516-y>
- [31] Marieke AG Martens, Angus Antley, Daniel Freeman, Mel Slater, Paul J Harrison, and Elizabeth M Tunbridge. 2019. It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology* 33, 10 (2019), 1264–1273.
- [32] Maryam Memar and Amin Mokaribolhassan. 2021. Stress level classification using statistical analysis of skin conductance signal while driving. *SN Applied Sciences* 3, 1 (Jan. 2021), 64. <https://doi.org/10.1007/s42452-020-04134-7>
- [33] Deniz Melevio glu, David Murphy, and Sabin Tabirca. 2021. Real-time Anxiety Prediction in Virtual Reality Exposure Therapy. *New York City* 0, 0 (2021), 5.
- [34] Christos Mousas, Dominic Kao, Alexandros Koilias, and Banafsheh Rekabdar. 2020. Real and virtual environment mismatching induces arousal and alters movement behavior. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Atlanta, GA, USA, 626–635.
- [35] T. C. Peck, M. Doan, K. A. Bourne, and J. J. Good. 2018. The Effect of Gender Body-Swap Illusions on Working Memory and Stereotype Threat. *IEEE Transactions on Visualization & Computer Graphics* 24, 04 (apr 2018), 1604–1612. <https://doi.org/10.1109/TVCG.2018.2793598>
- [36] Mary Pidgion, Nadia Kanwal, and Niall Murray. 2022. Federated Learning to Understand Human Emotions via Smart Clothing: Research Proposal. In *Proceedings of the 13th ACM Multimedia Systems Conference (Athlone, Ireland) (MM-Sys '22)*. Association for Computing Machinery, New York, NY, USA, 408–411. <https://doi.org/10.1145/3524273.3533936>
- [37] Aylene Ricca, Amine Chellali, and Samir Otrnane. 2021. The influence of hand visualization in tool-based motor-skills training, a longitudinal study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, Lisboa, Portugal, 103–112. <https://doi.org/10.1109/VR50410.2021.00031> ISSN: 2642-5254.
- [38] Stephanie Schulze, Toni Pence, Ned Irvine, and Curry Guinn. 2019. The effects of embodiment in virtual reality on implicit gender bias. In *Virtual, Augmented and Mixed Reality. Multimodal Interaction: 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21*. Springer, USA, 361–374.
- [39] Sofia Seinfeld, Tiare Feuchtner, Johannes Pinzek, and J rg M ller. 2020. *Impact of Information Placement and User Representations in VR on Performance and Embodiment*. Vol. PP. IEEE, remote. <https://doi.org/10.1109/TVCG.2020.3021342> Journal Abbreviation: IEEE transactions on visualization and computer graphics Publication Title: IEEE transactions on visualization and computer graphics.
- [40] Pavel Ugwitz, Al b ta Sa inkov , Ce ek Sa inka, Zden k Stacho n, and Vojt ch Ju ik. 2021. Toggle toolkit: A tool for conducting experiments in unity virtual environments. *Behavior research methods* 0, 0 (2021), 1–11.
- [41] Viswanath Venkatesh, James YL Thong, and Xin Xu. 2012. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly* 0, 0 (2012), 157–178.
- [42] Jhon Bueno Vesga, Xinhao Xu, and Hao He. 2021. The Effects of Cognitive Load on Engagement in a Virtual Reality Learning Environment. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, Lisboa, Portugal, 645–652. <https://doi.org/10.1109/VR50410.2021.00090> ISSN: 2642-5254.
- [43] Sophie Villenave, Jonathan Cabezas, Patrick Baert, Florent Dupont, and Guillaume Lavou . 2022. XREcho: A Unity plug-in to record and visualize user behavior during XR sessions. In *Proceedings of the 13th ACM Multimedia Systems Conference*. ACM, Athlone, Ireland, 341–346.
- [44] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240.
- [45] Erwin Wu, Mitski Piekenbrock, Takuto Nakamura, and Hideki Koike. 2021. SPin-Pong - Virtual Reality Table Tennis Skill Acquisition using Visual, Haptic and Temporal Cues. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (May 2021), 2566–2576. <https://doi.org/10.1109/TVCG.2021.3067761> Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- [46] Hui-Yin Wu, Florent Robert, Th o Fafet, Brice Graulier, Barth lemy Passin-Cauneau, Lucile Sassatelli, and Marco Winckler. 2022. Designing Guided User Tasks in VR Embodied Experiences. *Proceedings of the ACM on Human-Computer Interaction* 6, EICS (2022), 1–24.
- [47] Tong Xue, Abdallah El Ali, Tianyi Zhang, Gangyi Ding, and Pablo Cesar. 2021. CEAP-360VR: A Continuous Physiological and Behavioral Emotion Annotation Dataset for 360 VR Videos. *IEEE Transactions on Multimedia* 0, 0 (2021), 1–1. <https://doi.org/10.1109/TMM.2021.3124080>