



HAL
open science

Analyzing and understanding embodied interactions in virtual reality systems

Florent Robert

► **To cite this version:**

Florent Robert. Analyzing and understanding embodied interactions in virtual reality systems. Graphics [cs.GR]. Université Côte d'Azur, 2024. English. NNT : 2024COAZ4044 . tel-04913490

HAL Id: tel-04913490

<https://theses.hal.science/tel-04913490v1>

Submitted on 27 Jan 2025

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.

THÈSE DE DOCTORAT

Analyser et comprendre les interactions incarnées en système de réalité virtuelle

Florent ROBERT

Université Côte d'Azur, I3S, SPARKS/Wimmics
Centre Inria d'Université Côte d'Azur, Biovision

**Présentée en vue de l'obtention
du grade de docteur en Informatique
d'Université Côte d'Azur**

Dirigée par : Marco WINCKLER, Professeur,
Université Côte d'Azur

Co-encadrée par : Hui-Yin WU, Chargée
de recherche, Centre Inria d'Université Côte
d'Azur **et par** Lucile SASSATELLI, Profes-
seure, Université Côte d'Azur

Soutenue le : 09 Decembre 2024

Devant le jury, composé de :

Adrien BOUSSEAU, Directeur de recherche,
Centre Inria d'Université Côte d'Azur

Luciana NEDEL, Professeure, Federal University
of Rio Grande do Sul

Jan GULLIKSEN, Professeur, KTH Royal Insti-
tute of Technology

Indira THOUVENIN, Professeure, Université de
technologie de Compiègne

Jean-François TRUBERT, Professeur, Université
Côte d'Azur

**ANALYZING AND UNDERSTANDING EMBODIED INTERACTIONS IN
VIRTUAL REALITY SYSTEMS**

*Analyser et comprendre les interactions incarnées en système de réalité
virtuelle*

Florent ROBERT



Jury :

Président du jury

Adrien BOUSSEAU, Directeur de recherche, Centre Inria d'Université Côte d'Azur

Rapporteurs

Luciana NEDEL, Professeure, Federal University of Rio Grande do Sul

Jan GULLIKSEN, Professeur, KTH Royal Institute of Technology

Examineurs

Indira THOUVENIN, Professeure, Université de technologie de Compiègne

Directeur de thèse

Marco WINCKLER, Professeur, Université Côte d'Azur

Co-encadrante de thèse

Hui-Yin WU, Chargée de recherche, Centre Inria d'Université Côte d'Azur

Co-encadrante de thèse, Membre invitée

Lucile SASSATELLI, Professeure, Université Côte d'Azur

Membre invité

Jean-François TRUBERT, Professeur, Université Côte d'Azur

Florent ROBERT

Analyzing and understanding embodied interactions in virtual reality systems

x+116 p.

Analyzing and understanding embodied interactions in virtual reality systems

Summary

Embodied interactions allow the creation of realistic and immersive 3D storytelling scenarios by giving users the possibility to engage with virtual environments through the use of their physical bodies, including gestures, posture, and locomotion. Nonetheless, new methodologies to capture and analyze experiences are required for the case of embodied VR experiences, because of the use of the body as a communication device, stimulating the brain and enabling reactions differently than traditional media. Communication problems are also likely to arise between the experience sought by the designer and the experience lived and interpreted by the user, due to the variety of scenarios and interaction modalities that make up a realistic embodied scenario. To address these challenges, this thesis contributes to the observation, analysis, and comprehension of the lived embodied experiences through the conception of frameworks, models, and methodologies, a subject that has not yet been fully explored in the literature applied to VR.

Paul Dourish's theory of embodiment highlights three key elements to take into account for an efficient comprehension of embodied interaction. Inspired by this theory, this thesis is composed of three components : (1) The ontology component, presenting an embodied experience creation framework, to address the creation, the management, and the observation of experiences using embodied interactions in immersive environments ; (2) the intentionality component, presenting an embodied interaction analysis methodology, to address the comprehension of the embodied interactions through the analysis of the correlations between the user behavioral data and the system contextual data ; and (3) the intersubjectivity component, presenting a designer-system-user communication model in embodied context, to address the communication issues existing between a designer building an embodied experience and a user living this experience. Each component represents a chapter of this thesis, taking a user-centered approach to support designers in their understanding of the embodied experience lived in an immersive context by proposing tools inspired by existing works in human-computer interactions (HCI). To validate our tools, three user studies were conducted. The first study evaluates the system implementation of the proposed creation framework, with participants who possess expertise in 3D experience development. The second study enables embodied interaction observation and analysis by collecting participant behavioral metrics (e.g., body motion, skin conductance) while they perform embodied interactions in VR. The third investigates the communication gulf existing between the designer and the user of an embodied experience by observing and classifying the issues happening during the design and the usage phases of the experience. This thesis can support research that aims at a better understanding of human behavior or the improvement of user experience (e.g., health, HCI) by submitting tools that can support designers in acquiring a better understanding of how embodied experiences are lived in VR.

Keywords: Embodied interactions, Virtual reality, User behavior, Ontology, Intersubjectivity, Intentionality

Analyser et comprendre les interactions incarnées en système de réalité virtuelle

Abstract

Les interactions incarnées permettent la création de scénarios narratifs 3D réalistes et immersifs en donnant aux utilisateurs la possibilité d'interagir avec des environnements virtuels grâce à l'utilisation de leur corps physique, notamment par les gestes, la posture et le déplacement. Néanmoins, de nouvelles méthodologies de capture et d'analyse des expériences sont nécessaires pour le cas des expériences VR incarnées, en raison de l'utilisation du corps comme dispositif de communication, stimulant le cerveau et permettant des réactions différentes des médias traditionnels. Des problèmes de communication sont également susceptibles de survenir entre l'expérience recherchée par le concepteur et l'expérience vécue et interprétée par l'utilisateur, en raison de la variété des scénarios et des modalités d'interaction qui composent un scénario incarné réaliste. Pour relever ces défis, cette thèse contribue à l'observation, à l'analyse et à la compréhension des expériences incarnées vécues à travers la conception de framework, de modèles et de méthodologies, un sujet qui n'a pas encore été pleinement exploré dans la littérature appliquée à la RV.

La théorie de l'incarnation de Paul Dourish met en évidence trois éléments clés à prendre en compte pour une compréhension efficace de l'interaction incarnée. Inspirée de cette théorie, cette thèse est composée de trois composantes : (1) La composante ontologie, présentant un framework de création d'expérience incarnée, pour aborder la création, la gestion et l'observation d'expériences utilisant des interactions incarnées dans des environnements immersifs ; (2) la composante intentionnalité, présentant une méthodologie d'analyse des interactions incarnées, pour aborder la compréhension des interactions incarnées à travers l'analyse des corrélations entre les données comportementales de l'utilisateur et les données contextuelles du système ; et (3) la composante intersubjectivité, présentant un modèle de communication concepteur-système-utilisateur en contexte incarné, pour spécifier les problèmes de communication existant entre un concepteur construisant une expérience incarnée et un utilisateur vivant cette expérience. Chaque composante représente un chapitre de cette thèse, adoptant une approche centrée utilisateur pour soutenir les concepteurs dans leur compréhension de l'expérience incarnée vécue dans un contexte immersif en proposant des outils inspirés des travaux existants dans les interactions homme-machine (IHM). Pour valider nos outils, trois études utilisateurs ont été menées. La première étude évalue la mise en œuvre du framework de création proposé, avec des participants possédant une expertise dans le développement d'expériences 3D. La deuxième étude permet l'observation et l'analyse des interactions incarnées en collectant des mesures comportementales des participants (par exemple, le mouvement du corps, la conductance de la peau) pendant qu'ils effectuent des interactions incarnées en RV. La troisième étudie le fossé de communication existant entre le concepteur et l'utilisateur d'une expérience incarnée en observant et en classant les problèmes qui se produisent pendant les phases de conception et d'utilisation de l'expérience. Cette thèse peut soutenir la recherche qui vise à une meilleure compréhension du comportement humain ou à l'amélioration de l'expérience utilisateur (par exemple, la santé, l'IHM) en soumettant des outils qui peuvent aider les concepteurs à acquérir une meilleure compréhension de la façon dont les expériences incarnées sont vécues en RV.

Keywords: Interaction incarnées, Réalité virtuelle, Comportement utilisateur, Ontologie, Intersubjectivité, Intentionnalité

Acknowledgements

I would like to extend my deepest gratitude to my three supervisors, Marco Winckler, Hui-Yin Wu, and Lucile Sassatelli, for their guidance and encouragement throughout this thesis journey. Their expertise and patience have been essential to shape both my research and my development as a doctoral student. I am deeply grateful for the time they invested in helping me navigate the multifaceted world of research.

Additionally, I would like to thank the two research teams I had the privilege to be part of, the Biovision and Wimmics teams from INRIA. I am truly grateful for the intellectual exchange that supported our research efforts and greatly contributed to the enjoyment in performing this work. I also express my appreciation to my fellow doctoral students, Clément Quere, Johanna Delachambre, Pauline Devictor, Franz Gallo, and Quentin Guimard. Working and chatting with you has been an enriching experience.

Contents

1	Introduction	1
1.1	Theoretical background	2
1.2	Motivation	3
1.3	Challenges	4
1.4	Contributions	5
1.4.1	Engineering of ontology systems	6
1.4.2	Analysis of Intentionality from large immersive dataset	6
1.4.3	Investigating inter user-designer intersubjectivity communication	7
2	State of the Art	9
2.1	Management of immersive experiences	9
2.1.1	Building semantically rich interactive 3D environments	9
2.1.2	Building studies enabling embodied interactions and analysis	12
2.2	Understanding user experiences in virtual reality	13
2.2.1	Building datasets of behavioral and contextual data	13
2.2.2	Behavior analysis of user experience in context	13
2.2.3	Task-based analysis outside of VR	14
2.3	Improving designer-user communication mediated by systems	15
2.3.1	Designer’s intention communication through a product	15
2.3.2	The designer-user communication in virtual reality	17
2.4	Summary	18
Contributions		
3	Engineering ontology systems	21
3.1	GUsT-3D : Framework for designing guided user tasks in embodied VR experiences	21
3.1.1	GUTask workflow and system implementation	23
3.1.2	Case studies	29
3.1.3	Evaluation	33
3.1.4	Discussion	36
3.2	Framework application for the observation and understanding of multimodal embodied VR experiences	36
3.2.1	Methods and Materials	37
3.2.2	Study	40
3.2.3	Discussion	48
3.3	Conclusion	48
4	Analysis of Intentionality from large immersive dataset	51
4.1	PEM360 : A dataset with emotion, motion and attention data	52
4.1.1	User experiment and dataset description	52
4.1.2	Dataset structure	53
4.1.3	Pre-processing software	53
4.1.4	Preliminary analysis of the data	54
4.1.5	Discussion	56

4.2	CREATTIVE3D : Exploring, walking, and interacting in virtual reality with simulated low vision	57
4.2.1	User experiment and dataset description	57
4.2.2	Dataset structure	57
4.2.3	Technical Validation	58
4.2.4	Pre-processing and analysis software	59
4.2.5	Preliminary analysis of the data	59
4.2.6	Discussion	61
4.3	Task-based methodology to characterize immersive user experience with multivariate data	62
4.3.1	Dataset	63
4.3.2	Task model	65
4.3.3	Results	70
4.3.4	Discussion	72
4.4	Conclusion	73
5	Investigating inter user-designer intersubjectivity communication	75
5.1	Overview	75
5.2	The triangle of misunderstanding in interactive virtual narratives : gulfs between system, designers and players	76
5.2.1	VR communication model	77
5.2.2	Methodology	78
5.2.3	Structured interview	78
5.2.4	User study	79
5.2.5	Results	81
5.2.6	Correlation between issues	86
5.2.7	Post-test insight	87
5.2.8	Discussion	88
5.3	Conclusion	89
6	Conclusion	91
6.1	Contributions	91
6.2	Limitations	92
6.3	Future work	93
6.3.1	Rich experience management and creation	93
6.3.2	Granular experience visualization	94
6.3.3	Deeper embodied experience analysis	94
6.3.4	Designer-user communication improvement	94
6.4	Summary	95
	Publications	97
	References	99
	List of Figures	111

CHAPTER 1

Introduction

As early as of 1997, [Biocca et al.](#) discussed the tendency of virtual reality (VR) media to evolve toward a tighter coupling of the body and the interface, toward a higher level of embodiment, using the body as a communication device to optimize the human-system communication. VR embodied interactions increase narrative and interactive possibilities by allowing users to engage with the virtual environment using their physical bodies, through gesture, posture, and locomotion. Embodied possibilities made VR media attractive, nowadays used in several applications and research domains to tell stories ([Hardie et al., 2020](#)), communicate ideas ([Wen & Gheisari, 2020](#)), support artistic projects ([Yu et al., 2024](#)), or simulate life-like situations ([Xie et al., 2021](#)).

However, VR media is still developing, with major evolutions every year as the media faces challenges in delivering an immersive embodied experience in contrast to traditional media. Research in progress explore ways to improve the feeling of presence ([Schuemie et al., 2001](#) ; [V. Souza et al., 2021](#)) or immersion ([Slater, 2018](#)), to improve the interactions or locomotion ([Van Gemert et al., 2024](#)), to improve the usage of resources in order to increase the screen resolution, the field of view ([Mantiuk et al., 2021](#)), the streaming quality ([Sassatelli et al., 2020](#)), or to solve adverse effects like the cyber sickness ([Pöhlmann et al., 2024](#) ; [Qiu et al., 2023](#) ; [Rebenitsch & Owen, 2021](#) ; [Weech et al., 2019](#)) and eyestrain ([Hirzle, 2023](#) ; [Schroeder & Biggs, 2023](#)) generated by VR usage. All of these challenges are very important for VR and must be addressed to improve the user acceptance ([Sagnier et al., 2020](#)) : the experience can be immersive and high-resolution, but if it generates motion sickness, if there is no perceived usefulness, or if the interaction controls are inaccessible to a certain population, users and designers might prefer to avoid VR media.

All of these challenges share one similarity : **User experience (UX) defines what needs to be removed, changed, or refined in order to create a better immersive experience.** Understanding UX in embodied media, however, differs from traditional media : in embodied context, the body can be seen as a semiotic vehicle that represents a person's mental states (e.g., emotions, observations, plans) ([Benthall, 1973](#) ; [Biocca, 1997](#)). Embodied interactions offered by VR are accompanied by multisensory engagement (e.g., sight, touch, sound) as well as rich interaction possibilities (e.g., physical movement, object manipulation, 3D interaction, visual exploration). These specificities of embodiment, using the body as a communication device, open the possibility of obtaining a better comprehension of the UX through the observation of behavioral metrics related to embodied interactions (e.g., hand gesture, gaze movement, body motion, heart rate).

This thesis contributes in three chapters to the creation, observation, and comprehension of the UX in embodied immersive contexts through the conception of tools : (1) conception of a framework to create and perform embodied interactive studies in rich, annotated 3D contexts ; (2) conception of an analysis methodology using multimodal metrics to enable the characterization of embodied experiences ; and (3) definition of a communication model on designer-user communication issues in embodied contexts.

1.1 Theoretical background

Research in neuroscience (Alcañiz et al., 2009) has shown that interactions in VR stimulate the brain in ways comparable to real-life, physical interactions. This makes VR particularly promising for applications and research in many domains including edutainment (J. Zhao & Klippel, 2019; Wu et al., 2021), rehabilitation (Sveistrup, 2004; W. Li et al., 2020), therapy (Lundin et al., 2022), or training (Clifford et al., 2019; Menin et al., 2022; Nordin Forsberg et al., 2023) for creating meaningful daily experiences.

However, using embodied experiences for research purposes presents multiple challenges. Good communication between the designer, the system, and the user is an important factor in enabling the use of personalized applications. By allowing designers to understand the user's interpretation and difficulties, they can more efficiently support them, which is a major component for training and rehabilitation purposes. To define how to analyze embodied interactions and enable good communication between the three actors (designer, system, and user) of embodied immersive experiences, we used the work of Paul Dourish (2004) and Don Norman (2013), both of which have the strong advantage of being comprehensive theories characterizing the interactions and perception gaps that take place between the actors and applicable to the context of embodied experiences in VR. In order to reduce the gaps in perception between the three actors, Dourish, one of the most influential researchers in the domain of embodied interactions applied to research in HCI, presented the theory of embodiment. This theory identifies three major aspects of meaning, each playing a different role in embodied interaction and raising different sets of issues for interactive system design :

- **Ontology** deals with how the world can be described as collections of entities with established meanings that can be related to each other. However, there isn't a single and static ontology to describe a system : the user, the system, and the designer of an experience might not share the same ontological model of the domain. In the case of immersive experiences, this can create a gap of comprehension between the representation of the 3D scene shown by the system and the interpretations done by the user and the designer.
- **Intentionality** deals with the meaning behind the user performing an embodied interaction, which is required to understand the intention the user has when using a system. Not understanding this meaning in immersive experiences could result in a gap between the user's intentions and interpretation when living the experience and the designer's comprehension of this lived experience.
- **Intersubjectivity** deals with how the meaning can be communicated by different actors, mediated by the system. The designer must, through the system, appropriately reveal the goals of its design and the ways in which it is intended to be used. An unsuccessful communication can result in a gap in the understanding from the designer to the user when communicating the task goals and constraints of the 3D environment through the system.

This theory of embodiment highlights the issues related to the three aspects of meaning (ontology, intentionality, and intersubjectivity) that must be addressed to prevent the designer-system-user gaps, in order to support a better understanding of embodied experiences.

To further investigate the topic of the designer-system-user communication gaps, the seminal work of Norman (2013) on *interaction gulfs* clearly illustrates what happens when users interact and interpret the signs produced by a system. The seven stages of action theory refer to the way to exchange with a system in seven stages, separated into two phases : the execution phase, during which the user will form an intention and interact with the system, and the evaluation phase, during which the user will examine the result of their action and compare it to the initial intention, as shown in Figure 1.1. It highlights four distances between the user and the system that can be a source of incomprehension or inefficient usage when interacting with a system. This theory allows designers to specify the different stages composing the intention-interaction-interpretation process, highlighting two key gulfs in the usage of a system : the execution gulf (i.e., the gap between the initial intention and the performed actions) and the evaluation gulf (i.e., the gap between what the system shows and what the user expects). By breaking down the interaction process into distinct stages, this theory improves the interactive system by exposing the potential issues behind a given stage. This allows the identification of the impact of an issue on either the good usage or the

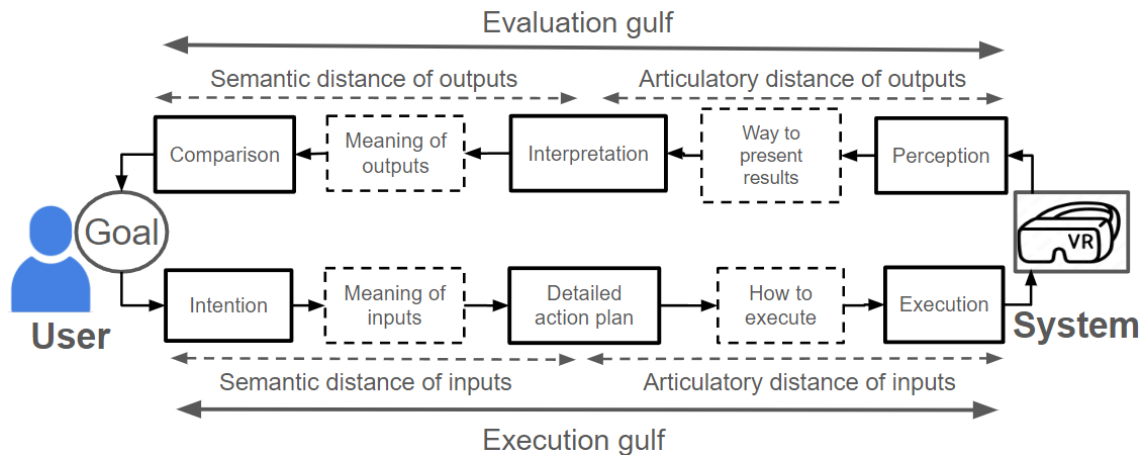


Figure 1.1 – The seven stage of action theory by D. Norman, composed of seven steps : (1) defining the goal to achieve, (2) defining the intention to interact with the system to achieve the goal, (3) defining the detailed plan to achieve the interaction, (4) executing the detailed plan, (5) perceiving the state of the system after executing the detailed plan, (6) interpreting the observed state of the system, (7) comparing the observed changes to the initial goal.

good comprehension of the system, as well as potential guidelines to prevent comparable issues. The theories of D. Norman and P. Dourish are used as a basis for multiple components conceived during this thesis.

1.2 Motivation

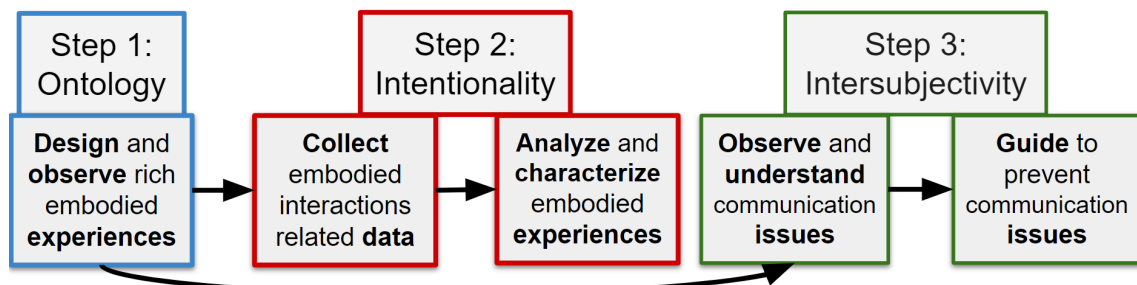


Figure 1.2 – We build systems to support designers of embodied experiences across three components. First, we propose systems to help with the design and observation of embodied experiences. Using this first system, we then construct frameworks to collect and analyze data to enable a deeper comprehension of embodied interactions. Finally, we introduce a methodology to understand designer-user communication issues and ways to guide designers and users in the design and usage phases of an embodied experience to avoid said issues. All systems are conceived using a user-centered approach, with user studies involving designers and users of embodied experiences to evaluate the system's usability.

This thesis aims to reduce the communication gap existing between the designer of an embodied experience, creating an experience with a design intention, and the user of the experience, building their own interpretation of the experience. The purpose is to help designers of embodied experiences obtain a better comprehension of the lived UX and the communication issues happening, to support the refinement of the experience. To achieve this communication gap reduction, our adopted methodology is to build systems that can support designers in the comprehension and reduction of communication gaps in the three

components presented previously : ontology, intentionality, and intersubjectivity, as shown on Figure 1.2. While this is a subject already widely explored in HCI research, it's still in a relatively young state of research in the embodied experience context. We position the thesis specifically for supporting designers for **creating, analyzing, and understanding embodied experiences in the context of virtual reality multitask embodied scenarios**. We present multiple frameworks, methodologies, and models, enabling the following characteristics :

- **Experience creation** : Creation of a system based on a modular ontology to create understandable 3D environments, interaction properties, and tasks of immersive scenarios
- **Study creation** : Creation of reproducible VR study framework in 6DoF enabling the usage of embodied interactions (e.g., 3D interaction, physical walk) in the virtual environment
- **Data collection** : Creation of study enabling the collection and temporal synchronization of multiple behavioral metrics (e.g., gaze motion, body motion, skin conductance) related to embodied interactions
- **Dataset construction** : Creation of a publicly available dataset containing multimodal data to enable fine-grained analysis of behavioral metrics in correlation with multitasks scenarios
- **Behavior analysis** : Conception of a virtual reality task-based methodology inspired by HCI research to characterize user behavior and define user behavior profiles in an embodied VR context
- **Communication issues comprehension** : Creation of a communication model specifying a classification of design and usage communication issues happening in the embodied VR context.

For each contributions, we take a user-centered approach. Actors, both designers and users, were involved in all processes of the tool's development through user studies, to evaluate the embodied experience creation usability, the multimodal data collection, or the observation of designer-user communication issues.

1.3 Challenges

To achieve the work presented on the three components (ontology, intentionality, intersubjectivity), we identified issues existing in the VR literature to work on :

Embodied experience creation using ontology : In immersive environments, a tedious process is represented by the creation of controlled experiences using a clear vocabulary defining the objects and interactive properties. We defined an **ontology issue** as an observed gulf between the system and an actor (designer, user). Ontology issues happen when the actor does not interact properly with the system or misinterprets a message communicated by the system. A vocabulary shared among designers and users is key to avoiding ontology issues. To verify the good communication between the actors, the solution commonly used in HCI works is to capture related contextual data (e.g., attention, interaction, task comprehension) to evaluate the quality of the designer-user communication. In a scenario in which a user misunderstands what the 3D object "black box" is or how to interact with it, we will observe a gulf between what the system expects and what the user actually performs. A vocabulary specifying clearly the rules to exchange with the 3D environments could prevent such issues, and support a better tracking of the objects and interactions. On the other hand, the collection of user data such as gaze and interaction could enable fine-grained analysis of user behavior to understand the issues encountered (e.g., the user didn't see the box, doesn't use the controller's joystick properly, is too stressed). These issues could then be addressed through guidance and experience-based refinement. We identified two key issues regarding this subject in the VR literature : (1) the lack of workflows using ontology to assist designers in the creation of rich embodied experiences with a clear communication of the properties to the actors ; and (2) the lack of framework to support designers in the creation of VR studies enabling execution and granular observation of embodied interaction.

Embodied interaction analysis for intentionality understanding : Granular analysis of the behavior is a key in obtaining a comprehension of the embodied experience lived by the user. We defined an **intentionality issue** as an observed gulf between the user experience and the designer's observations. Intentionality issues happen when a designer misinterprets the user's lived experience. However, understanding this user experience in the case of embodied context without a methodology is very demanding, due to the need to

observe and analyze temporally synchronized multimodal data (e.g., motion, emotion, attention, interaction, context) to enable the analysis of the behavior in correlation with the performances. In a scenario in which a stressed user tries to rapidly perform a task, the designer might misinterpret this behavior as overconfidence or boredom. The designer might need to analyze the mental state of the user (e.g., skin conductance, heart rate) in correlation with the context (e.g., interactions, attention) to fully understand the situation in which the user is. We identified two key issues regarding this subject in the VR literature : (1) the lack of openly available large multimodal datasets in embodied environments enabling comparable analysis, and (2) the lack of methodologies supporting works for granular analysis of multimodal datasets applied to the case of VR embodied experiences.

Designer-User intersubjective communication mediated by embodied experience : The issues related to communication between the designer and the user of an experience through a system are a common subject in HCI research. We defined an **intersubjectivity issue** as an observed gulf between the designer’s intention and the user’s interpretation. Intersubjectivity issues happen when the designer, through the system, doesn’t successfully communicate an idea to the user. A user’s misunderstanding of the experience conditions during a research study can have a negative impact on the designer’s experience comprehension by creating a mismatch between what the designer expects to communicate and observe and what the user actually interprets. For example, a sound signal intended as guidance could be interpreted by the user as a signal indicating a warning or error that is meant to increase the level of stress. As the interpreted experience doesn’t correspond to the intended experience, such issues will impact the investigation of the research questions explored by the designer of the study. For this reason, models that can support designers in identifying the communication issues happening during the complete design-usage process could strongly help the refinement of embodied experiences. We identified one major issue in the literature on this subject : while models widely relevant for understanding communication issues exist in HCI (Norman, 2013) and Semiotic engineering research (C. S. D. Souza & Leitão, 2009), these works haven’t explored in detail the potential communication issues associated with the design and usage of embodied immersive experiences.

1.4 Contributions

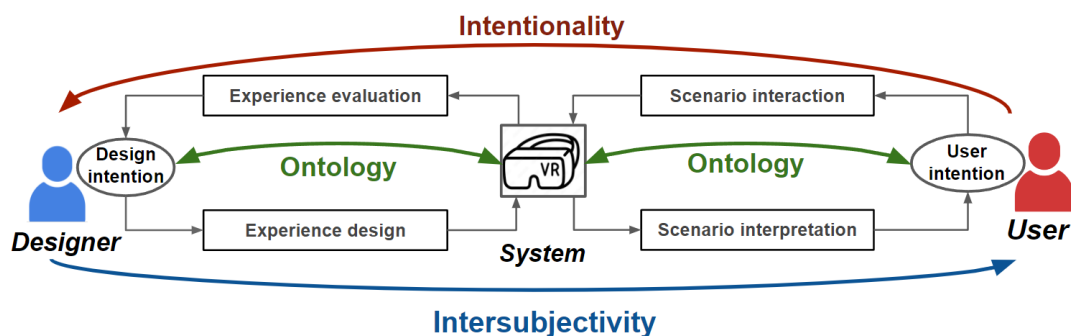


Figure 1.3 – This model represents the designer-system-user relationship when creating and using embodied experiences in VR. The left side represents the design process between the designer and the system, while the right side represents the usage process between the user and the system. We specify three components to address to avoid the creation of communication gulfs between the actors : the ontology, intersubjectivity, and intentionality.

This thesis is structured around a model representing the designer-system-user communication in embodied immersive experiences, shown on Figure 1.3 (inspired by the presented theories of P. Dourish and D. Norman theories). The model highlights three steps in the design process : (1) the designer will form a design intention ; (2) create an experience based on the intention ; and (3) evaluate the resulting experience to compare it with the initial intention. On the usage side, we also observe three steps : (1) the

user will establish an intention to interact with the system ; (2) perform the interaction with the system ; and (3) interpret the consequences of the interaction on the 3D environment compared to the initial intention. During the design and the usage process of the embodied experience, we specify issues with the potential to create a gulf between designer and user by affecting the comprehension in three categories presented previously : ontology issues, intentionality issues, and intersubjectivity issues.

In an embodied immersive experience, the three categories represent subjects we have to address if we want to avoid communication gulfs between designers and users. Based on these observations, we defined three chapters for this thesis, addressing each of the three categories of issues, to achieve the aim of reducing the designer-system-user gulf and provide designers tools to support their comprehension of embodied interactions in virtual reality environments : (1) engineering of ontology systems ; (2) analysis of intentionality from a large immersive dataset ; and (3) investigating inter user-designer intersubjectivity communication.

1.4.1 Engineering of ontology systems

The first contribution is a framework to create understandable interactive environments and enable the collection of behavioral and contextual metrics. The purpose is to better communicate the immersive environment properties to reduce the ontology issues, as well as to open up in-depth analysis possibilities for the next chapter through metrics collection. This chapter is structured around the research question 1 :

- **RQ1 : How can we support designers in the creation and observation of embodied interactions in rich, annotated 3D contexts ?**

To address RQ1 of this thesis, we conceive a framework to support the creation, management, and analysis of rich embodied VR experiences. The framework *Guided Users Task for 3D Environments* (GUsT-3D) allows the design of 3D experiences using a domain specific language (DSL) to create tasks and annotate entities composing the environments with positional and interactive properties. This framework is presented in detail in the section 3.1. Using the GUsT-3D framework as a basis, we then designed a technical platform to perform an embodied immersive study in diverse conditions with the specificity to enable the collection of synchronized behavioral and contextual metrics related to embodied interactions. This framework is presented in more detail in the section 3.2.

1.4.2 Analysis of Intentionality from large immersive dataset

The second contribution is the conception of large immersive datasets as well as our conceived analysis methodology to characterize the user's experience. The aim of this analysis methodology is to provide designers with a better understanding of the user embodied experience and reduce the intentionality issues. This chapter is structured around the research question 2 :

- **RQ2 : How can we analyze intentionality from the user's embodied interactions in immersive experiences using multimodal metrics, and how can this analysis be used to provide feedback to designers ?**

To address RQ2, we use the presented work on system engineering to build two publicly available large immersive datasets : the PEM360 dataset and the CREATTIVE3D dataset. The first dataset, PEM360, consists of participants watching 360-degree videos in VR with a capture and analysis of the emotional and attentional data, presented more in detail in Section 4.1. The second dataset, CREATTIVE3D, consists of participants performing road crossing tasks in VR, with a capture of emotion, attention, motion and contextual data, presented more in detail in Section 4.2. To characterize the embodied experiences observed in the CREATTIVE3D dataset, we created a task-based methodology, inspired by works from HCI research, to analyze embodied interactions on a granular level using multimodal data, presented more in detail in Section 4.3.

1.4.3 Investigating inter user-designer intersubjectivity communication

The third contribution is a communication model to support the specification of a taxonomy of communication issues happening during the design and usage of embodied experiences. The aim is to support designers in understanding and preventing communication issues during both phases to reduce the intersubjectivity issues. This chapter is structured around the research question 3 :

- **RQ3 : How can we specify the intersubjectivity gap between design intention and actual usage in embodied interactive experiences, and how can we use this specification to provide feedback to the designers ?**

To investigate RQ3, we build a communication model inspired by D. Norman’s seven stages of action theory, which describe how a user interacts with a system. This model specifies the issues that can happen when a designer creates an experience (e.g., design issue, evaluation issue) and when a user engages with the experience (i.e., performance issue, comprehension issue). We performed a qualitative study with participants that implemented the full designer-user workflow to collect and classify communication issues and their impact on designer-user communication. This work is presented more in detail in Section 5.2.

Each research question is investigated in a chapter of this thesis through the creation of frameworks, datasets, methodologies, and models, followed by user studies in immersive environments to validate the feasibility and the efficiency of the proposed solutions.

Engineering of ontology systems contributions :



F. Robert, H.-Y. Wu, L. Sassatelli, and M. Winckler. “An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual Reality”. ACM IMX 2023, Nantes, France. **Best Paper Award** 🏆 [hal-04102737](https://hal.archives-ouvertes.fr/hal-04102737)

H.-Y. Wu, **F. Robert**, T. Fafet, B. Graulier, B. Passin-Caneau, L. Sassatelli, M. Winckler. “Designing guided user tasks in VR embodied experiences”. ACM EICS 2022, Biot, France. [hal-03635452](https://hal.archives-ouvertes.fr/hal-03635452)



GUsT-3D Framework : <https://project.inria.fr/creative3d/gust-3d/>

Analysis of Intentionality from large immersive dataset contributions :

F. Robert, H.-Y. Wu, L. Sassatelli, and M. Winckler. “Task-based methodology to characterise immersive user experience with multivariate data”. IEEE VR 2024, Orlando, Florida. [hal-04446066](#)

H.-Y. Wu, **F. Robert**, F. Franco Gallo, K. Pirkovets, C. Quéré, J. Delachambre, S. Ramanoël, A. Gros, M. Winckler, L. Sassatelli, M. Hayotte, A. Menin, P. Kornprobst. “Exploring, walking, and interacting in virtual reality with simulated low vision : a living contextual dataset”. (Under review for Nature Scientific Data). [hal-04429351](#)

Q. Guimard, **F. Robert**, C. Bauce, A. Ducreux, H.-Y. Wu, L. Sassatelli, M. Winckler, A. Gros. “PEM360 : a dataset of 360° videos with continuous physiological measurements, subjective emotional ratings and motion traces”. MMSys 2022, Athlone, Ireland. [hal-03710323](#)

Q. Guimard, **F. Robert**, C. Bauce, A. Ducreux, H.-Y. Wu, L. Sassatelli, M. Winckler, A. Gros. “On the link between emotion, attention and content in virtual immersive environments”. MMSys 2022, Athlone, Ireland. [hal-03825059](#)



CREATTIVE3D dataset : <https://zenodo.org/records/10406560>

PEM360 dataset : <https://gitlab.com/PEM360/PEM360/>

Investigating inter user-designer intersubjectivity communication contribution :

F. Robert, H.-Y. Wu, L. Sassatelli, and M. Winckler. “The triangle of communication in interactive virtual narratives : gulfs between designers, system and players”. (Under review for SIGCHI 2025).

CHAPTER 2

State of the Art

Many systems for creating experiences proposing new interactive and narrative possibilities have emerged in recent decades. Virtual reality (VR) is one of the most recent evolving every year, and seeing uses from entertainment (Wu et al., 2021) to medical rehabilitation (Sveistrup, 2004). To create and refine a system according to user needs, a wide range of HCI and semiotic engineering models and methodologies exist, allowing the observation and improvement of communication done through a system's interface. The usage of such models and methodologies to improve the understanding and refinement of experiences lived in embodied VR context is, however, a subject still in development.

The aim of this chapter is to review the work that has made fundamental contributions to the understanding of user experience in interaction with a system, with a particular emphasis on VR systems. The following sections review work in relation to the three main contributions of this thesis : embodied experience creation, observation, analysis and communication. We first review the work done to create, manage and observe immersive experiences in Section 2.1 ; we then review works on the in-depth analysis of user experience and behavior in immersive contexts and the existing dataset enabling such analysis in Section 2.2 ; and finally, we review works exploring the designer-user communication mediated by a system and ways to improve this communication in Section 2.3.

2.1 Management of immersive experiences

Creating workflows for semantically rich 3D environments in an immersive context is a complex but important task. An environment with defined object and interaction properties can support designers in the creation, but also in the observation on a granular level of the rich interactions composing scenarios. To enrich the object and interaction properties, works usually create ontologies, domain-specific languages, or scene graphs to represent the environment and interactive possibilities relevant to designers and users of a 3D experience. In this section, we explore the subject of embodied experience management through two different directions : (1) the enrichment of the environment and interactions to support designers on the creation and observation of complex experiences as well as the actor-system communication ; and (2) the creation of rich immersive experiences enabling the execution and observation of interactions in the embodied context.

2.1.1 Building semantically rich interactive 3D environments

To design and manage user experiences in interactive virtual environments, as well as to collect data that can support the user experience comprehension, the creation of tools with a fully integrated workflow (i.e., including the phases of creation, validation, execution, and analysis of a VR experience) is an efficient solution. Vergara et al. (2017) built a fully integrated workflow for VR experiences, for the design of applications in engineering education, coupled with suggestions of suitable tools for each step of the workflow. The Unity game engine is one of the most popular applications used to build workflows to support researchers in the design, management and observation of studies in VR, with systems done by Ugwitz et al. (2021) allowing the trigger and the record of various interactive events in the environment, or the work done by Javerliat and Villenave (2022 ; 2024), allowing the execution, collection, and visualization of multimodal data for a more comprehensive view of the user behavior in immersive experiences. In the augmented

reality case, Gottschalk et al. (2020) proposed a product-user-environment framework for e-commerce in the interior design and furniture sector, with a specific focus on the enrichment of the product properties, modeling product features (e.g., textures and components), user models to interact with the product, and the configuration of the product in the environment.

Semantically enriching the environment is a solution often adopted to improve the user-system and designer-system communication, by clarifying the environment and interaction properties and possibilities. A full comparison of the relevant ontologies and frameworks aiming at enriching a part of the conception process are summarized in Table 2.1. The most widely known domain-specific language for 3D content is X3D (Brutzman & Daly, 2010), formerly known as VRML. It is an XML-based syntax that focuses on the information on the 3D model itself (e.g., vertices, texture, material), its 3D coordinates (e.g. position, rotation), and animation as trajectories over a specific time frame. A detailed review of similar ontologies for 3D scenes has been done by Flotyński (2017), but they have a primary interest in 3D modeling applications and not the creation of 3D interactive experiences. A number of extensions to X3D have allowed it to incorporate more contextual, multi-hierarchical, and ontological representations of scenes (Pittarello & De Faveri, 2006), as well as sensors and controller input for virtual applications (Behr et al., 2004). Buffa and Lafon (2000) have also demonstrated its use for an online interactive 3D warehouse application. Another extension by Flotyński et al. (2018) allows the logging of user interactions in VRML applications.

Robotics and procedural content generation domains also explored ontologies as a way to enrich the generation of environments or the granular observation of the robot’s interactions. The RSG-DSL (Blumenthal & Bruyninckx, 2014) was developed in order to represent the robot world model as a scene graph, which allows the attachment of functional blocks to the scene graph to establish the robot’s workflow. Beßler et al. (2020) proposed an ontology for modeling affordance in robotics tasks coupled with the capabilities for logical reasoning and question answering. The Scenic language (Fremont et al., 2019) was developed for machine learning applications in 3D environments to generate datasets from a high-level scene description to train perception systems such as those for autonomous driving. Similarly, Plan-It (Wang et al., 2019) allows the generation of 3D indoor scenes from a high-level scene graph definition. Liu et al. (2019) introduce the idea of *scene programs* that allow the generation of 3D images from layout descriptions and conversely reason about the layouts of objects from images. For game applications, the GIGL language was recently introduced (T. Chen & Guy, 2018) to procedurally generate game maps, such as for dungeons in RPG games.

More precisely applied to the context of virtual reality, ontologies can enrich the 3D objects and interactive properties, in order to support more granular observations of immersive experiences. One of the first formal representations of user interactions was the Interactive Cooperative Objects (ICO) description proposed by Navarre et al. (2005). The work decomposes gestures into a granular graph of states to define selection, rotation, and movement of objects at a low level for specific interactive devices, using a chess game as a case study. Vanacken (2007) introduces the concept of tasks by designing the “concept” datatype that can be attached to objects to label their interaction possibilities. In this work, *task* refers to an interaction technique (such as selecting an object) that is triggered by an event and results in a state change. More recently, a number of ontologies have been proposed for the semantic modeling of virtual environments (Saunier et al., 2016; Chevaillier et al., 2012) with the goal of improving the richness of information about the virtual environment in a database for multi-agent systems.

Another use case of ontologies is to propose ontology-based methods for context representation in interactive 3D applications. One notable example is that of Bouville et al. (2015) who built the system #FIVE on modeling interactive VR environments through the annotation of elements composing the environment with properties, though with a fixed predefined ontology. Kim et al. (2020) 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.

Based on the previous works presenting fully integrated workflows, and ontologies for experience enrichment, we believe designers of embodied experiences could be strongly supported by tools incorporating both components : a fully integrated workflow with a consistent vocabulary throughout all phases of the embodied experience management workflow, including the scene design, annotation, task description, logging, real-time validation of tasks, and post-scenario UX analysis. This would facilitate the smooth

TABLE 2.1 – Summary table for comparing related contributions including their focus (ontology or framework), application scenarios, and other observations. System implementations (with source code) were found for (Beßler et al., 2020; Brutzman & Daly, 2010; Speicher et al., 2021; Raffailac & Huot, 2019; T. Chen & Guy, 2018; Gottschalk et al., 2020).

Focus	Related work	Application scenario	Comments
Ontology	(Behr et al., 2004; Brutzman & Daly, 2010; Buffa & Lafon, 2000; Flotyński & Sobociński, 2018; Pittarello & De Faveri, 2006)	Web 3D; e-commerce, other varied applications	X3D and its extensions for modeling of 3D scenes, as reviewed by (Flotyński & Walczak, 2017)
	(Beßler et al., 2020)	robotics	affordance models for Q&A systems
	(Blumenthal & Bruyninckx, 2014)		robot world models
	(Y. Zhao et al., 2024)	3D environments and robotics	contextual, topological, and geometric information of 3D scenes
	(Wang et al., 2019; Y. Liu et al., 2019; T. Chen & Guy, 2018; Fremont et al., 2019)	games and learning systems	Procedural content generation
Framework	(Speicher et al., 2021)	virtual and augmented reality	wizard of oz system with medium fidelity paper prototyping
	(Gottschalk et al., 2020)	augmented reality; e-commerce	
	(Górski, 2017)	industrial VR applications	knowledge-based approach
	(W. Li et al., 2020)	rehabilitation for wheelchair users in VR	optimization approach to generating constrained 3D scenes
	(Borras et al., 2023)	cloth manipulation dataset creation in VR	cloth state semantically labeled
	(Yang, 2023)	education in VR	probabilistic signature scheme
	(Raffailac & Huot, 2019)	2D interfaces; based on entity-component system	

transition of scenes, assets, tasks, and interpretations between different steps of a complex workflow for VR experience creation (Vergara et al., 2017) and open possibilities for in-depth interaction analysis. This is the subject explored in the contribution presented during the Section 3.1.

2.1.2 Building studies enabling embodied interactions and analysis

For the research using VR media, understanding embodied experience is an important topic. The personalized training possibilities offered by embodied VR experiences have potential for application in contexts that may be considered highly dangerous or require expertise. This has attracted research into understanding users' behavior in specific contexts, using methodologies for analyzing perception, emotions, and behavior from neurocognition. In the literature, we see many applications of VR for training scenarios, such as understanding driving habits (Y. Lang et al., 2018), sports learning (Wu et al., 2021), or firefighter training (Clifford et al., 2019).

However, due to embodied interaction's specificities (i.e., physical interaction and multisensory engagement), a whole new set of interactive metaphors is required to observe behavior in contrast to traditional 2D media. Alcañiz et al. (2009) applied transcranial Doppler sonography to investigate how VR experiences stimulate the brain and found them comparable to real-life experiences. Dickinson et al. (2021) investigated the usage of diegetic interfaces (interfaces integrated and adapted to the virtual environment) in VR. While these types of interfaces show good efficiency in traditional 2D media, in VR, such interfaces do not bring immediate benefits and did even increase the workload and completion time of users. Many works (Peck et al., 2018; Ricca et al., 2021; Dewez et al., 2021; Aseeri & Interrante, 2021) show how embodiment, through body or hands visualization, is an important aspect of VR interaction, affecting how users will interact with the environment and interact with other users.

Performing studies using embodied interactions such as physical walk, more than just to allow to create life-like experiences, also shows improvement on multiple factors like the feeling of presence or the motion sickness generated compared (Caserman et al., 2021; Christensen et al., 2018). Weser et al., in their works (2024) also found that walking in VR is superior to all other locomotion techniques. However, using physical walk as the locomotion technique for a study greatly complexifies the design : due to the technical difficulties of setting it up (e.g., room size, cable length, headset battery). Few works involve long-range (greater than in a 4m x 3m space) navigation tasks with real walking. Boldt et al.(2018) mention room-scale navigation using the HTC Vive Pro's 12m² space limitation, but the scale of the study design isn't detailed. The designed task involved walking to a target and interacting with it, with the goal of investigating the effect of haptic and auditory feedback on wall perception. The backpack is often a solution selected for long-range navigation, such as Mousas et al.(2020) conducting a path following task of 150 meters with a VR backpack and analyzing physiological and pose data to investigate how mismatching virtual and real environments can strongly affect user behavior. This paper brings to light the fact that both virtual and real environments matter in VR experiences, adding a layer of complexity to the design of a proper VR experience in which real walking is included. VRoamer (Cheng et al., 2019) using a VR backpack allowed navigation within a 40m x 40m space, though the work focuses on enabling free navigation and is not task-oriented. Delaux et al. (2021) in their studies on perception during navigation in VR with EEG also used a VR backpack to allow users to freely walk physically in the study environment. While the backpack is a solution enabling free walk in the environment without a cable length limitation, some existing works also underline flaws in the usage of backpacks, such as the cost, the weight, and it's effect on user motion (Hughes & Kanakri, 2020; Wagner & Schmalstieg, 2006).

All these works highlight the complexity that surrounds the creation of an embodied VR experience. The creation of new interactive metaphors and systems allowing the creation of better embodied VR experiences is still a subject in development. We believe a workflow allowing designers to observe tasks and analyze embodied interaction on a granular level, by providing a better understanding of the lived experience, could not only support research on human behavior analysis applied to VR but could also support the improvement of embodied locomotion and interaction techniques. This is the subject explored in the contribution presented during the Section 3.2.

2.2 Understanding user experiences in virtual reality

In studies using embodied interactions such as physical walk and 3D object manipulation, observing and analyzing the UX becomes a challenge. As the body is used as a communication device, it becomes valuable to observe the body, behavior, and interaction to the user's UX, intentions and decisions. In this section, we explore the complexity represented by the comprehension of the embodied experience in VR through three subjects : (1) the creation of datasets enabling embodied interaction analysis in context ; (2) the analysis of user experience through behavioral metrics in context ; and (3) existing models in HCI research on task-based analysis to support granular behavior analysis.

2.2.1 Building datasets of behavioral and contextual data

To enable in-depth investigations of the user cognition and behavior in VR, datasets with rich behavioral and contextual data have been constructed. Many datasets in 3DoF (3 degrees of freedom, allowing users to rotate their head and body) were created in recent years. In 2017, Li et al. (2017) introduced the first reference 3DoF database obtained from 95 users freely watching 73 videos in 360° in VR, providing the users valence and arousal ratings using the self-assessment manikin (SAM) tool (Bradley & Lang, 1994) as well as the head positions' data. The material publicly available, however, only consists of the videos and the average valence and arousal value pair for each video (averaged over all users). Subsequently, Tang et al. (2020) presented an experiment where 19 users watched 36 images in 360 degrees while their self-reported emotions and eye motion were collected. The data collected by Toet et al. (2020) and Xue et al. (2021) also comprise physiological measurements of heart rate and electrodermal activity (EDA, as skin conductance), which have been shown to reliably represent user instantaneous arousal (Boucsein, 2012). They provide a dataset of 11 immersive videos from the same database experienced by 32 users, with subjective emotional ratings, physiological measurements, and head and gaze movements.

To observe embodied interactions such as the physical walk, a study 6DoF (6 Degrees of Freedom, allowing users to move in the environment) is required. A large number of 6DoF datasets were built using human motion data and behavior capture in context. A major one is the CIRCLE dataset (Araújo et al., 2023) published in 2023, which involved around 10 hours of human motion and egocentric data captured in VR scenes using a 12 Vicon camera setup. The actions concern reaching tasks in scenes annotated with the initial state and goals. For full-body motion capture, the Human3.6m (Ionescu et al., 2014) dataset is the most widely used, and up until recently the largest, featuring 3.6 million poses with RGB camera filming as well as actor lidar scans. The focus of Human3.6m is to provide a wide variety of human motions by professional actors, however, object interactions are not tracked. The GIMO (Zheng et al., 2022) and EgoBody (S. Zhang et al., 2022) datasets on the other hand combine motion capture and augmented reality headsets – the HoloLens – to capture gaze and motion in context. The physical environments are scanned as 3D meshes and calibrated to have the motion and scene in the same coordinate system.

The presented datasets are valuable contributions to work on immersive experiences, however they are mostly created for specific studies with capture setups that are not made available for reproducibility or extension. We explore the creation of datasets for embodied interactions observation and analysis in the contribution presented during the Section 4.2.

2.2.2 Behavior analysis of user experience in context

While a number of works in the domain of human-computer interaction and cognitive sciences have studied behavior in immersive environments (Baños et al., 2008; Felnhofer et al., 2015; Pallavicini et al., 2019; Voigt-Antons et al., 2020), it has been only more recently that the multimedia community in particular has considered sensing and recording behavioral metrics such as emotions, head and, gaze or body motion (B. J. Li et al., 2017; Tang et al., 2020; Toet et al., 2020; Xue, El Ali, et al., 2021; Xue, Ali, et al., 2021).

Physiological analysis of VR experiences in 3D environments is nowadays commonly used to refine experiences or understand the user behavior. Keighrey et al. (2021) investigate the possibility of evaluating

the user's perceived quality of experience (QoE) for virtual speech and language assessment applications on different interfaces (i.e., tablet, VR, AR) using heart rate and electrodermal activity measurements in addition to post-test questionnaires. Bender et al. (2021) quantified 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 has been used in clinical settings to analyze the physiological and mental state of the user for medical purposes such as anxiety prediction (Mevlevioğlu et al., 2021; Hawes & Arya, 2021) as well as stress and cognitive load assessment (Vesga et al., 2021). Other applications use physiological data for driver stress detection (Memar & Mocaribolhassan, 2021) and emotion evaluation with smart clothing (Pidgeon et al., 2022).

To understand the embodied experiences lived by users in VR, measures of motion, perception, and emotion can support the validation of observations on user behavior, providing insight into the influence of the content on the affective user experience, as surveyed by Luong et al. (2021). As early as 1999, motion, specifically posture, has been a subject of interest (Kuno et al., 1999) to measure the level of body sway as a function of visual motion in VR. Jicol et al. (2021) adopted questionnaires for emotional intensity and positivity and employed structural equation models to find complex correlations between emotion, sense of presence, and agency, such as the observation that sense of agency augmented sense of presence only while mediated by an emotional fear condition. Different types and levels of emotions can be correlated to specific types of motion (B. J. Li et al., 2017; Tang et al., 2020; Xue, Ali, et al., 2021), and results from Li et al. (2017) show more precisely some level of correlation between emotion arousal, valence, and head angle. On the other hand, results from Tang et al. (2020) show a significant link between valence and gaze behavior. Seinfeld et al. (2020) measured performance and the sense of embodiment as a function of object placement and modalities of interaction (e.g., hand gestures, joysticks, keyboards), finding a positive increase in performance and sense of embodiment for virtual hand paradigms. These studies brought about new ways to analyze the user behavior within a scenario and measure the behavior variation between scenario conditions using modern cognitive science approaches.

A major feature of embodied VR training frameworks is their capacity to adapt the scenarios to each person's needs, by analyzing and understanding the relationship between the user behavior and the environment. Rule-based performance metrics and baselines such as characteristics of the training scenario (W. Li et al., 2020), completion time (Rossol et al., 2011; Siu et al., 2016), and error rate (Rossol et al., 2011) have been dominantly used as parameters for such adaptations. The more recent inclusion of user behavior metrics as parameters for synthesizing personalized training scenarios or adapting existing ones is gaining traction. Chen et al. (2022) designed a framework that analyzes locomotion parameters such as pose accuracy and motion speed to generate new target points that train specific stances. Dey et al. (2019) explored the use of EEG to estimate task load from alpha peaks and adapt the difficulty level of a target selection task. Lang et al. (2018) used gaze tracking and noted event baselines such as improper habits in driving scenarios, then used an optimization approach to generate personalized map layouts for improving driving habits.

While these works present valuable analysis of the user experience, we believe the understanding of embodied experiences could be deepened by analyzing behavioral and contextual data on a more granular level by coupling such works with task-based analysis. We explore the task-based analysis of embodied interactions in the contribution presented during the Section 4.3.

2.2.3 Task-based analysis outside of VR

Frameworks and methodologies enabling task-based observation of an experience are commonly used in HCI and Robotic research domains to obtain a fine-grained understanding of the UX or performed actions by the robot. In HCI, task analysis is a cornerstone process for finely understand how users perform tasks to achieve their intended goals (Diaper & Stanton, 2003). Task analysis is often done through direct observation of users interacting with the system. Task models provide an abstract structure for the analyst to organize information gathered during task analysis that can be further detailed where needed (P. A. Palanque et al., 2011). Santoro (2005) characterized task models representations of the human activity in a hierarchical structured way (such as a state machine, action tree, or graph) allowing the analysis of task feasibility (formal demonstration that a task can be achieved) and (estimated) user performance. Task models have strong

advantages to support task analysis, including coping with complex scenarios (by structuring tasks and sub-tasks), supporting abstract and generalized reasoning about tasks (beyond individual's cases of use), and serving as non-ambiguous documentation of tasks and observations made (Vigo et al., 2017; Diaper & Stanton, 2003).

Task models are essential in HCI for a fine-grained analysis of user tasks. They bind together a deep understanding and characterization of the interactive system constraints, the user performance, and the qualitative aspects of the experience. Various works have adopted task models for evaluation on various levels of granularity. KLM-GOMS (Sauro, 2009) is a well known method used for traditional media to quantify how much time users take to perform a given task by aggregating the duration of low-level tasks (e.g., point mouse to a target = 1.1 seconds, move hand to keyboard from mouse = 0.4 seconds) that compose more complex interactions.

In robotics, there is a strong interest in performance analysis on fine-grained slicing of a scenario, such as to pinpoint performance flaws. Analyzing and identifying when there is a flaw and the cause are crucial to improving the robot and avoiding future failures, corresponding to a similar need in VR applications where adapting and personalizing UX are concerned. For example, Lee and Lozano-Pérez (2015) focus on the design of interaction graphs that finely segment tasks carried out by a robot, in order to precisely detect when an error occurs and proceed to a failure mode. Kroemer et al. (2014) focused on the prediction of phases of manipulation by a robot using a probabilistic model to represent the steps composing a manipulation task.

Moving to works using task based methods on recent medias, we find Rice et al. (2014) who investigate the description of different inputs composing touchscreen and mobile device interactions (e.g., pinch, zoom, tilt). Guerra et al. (2022) extended the available interactions and measured interaction time to provide new GOMS methods to evaluate task performances (e.g., grab an object, using teleportation movement) for augmented and extended reality systems. Zhou et al. (2022) built a method focused on quantifying user performances using a virtual hand. These works aim primarily at the description of a scenario purely based on the time spent on each task, to measure the efficiency of performance, without considering behavioral performance components such as the user's attention or emotion.

While these works show effective ways to observe and analyze finely an experience based on the tasks performed in the context, few address the embodied interactions encountered in immersive systems. We believe that task models have strong advantages in evaluating the usability and experiential UX of systems, and could be adapted to the embodied context to support more granular analysis and understanding of embodied interactions. This could open opportunity to finely observe the user's performance in correlation with their current mental state (e.g., stress, emotions) and physical state (e.g., attention, motion) in embodied context.

2.3 Improving designer-user communication mediated by systems

In embodied VR experiences, communication problems are likely to arise between the experience sought by the designer and the experience lived and interpreted by the user, due to the variety of scenarios and interaction modalities that make up a realistic embodied scenario. In this section, we explore the complexity represented by the designer-user communication mediated by systems through two subjects : (1) the designer's intention of communication through products and systems ; and (2) this communication mediated by virtual reality systems.

2.3.1 Designer's intention communication through a product

Understanding the gulf between the designer's expected usage of a product and its actual usage in real-life situations is key to the design of consumer products, human-computer interfaces, and even virtual reality experiences. Such gulf has multiple impacts on the product depending on the usage context : using the product the wrong way can reduce its usability, directly affecting its efficiency and user satisfaction (Ferreira et al., 2020). In various research domains, it can be important for a user's studies to be interpreted and lived the way the designer intended them to be, to align the designer's intention with the lived experience. A gap

unintended by the designer could influence the way the data can be interpreted. Diverse research domains have investigated ways to formalize and reduce this designer-user gulf by defining models, taxonomies of issues, or exploring means to more efficiently communicate an intention to a user through a product, a system, or an experience.

Communication through a product has been a key subject in the industry for a long time, to maximize a product's efficiency or attractiveness. Norman (2010) described a product as a service about the discovery, the anticipation, and the first usage. He presents a product as more than a product; it's an experience built by the designer and lived by the consumer. Crilly et al. (2008) in their work investigate the nature of the relation designer-user around a product, presenting the design of a product as a way for the designer to communicate an intention, which will then be interpreted by the consumer and used only based on their own comprehension. Norman (2008) highlight how good designers can efficiently communicate with their users, using visible clues, hints, and affordances, to maximize their chances for the user to understand and use the product as intended by the designer. These different works set the basis for the importance of intention communication when designing a product, nonetheless the improvement of this communication is still an open question.

Multiple research domains explore ways to reduce the designer-user communication gulfs, highlighting two ways to proceed : by evaluating the quality of the intention communication, or by categorizing and proposing guidelines to prevent the issues in the design or usage process that could be the source of communication gulfs. Souza and Leitão (2009) presented how semiotic engineering views human computer interaction (HCI) as a computer-mediated communication between designers and users, where the system speaks for its designer to communicate an intention by using diverse means of affordance. It presents methods such as the semiotic inspection method (SIM) to appreciate the quality of the meta communication and the communicability evaluation method (CEM) to analyze the quality of the communication done through the user interface. In his previous works, Norman (2013) did present a categorization of issues that can happen when a user interacts with a product as slips and mistakes : slips represent a conscious issue while performing the correct action, while mistakes represent an unconscious issue in understanding the correct goal. Laubheimer (2015) applied this categorization particularly to the human-computer interaction domain, presenting strategies and guidelines to prevent slips and mistakes during website or application design.

A common way to investigate communication issues is through the creation of a human errors' taxonomy while interacting with a system. Bolton (2017) establish a taxonomy to classify the different erroneous human behaviors. It describes where a user diverges from a normative task and how to identify the information the user failed to properly understand or attend to. Multiple taxonomies were built pursuing comparable purposes, such as Paek's taxonomy (2003) on various types of communication errors in linguistic or conversation analysis. This was done in order to highlight the different types of failure and present guidelines to handle such errors for automated systems. Gibson et al. (2006) presented a taxonomy on communication issues related to railway track maintenance, classifying the observed issues according to the communication topic and the type of error in question. Tian and Oviatt (2021) presented a taxonomy focused on human-robot social interaction error, investigating the impact of such issues and ways to reduce issues while interacting with a system, which can degrade the user's perception of the system's capability to achieve a task or to properly achieve a socio-affective interaction. Such a taxonomy can be very valuable, as preventing or avoiding any user erroneous action is a main concern for system designers.

To reduce the designer-user gulf, another solution is to explore efficient ways to communicate an intention to the user or reorient users if they deviate from the intended experience. Matthews et al. (2017) investigated how to communicate an intention to let a pedestrian cross the road when using autonomous cars without drivers. By showing a textual message to explicitly communicate the intent, users reacted more predictably compared to situations without explicit communication. This serves to enable both better communication and a better understanding of the user's behavior. Li and Zhang (2017) looked at ways in robot-human interactions to implicitly communicate a user intention to a robot by proactively analyzing user gaze. They found positive results on the usage of gaze data to improve user-system communication, opening possibilities for the system to better capture the user's intention. These works highlight interesting ways to communicate intentions and communicate with users. While these works weren't built for VR, we believe that comparable application could be proposed and applied for the VR context.

While these works show effective solutions for communication refinement in product design, HCI, or robotics, they are neither built nor applicable for the observation of embodied experiences, which display different interactive properties and usage cycles.

2.3.2 The designer-user communication in virtual reality

In virtual reality, communicating an intention can be complex, due to the offered freedom and interactivity. The work of Larusdottir et al. (2019) highlight the complexity that can represent the creation of a VR experience to be lived the intended way, listing numerous issues encountered during a user test session. Speicher et al. (2021) presented a prototyping application for mixed reality (XR) to simulate spatial interactions via a Wizard of Oz system in a minimalist version of the 3D environment captured by a camera. This prototyping methodology increased realism and a reduced gap between the prototype tested by the designer and the actual experience proposed by the system, to support designers on the refinement of their built experiences. Wolfartsberger and Niedermayr (2020) explored ways in VR to reduce the deviation of the player from the intended path in a learning tool. They presented a virtual “ghost user” performing the tasks to help the player learn. David-John et al. (2021b) used gaze data in order to investigate the prediction of the player’s interactions with the 3D environment, which could open possibilities for using passive, biological responses to reorient players at the right moment who are deviating from the intended experience. To reduce the designer-player gulf in VR, Oh and Kim tried to (2019) use onomatopoeia and sound effect feedback to emphasize or draw attention to a precise element. Through a study with and without onomatopoeia, onomatopoeia showed a higher success in communicating the designer’s intention and offer more object affordance.

Understanding and avoiding issues in the communication mediated by VR is a key to improving immersive experiences. Vuarnesson et al. (2021) looked into how to assist multiple users in communicating their intentions through a VR experience. In the study case of dance improvisation, they presented a minimalist avatar representation of users outside of VR to make them co-present and improve the user’s intention communication. Vignais et al. (2009) investigated the gulf in sport science that can exist between the experimental protocol and the subject behavior in real conditions. Virtual reality was used in a simulated handball game context to compare two conditions : the judgment task (where the user has to predict the direction to go to stop the ball) and the motor task (where the user has to physically stop the ball). VR, by allowing fine-tuning of the experiment’s usage in two different conditions, allowed the observation of the perception-action coupling. Cavalcanti et al. (2021) investigated ways to train participants for the evaluation of their choices using simulated dangerous environments in VR. In a gamified environment, users had to cross a maze while avoiding multiple dangers. The dangers varied from low to high, with multiple warning conditions. While designers expected the user to avoid the danger, the users, as they knew they were in a virtual environment, had the tendency to test the consequences of their actions rather than maintaining safety. This highlights a key gulf between a designer’s intention when building a study and the actual user experience.

The works presented in this section highlight the various ways in which gulfs in designer-user communication in VR occur and the importance of reducing this gulf. As a media increasingly used for scientific research, the rich interactive possibilities and 6 DoF exploration allowed can widen the gulf of comprehension between the actors. We believe a model specifying the communication issues that can happen during the parts of the design and usage process of embodied experiences could support designers. By knowing what issues can happen and how to prevent them, designers could improve their VR-mediated communication as well as their understanding of the user’s lived experience. This is the subject explored in the contribution presented during the Section 5.2.

2.4 Summary

In this chapter, we focused our attention on the existing work that creates and analyzes immersive experiences, and on works that explore ways to provide feedback to designers to support the refinement of the experience. VR studies are employed for a variety of purposes, from dangerous scenario simulation to human behavior analysis in specific contexts. A wide number of factors can complicate both the creation and the understanding of comparable studies (e.g., a setup allowing physical interactions, the observation of relevant behavioral metrics, the observation of designer-user communication issues).

On the topic of **embodied experiences creation and observation**, we identified three main challenges : (1) the creation of semantically understandable ontology for immersive interactive environments ; (2) the observation of multiple behavioral metrics related to embodied interactions ; and (3) the creation of 6DoF experiences enabling the user's embodied interaction, such as physical walk. Existing works did explore these challenges, however none present a reproducible fully-integrated workflow enabling the creation and task-based observation of rich embodied experiences in VR.

On the topic of **embodied experiences analysis**, we identified two main challenges : (1) the availability of large immersive datasets composed of multimodal metrics to enable embodied interactions analysis ; and (2) the performance of behavior analysis in correlation with the contextual data of a multitask scenario on a granular task level to allow embodied interactions comprehension. While many works explore ways to analyze the user experience in VR, the analysis is often performed through the usage of a single behavioral metric, in 3DoF context, or in a scenario composed of a single task. While these choices allow for simplified analysis, they don't support the ability to finely analyze embodied interactions in multitask scenario contexts, composing a large part of the immersive training scenarios.

On the topic of VR mediated **designer-system-user communication**, we identified two main challenges : (1) the lack of methodology to support the identification of communication issues happening during the design and usage phases of embodied experiences ; and (2) the creation of relevant feedback for designers to contribute to the refinement of the experience. While a wide variety of works and models were built for this subject in HCI, semiotic engineering, or robotics work, the existing communication models weren't designed nor applied yet to VR studies.

In this thesis, we aim at addressing these challenges to support designer on embodied experience creation, analysis and understanding.

Contributions

CHAPTER 3

Engineering ontology systems

The first subject investigated in this thesis concerns ways to conceive virtual reality studies, enabling the execution and observation of embodied interactions. A major challenge in designing 3D experiences and user tasks lies in bridging the inter-relational gaps of perception between the comprehension by the designer and the user and the actual 3D scene. Based on Paul Dourish's theory of embodiment, issues with the **ontology** components result in gaps between the scene representation and the user and designer's perception. To bridge this gap between users and systems, we focused our work on a workflow to assist designers in the creation of semantically rich experiences as well as on the collection of relevant metrics regarding embodied experiences.

This chapter investigates the *RQ1 : How can we support designers in the creation and observation of embodied interactions in rich, annotated 3D contexts ?* through two sub-questions :

- **RQ1A** : How to create an ontology able to enrich the design and the usage of VR experiences to improve the user-system and designer-system communications ?
- **RQ1B** : How to build VR experiences efficiently that can enable the capture of metrics in the context of the scene relevant to embodied interaction analysis ?

We present two contributions. We first address the challenge of building scenarios with rich semantics, addressing RQ1A, to improve the creation and management of experience through the conception of the framework GUsT-3D presented in the section 3.1. The second contribution addresses the challenge of conceiving embodied experiences and capturing metrics related to embodied interactions in response to RQ1B. We applied in real conditions the framework GUsT-3D and a technical platform for an embodied VR study, presented in the section 3.2.

3.1 GUsT-3D : Framework for designing guided user tasks in embodied VR experiences

The first major challenge investigated during this chapter, related to the **RQ1A**, concerns the creation of systems to assist designers in building semantically understandable environments, enabling the gain of insight into user behavior, perception, and efficacy in a task in VR. These insights are keys to user experience understanding and the measuring of the usability of interactive systems. This has been addressed in research in user-centric design and task modeling (Bernhaupt et al., 2020). However, in the design of immersive interactive systems, gaps of perception may appear in any pairwise relation between the designer, the user, and the system itself. In VR, the freedom of movement to visually explore and interact with an immersive 360-degree environment further widens these inter-relational gaps between the designer's desired user experience, the affordances of the 3D scene in VR, and how the user intends to act and react throughout the experience.

To create and observe a VR embodied experience, we designed the GUsT-3D framework which defines and use a consistent set of vocabulary to enables a workflow consisting of 4 steps : (1) the annotation of

various entities in the scene, including navigation spaces, interaction possibilities, object properties, and relations; (2) the definition of *GUTasks* and their constraints for completion; (3) running the *GUTask* scenario with real-time interaction, logging, and validation of the task; and (4) analysis of user experience through the construction of spatio-temporal queries and scene graph visualization. This workflow is implemented as a set of tools in Unity to create and design embodied, interactive experiences in VR. The export of the scenario in GUsT-3D representation contains sufficient information to recreate the entire 3D scene with all the original geometric, interactive, and navigational properties. Conversely, the *GUTasks* can be exported and imported into a new scene, holding the same semantic properties required by the task as the original scene. We validate these contributions with two case studies and a formative evaluation through expert interviews with professional Unity users. The rest of Section 3.1 describes our framework, and its implementation is discussed in Sections 3.1.1.

The framework GUsT-3D for creating 3D embodied experiences, shown in Figure 3.1 is comprised of three main components, each used as a technical basis for chapters of this thesis :

- **GUsT-3D** : representing the **ontology** of the scene, implemented in JSON. It is used to represent the export of the annotated 3D scene, including all of the geometric properties, interactive and navigation possibilities, and their interrelations as a spatio-temporal scene graph. It also serves as the vocabulary set used consistently with the *GUTask* definition and user logging.
- **GUTasks** : representing the **intersubjectivity** between the designer and the user, also implemented in JSON using the same vocabulary defined in GUsT-3D. Allows the definition of user tasks to be carried out in the VR scenario, their interactive and completion constraints, and management and validation in real-time.
- **Logs and query language** : representing the user **intentionality**, implemented in LINQ (Meijer et al., 2006). By capturing multiple contextual metrics, it allows the designer to construct spatio-temporal queries on any annotated object or user in the 3D scene.

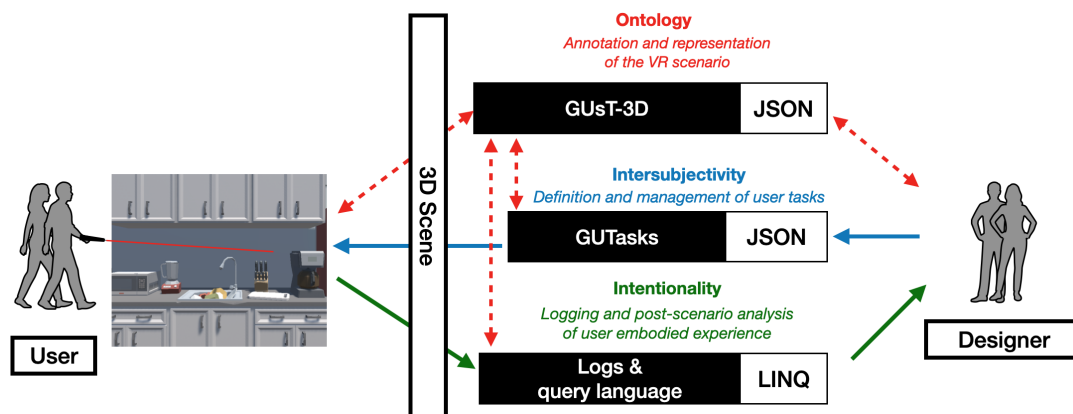


Figure 3.1 – **3-Component framework** for designing *GUTasks* in 3D virtual environments based on embodiment theory : the GUsT-3D vocabulary which is the full representation of the scene *ontology*, the definition and management of *GUTasks* communicated between the Designer and User for *intersubjectivity*, and the logging and query of user experience to allow Designers to analyze user *intentionality*.

The implementation of this framework as a unified toolkit in Unity allows the realization of a typical workflow for the design of embodied experiences involving four steps : (1) the design of the 3D scene and annotation of various properties (e.g., interaction, navigation) and calculation of the scene graph (relations between objects), (2) the definition of *GUTasks* that the user should carry out, their constraints and interaction design, (3) the running of the task scenario with real-time guidance, task validation, and logging, and (4) post-scenario analysis of the scene evolution and user behavior through query and visualization tools. The detailed task model for this workflow is depicted in Figure 3.2.

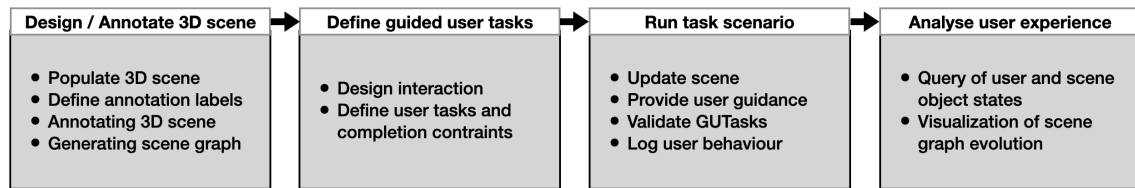


Figure 3.2 – **4-Step workflow** to create *Guided User Tasks* in 3D environments involves : design and annotation of the 3D scene, defining *GUTasks*, running the task scenario, and post-scenario analysis of the user experience. Each step of the workflow contains a number of sub-tasks that the designer may need to do.

The following section will introduce the tools of this framework and how it integrates into a design workflow for embodied experiences.

3.1.1 GUTask workflow and system implementation

This section details the implementation of our framework for the creation of 3D embodied experiences, shown in Figure 3.1. We first introduce each of the three elements of the framework corresponding to Dourish’s theory of embodiment – the GUsT-3D ontology (Section 3.1.1.1) for annotation and scene graph visualization of entity relations, definition and management of intersubjectivity (Section 3.1.1.5) through *GUTasks*, and the analysis of user behavior for intentionality (Section 3.1.1.6) through query and scene graph analysis. We then summarize how this framework fits into the task model depicted in Figure 3.2.

3.1.1.1 Ontology : the GUsT-3D representation

We refer to the elements in the scene as entities, which include all objects, terrain, and the user. The first step is for the designer to establish the ontology of the 3D scene. This includes (1) defining and assigning various properties (e.g., interactive, navigational) to entities in the scene, (2) defining the constraints for interactions between entities, and (3) generating a global view of entity relations within the scene.

3.1.1.2 Layers for defining and annotating entities :

We introduce the idea of *layers*, which represent a category of an object in the scene and/or its interactive properties. Currently four categories of *layers* are provided in our framework : navigation, object, interactive, and environment layers.

Navigation layers : structure the space in which the user can move, and defines all the navigation possibilities and constraints within the 3D scene. Three types of navigation layers are currently defined : *ground* for the *navigable spaces*, such as a room or the strip of sidewalk of a street ; *entryways* such as doors and passages that serve as connections between these navigable spaces ; and *obstacles* such as walls that partition these spaces. The designer can define new layers that inherit the *ground* property to create a named navigable space.

Interactive property layers : describe the relationship between two or more entities in the scene, and indicate how these entities can interact within the 3D scene. These properties exist between two scene objects : *support*, indicating that an object can support another ; *placed-on* the opposite, signifying whether an object can be placed on another *support* object ; *container* indicating if an object can contain another object. These properties also exist between an object and user : *movable*, indicating whether an object can be taken and moved around by the user ; *interactable*, indicating whether the object can be activated, such as a light switch, button, or a remote control.

Object layers : layers that have structural or interactive functionalities within the scenes – furniture that can serve as supports or containers, props that can be moved and manipulated. They can inherit other layers such as to take on interactive properties.

Environment layers : are used to annotated entities within the 3D scene that do not change the topology nor structure of the 3D scene, but through environmental parameters, allow users to change the perceptions of the environment. For example, all scenes have a camera, which can be extended to be a perceptive agent, such as the user. The camera is often used to represent the user’s visual range (i.e. field of view), current position, movement direction, and gaze. Another layer is light, which contributes to environmental lighting of the scene.

The flexibility of our layer approach is that objects can be assigned multiple layers, which can also inherit other layers, and thus gaining their annotated properties. For example a shelf layer can be defined as inheriting both support (e.g. placing a book on it) and container since it can hold other smaller items on its shelves. The ground layer also inherits support, since it has to support furniture and other items.

3.1.1.3 Interactive constraints

While the definition of an interactive property layer only indicates an interaction possibility, certain constraints to the interaction may exist. For example, one may wish to enforce that during the interaction with certain objects, the target must be in the visual field and within reaching distance of the user.

Within GUsT-3D, the designer can redefine the constraints by directly using the pre-defined layers. In GUsT-3D a constraint is a tuple of three entries :

- the *(Source,Target)* pair of the interaction, with layers.
- *Criterion* on under which the interaction can take place, which is a boolean *evaluation* on a *property* for a certain *value*. The property can be further of three types :
 1. Individual property : **a)** a geometric property (i.e., *position, rotation*), **b)** the localization of the entity (within a navigable space), and **c)** the layer which the entity belongs to,
 2. Relational property : **a)** the Euclidean *distance* or designer added definitions such as *near* or *far* based on the Euclidean distance, **b)** the relative position between entities including *above, below, contains*, and *holding* (e.g., in the case the user is holding an object), and **c)** visibility constraints between the camera and another entity such as *visible, occluded*, and *facing*.
- *State_change* resulting from the interaction

Using this vocabulary, we can then define an interaction constraint, such as the following for *take_object* movable and interactable :

```

1 "unitSpace": { "name": "Meter", "value": 1.0 },
2 ...
3 "distance_variables": [
4   { "name": "Far", "value" >= 8 },
5   { "name": "Near", "value" <= 3 } ],
6 "interaction_criteria": [
7   { "name": "Movable",
8     "range": { "name": "Near" },
9     "visibility": true },
10  { "name": "Interactable",
11    "range": { "name": "Near" },
12    "visibility": true } ]

```

In this example, to perform an interaction with an object of type interactable or movable, the criteria are defined to require the user to be near (i.e., within 3 meters) and have the object in their field of view objects. These are currently limited to relations between at most two entities, but we can imagine incorporating lambda functions to have more complex constraints.

3.1.1.4 3D Scene graph generation

One of the advantages of having entities in the 3D scene annotated with an ontology is to extract and represent the relations between entities as a scene graph. The scene graph becomes a model that can be used to describe the state of the system from a given perspective in a moment of time. To do this, we take the annotated scene, layer definitions, and the defined interactive constraints, and calculate the scene graph in three steps :

1. calculation of *navigable spaces*. Ground tiles labelled with the same location are grouped together to define a contiguous space.
2. calculation of the navigational graph based on the positioning of entryways between the navigable spaces, and
3. reasoning of *interactive constraints* between all pairs of entities to extract the interactive relations

These are then exported as a .dot file which can be visualized as a graph where the nodes denote various entities or entity groups, and edges denote their relations. An example of a scene graph calculated from the annotated scene can be found in Figure 3.3.

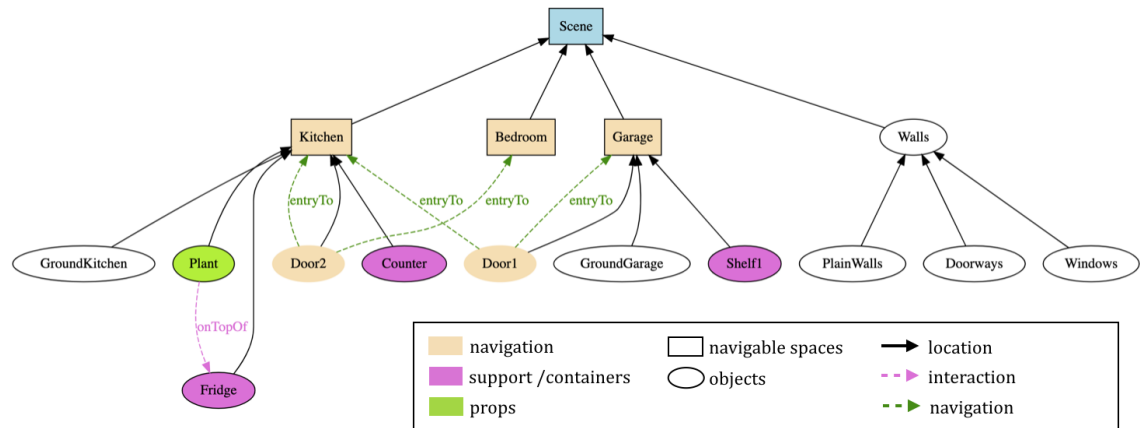


Figure 3.3 – A partial scene graph generated from the scene to show the relations between various entities in the scene, including navigable spaces and entryways, movable props, objects that serve as support or containers, and other elements such as ground and wall entities. The hard edges denote location relations, whereas dashed edges of different colours indicate interactive and navigation relations between entities.

3.1.1.5 Intersubjectivity : defining and managing real-time GUTasks

Setting out from an annotated 3D scene, the designer must now define the target tasks for the user to carry out, which we term *Guided User Tasks (GUTasks)* that can be managed in real-time.

GUTask definition Based on the grammar for interactive constraints and state changes, defining a *GUTask* comprises of specifying a list of constraints for the success of the task. Specifically, each constraint is composed of a time frame, and the desired state to be achieved. Assuming the task scenario is run from time t_0 to t_n , expressed in first order logic, a state can be evaluated at : a specific time $State(t_m, [parameters])$; at some arbitrary point during the task $\exists m(State(t_m, [parameters]))$; or at all time steps during the task $\forall m < n \rightarrow State(t_m, [parameters])$

Suppose we would like to define a *GUTask* that requires the user to place a book on the bedroom table, the definition of this *GUTask* would be :

```

1  "unitTime": {"name": "Second","value": 1.0},
2  ...
3  "sceneName": "Road-crossing-oneway",
4  "gutask": [
5      {"taskID": 1, "goal": "Take", "goalItem": "Book",
6       "targetHelp": "PathGuide", "baselineHelp": "HelpText",
7       "TargetTime": 10, "BaselineTime": 20,
8       "targetText": "", "baselineText": "The book is on the ground",
9       },
10     {"taskID": 2, "goal": "GoTo", "goalItem": "Bedroom",
11      "targetHelp": "PathGuide", "baselineHelp": "None",
12      "TargetTime": 10, "BaselineTime": 20,
13      "targetText": "", "baselineText": "",
14      },
15     {"taskID": 3, "goal": "PlaceOn", "goalItem": ["Book", "Bedroom table"],
16      "targetHelp": "PathGuide", "baselineHelp": "HelpText",
17      "TargetTime": 10, "BaselineTime": 20,
18      "targetText": "", "baselineText": "The table is next to the bed",
19      } ] }

```

When the target time or baseline time is exceeded for a task, a help can be provided to the user. In the above example, we can observe two different types of help : (1) *HelpText*, displaying a text defined by the designer to provide hints to users; and (2) *PathGuide*, showing the path that leads to the target object or location to the user in the form of a red line between the user and the target. Figure 3.7 shows the interface created for the design of *GUTasks*. The definition of a *GUTask* is independent of any specific scene geometry or configuration : the same *GUTask* could be applied to two different scenes containing objects and entities annotated with the same *layers*. However, if the scene doesn't contain objects or layers defined in the *GUTask*, then the *GUTask* is not relevant to the scene, and thus there is no way for the user to achieve the task.

Real-time *GUTask* management Our framework also provides easy ways to record user behavior, and manage and validate *GUTasks* when the scenario is run.

The first such mechanism is the recording of logs, at a granularity that can be set to n logs per second in a Unity configuration window. Two types of logs are included : *LogO* created for objects, and *LogU* for the user. Each newly launched instance of the scene creates a separate log file. Each log contains the following information : **Movement** of the position and rotation of the entity, **State Changes** of any defined interactive property (from Section 3.1.1.3). Since these can be numerous, the designer can define a “watchlist” of those that should be logged, and **Interaction** that results in a state change to a scene entity, as a target-source pair.

Here is an example of a *LogU* during a scenario :

```

1  {  "logID": "c2s1r14",
2     "timeUnit": { "name": "Decisecond", "value": 0.1},
3     "playerHeigh": 1.75,
4     "logs": [{
5         "unixTime": 1675159973585, "timeID": 1,
6         "position": {"x": -1.9,"y": 3.4,"z": -4},
7         "rotation": {"x": 0.02,"y": 0,"z": 0.},
8         "holdingItem": "",
9         "localisation": "House",
10        "inViewItems": ["brown_table"],
11        "lookedAtItemName": "brown_table",
12        "currentTask": "Step 1 : Go to Outside",
13        },
14        {"unixTime": 1675159973605, "timeID": 2,
15         "position": {"x": -1.9,"y": 3.4,"z": -4.4},
16         "rotation": {"x": 0.01,"y": 0,"z": 0.},

```

```

17     "holdingItem": "",
18     "localisation": "House",
19     "inViewItems": ["black_box", "brown_table"],
20     "lookedAtItemName": "brown_table",
21     "currentTask": "Step 1 : Go Outside",
22     ...
23 },

```

In this example, the logs of the player were recorded at a frequency of 10Hz (one frame every decisecond). The logs include the user position, height, interactions, attention, and current task. In this example, the player currently at the localization House have the task asking to go to the localization Outside.

```

1  { "logID": "c2s1r14",
2    "timeUnit": { "name": "Decisecond", "value": 0.1},
3    "logs": [{
4      "unixTime": 1678368892245, "timeID": 1,
5      "item": {
6        "objRef": 1,
7        "objName": "golden_key",
8        "position": {"x": -4, "y": 0.9, "z": -4},
9        "rotation": {"x": 0.1, "y": -0.7, "z": 0.6},
10       "layer": "Movable",
11       "pair": ["coffeeTable", "OnTopTarget"]
12       "localisation": "House"
13     "status": "dropped"
14     },
15     { "unixTime": 1678368892245, "timeID": 1,
16     "item": {
17       "objRef": 2,
18       "objName": "garbage_bag",
19       "position": {"x": -2.7, "y": 0, "z": -4},
20       "rotation": {"x": 0.0, "y": 0.0, "z": -0.1},
21       "layer": "Movable",
22       "interactions": ["user", "holding"]
23       "localisation": "House"
24     "status": "holding"
25     },
26     ...

```

In this example, the logs of the objects composing the scene "golden_key" and "garbage_bag" were recorded at 10Hz. The logs include the object's position, layer, localization, and current interactions with either the environment or the user. We can observe the fact that the golden_key has been dropped by the user on top of the object coffeeTable, while the user is currently holding the object garbage_bag. These logs contain sufficient information to re-calculate the scene graph for each time step, updating localization and interaction states. The scene graph would allow one to deduce as much fine-grained information as all the entities visible to the user, as well as high-level changes such as an object going from being placed_on another object to being held by the user, which is key to analyzing the user experience, presented in the next section.

3.1.1.6 Intentionality : reasoning on and visualizing user experience through spatio-temporal queries

The query component of the framework is designed to allow the construction of diverse queries about the scene using a fixed set of vocabulary. Queries can be made on any entity or layer name. The syntax for possible queries under our framework is shown in Figure 3.4

Queries can be constructed on entities in real-time or on the logs post-experience. In real time, entities that match the query are highlighted in the 3D scene. On the logs post-experience, the designer can use this same component to construct spatio-temporal queries, and to observe at each timestamp the temporal

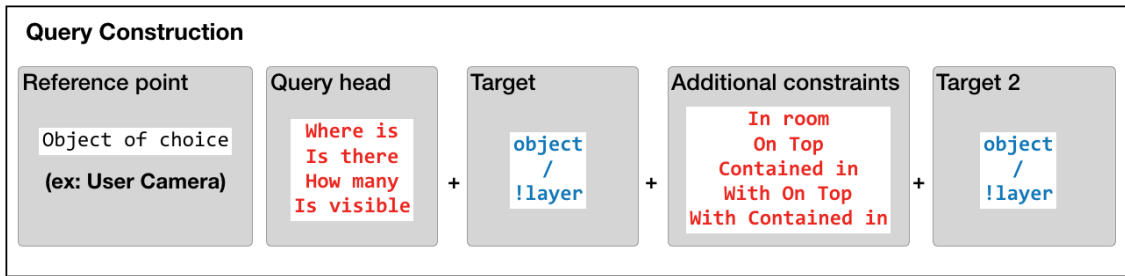


Figure 3.4 – The syntax of the query is composed of keywords from the vocabulary defined in GUsT-3D including a minimum of three keywords – the reference point (usually the user camera), query head, and target – and potentially two additional keywords that can further limit the search to entities that fulfil specific location or interaction constraints. The result of a runtime query can be seen in Figure 3.8.

evolution of positions of objects, the user, and the interactions that take place within the scene between these entities. This query component thus allows the designer or observer to monitor all of the entities of interest in the 3D scene. Object and path highlighting aids the designer in clearly defining the objects and layers related to the *GUTasks*, as well as add suitable indications which can be used to guide the user in completing the task.

3.1.1.7 Workflow and system integration in Unity

We can now combine the framework and the workflow from Figures 3.1 and 3.2 into Figure 3.5 to show how each element of the framework introduced in this section intervenes with the designer workflow, and the flows of information between the Designer, the User, and the 3D scene.

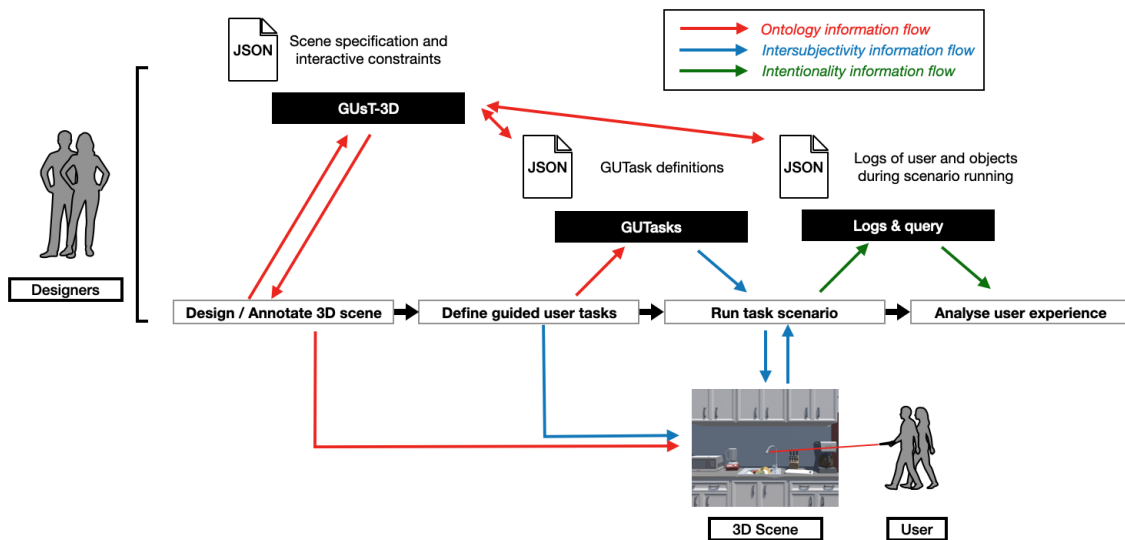


Figure 3.5 – The combined framework and task model to indicate how each element of the framework intervenes on the workflow of creating a 3D embodied experience. The arrows indicate the information flow on the three embodied experiences aspects : ontology (red), intersubjectivity (blue), and intentionality (green)

This framework is realized as a set of editor interfaces in Unity. It realizes our 4-step workflow of the design of an embodied user experience depicted in Figure 3.2. We detail below concretely how the framework integrates into Unity for each step of this workflow.

Design and Annotate of 3D scene ontology Scene annotation provides the first step to having a context-aware 3D environment that allows one to make meaningful inferences as to where the user is, what actions they are carrying out and how, and from there analyze the perceptions and intentions of the user.

We designed an interface that allows the easy addition of layers, interactive constraints, and units. The design schema for the interface is shown in Figure 3.6, which also demonstrates how layers are structured hierarchically to allow layers inheriting properties of other layers.

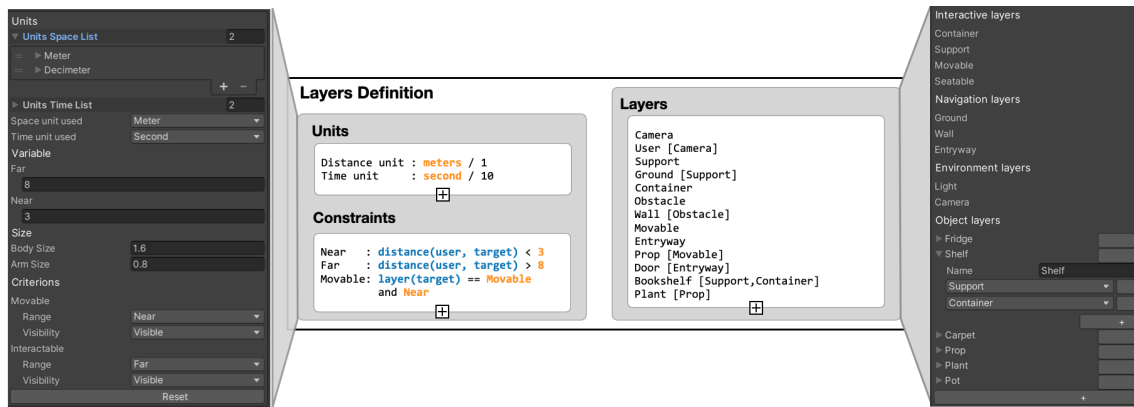


Figure 3.6 – Schema of the layers definition interface, which allows the scene designer to add and edit the units, constraints, and layers either through point-and-click or importing a configuration file. The actual interface we designed based on the schema is shown on the two sides.

When the scene has been fully annotated, the navigable spaces and scene graph are calculated with an interpreter for the layer relations and the interactive constraints. The navigable spaces are calculated based on connectivity between the ground tiles and the designated entryways. All entity properties and relations are then encoded by color and style respectively, and exported to a .dot file that provides a scene graph visualization as depicted in Figure 3.3.

Define *GUTasks* for intersubjectivity The next step of the workflow is to establish the intersubjectivity through defining *GUTasks* for the scenario to be communicated from the designer to the user, based on the pre-defined ontology. The interface component in Unity for specifying *GUTasks* is shown in Figure 3.7, allowing the easy modification of the task goal, constraints, as well as user guidance in the form of text or path indications that can be communicated to the user at specific time points when they have difficulty accomplishing the task.

The *GUTask* list is also saved in a JSON file in a separate file from the scene representation, using the same vocabulary defined in the GUsT-3D. They use the same layer and annotations from GUsT-3D ontology, which means the designer can easily create and import new *GUTask* lists for the same 3D scene, allowing the same scene to be reused easily for different task scenarios.

Analyze user experience The analysis of user experience is facilitated through the loading of the user logs, in which various information is then visualized. This visualization involves composing queries on specific entities in the scene. The Query Log window allows one to load a log, and drag the time slider to enumerate all the state changes for the specific entity. At the same time, the scene graph is generated for each time stamp, and can be viewed directly in the log window. These elements are shown in Figure 3.9.

3.1.2 Case studies

We present here two case studies – an indoor and outdoor scene – to show the wide range of possible 3D embodied experiences that can be created with our system.

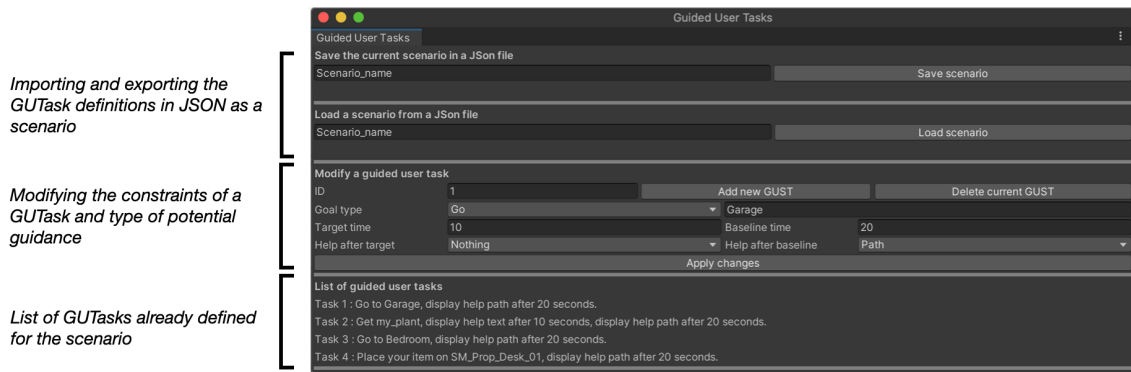


Figure 3.7 – This interface allows designers to define *GUTasks*, add and edit task constraints, view the list of existing tasks. The tasks can be easily loaded and exported in the GUS-T-3D representation for reuse in various scenarios.



Figure 3.8 – Here we show the window to construct a simple query using the syntax from Figure 3.4 that counts the number of props within a near distance (3 meters) of the player camera. The two views show the highlighting in the scene from the player viewpoint (left) and the Unity scene view from above (right).

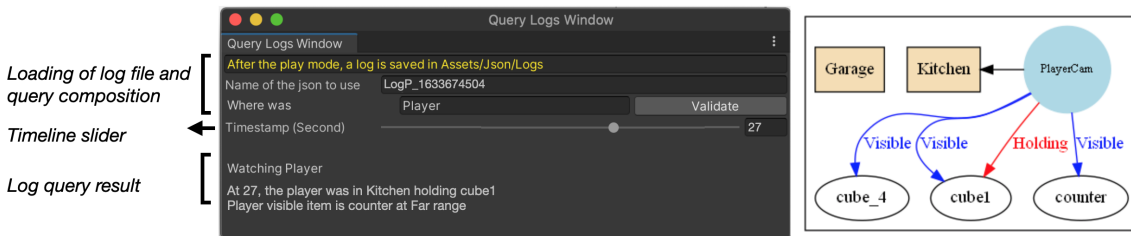


Figure 3.9 – Window for querying the logs of a previously run scenario, including loading a log file, a slider to show all the state changes recorded to an entity, and the descriptive text on the state at a time point. A generated scene graph of the user’s interaction state at a specific time point is shown on the right, indicating, for example, visibility of entities and objects being held.

3.1.2.1 Indoor scene : navigation and object interaction

The first case study uses the indoor scene seen on Figure 3.8 with three navigable spaces – Garage, Kitchen, and Bedroom – separated by walls, and connected by doors. We design a simple *GUTask* for this scene that is similar to an escape game task. The user must (1) locate and retrieve a stack of books from the Bedroom, and (2) navigate to the Garage and place them on the bookshelf to reveal a hidden passage.

The annotation of the scene ontology involves the navigable spaces (ground tiles for *Bedroom*, *Garage*, and *Kitchen*), and the assigning of object and interaction layers to other entities (*movable* such as book stacks or *support* like bookshelves). Multiple elements can be selected and annotated simultaneously with the desired layers and localization using the interfaces we designed in Section 3.1.1.7. A previously annotated scene that was exported in the GUST-3D representation can also be loaded using the interface, which facilitates the loading of ontologies already pre-defined by other designers.

Next, we define the *GUTask* that facilitates the intersubjectivity communication between the designer and user on what the user should do in the scene. The user starts out in the Bedroom where the stack of books is. Assuming the task is carried out between time t_0 and t_n , the constraints for correctly carrying out this *GUTask* described in first-order logic include :

1. User at some point must be in the bedroom, near enough to the stack of books, the stack of books must be visible from the user point of view, and the user must interact with the stack of books :

$$\exists m (Localization(t_m, User, Bedroom) \wedge Near(t_m, User, BookStack) \wedge Visible(t_m, User, BookStack) \wedge Interact(t_m, User, Take, BookStack))$$
2. At the end, the user must be in the Garage and the stack of books must be on the garage bookshelf :

$$Localization(t_n, User, Garage) \wedge onTopOf(t_n, BookStack, GarageBookshelf)$$

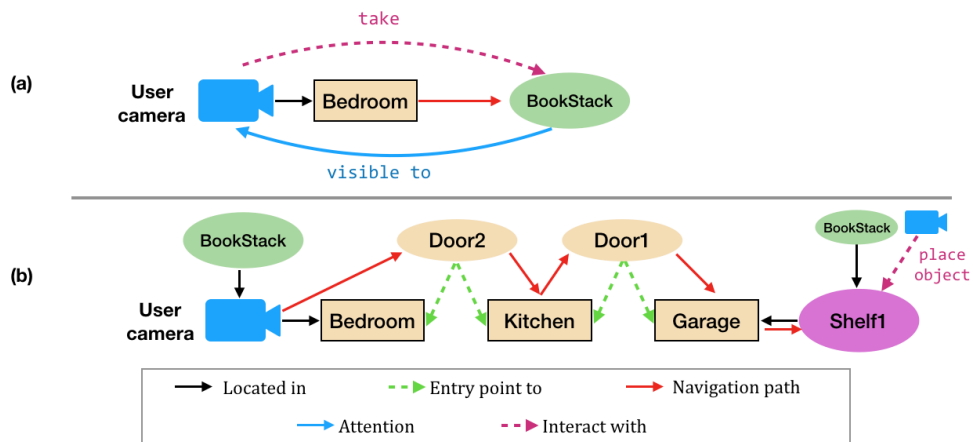


Figure 3.10 – Given the *GUTask* of placing a stack of books on the bookshelf in the garage, this figure depicts the interaction and navigation paths of the user for two constraints : (1) locating and retrieving the stack of books, and (2) navigating to the garage, and placing the stack of books on the bookshelf. We use the same color and shape representation of entities as Figure 3.3. It is worth noting that the interaction with the stack of books in (1) requires evaluating an attention constraint – that the object has to be in the user’s visual field.

The navigation paths and interactions for the two constraints of the *GUTask* are depicted in Figure 3.10 in two parts following the above constraints : (1) locating and retrieving the stack of books, and (2) navigating to the Garage to place the stack of books. Through user logs, we can observe which objects the user tries to interact with before finding the correct ones, and how much time the user takes to navigate the environment. It is an example of how a simple *GUTask* can be defined in a complex 3D scene with various interactive constraints.

3.1.2.2 Outdoor scene : traversing a traffic crossing

The second case study involves a pedestrian crossing as depicted in Figure 3.11 with a road lined by two sidewalk strips, and connected by a pedestrian crossing.

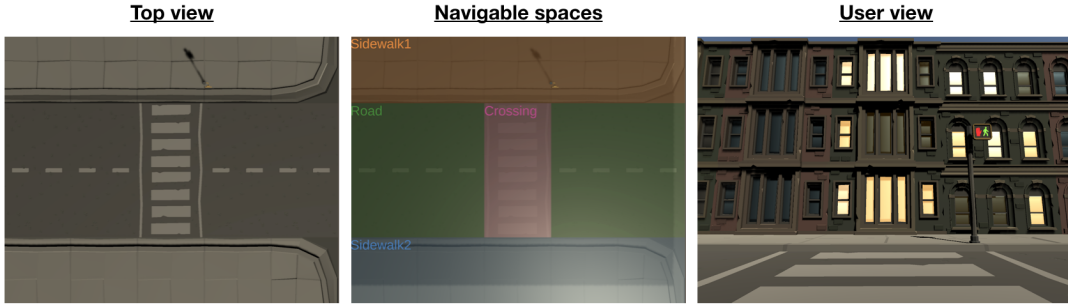


Figure 3.11 – The second case study features a basic road crossing scene in an open space. The scene has four navigable spaces : a road and a crossing that overlap, and two sidewalks. From left to right, this figure shows the top view, the navigable spaces, and a viewpoint from the user camera.

The *GUTask* defined for this scene is that the user must cross from Sidewalk2 to Sidewalk1 in security, involving two sub-constraints : (1) crossing through the pedestrian crossing and not anywhere else on the road, and (2) verifying that the crossing light is green before and during crossing.

The first step is the annotation of the scene ontology. The resulting calculation of navigable spaces is shown in Figure 3.11.

The second step in the workflow is to define the *GUTask* to communicate intersubjectivity from the designer to the user. The initial position of the camera is on Sidewalk2. Assuming the task is carried out between time t_0 and t_n , the constraints for correctly carrying out this *GUTask* described in first-order logic include :

1. User does not at any point appear on the Road :
 $\forall m < n \rightarrow \neg Localisation(t_m, User, Road)$
2. User must at some point look at the light when approaching the crossing from Sidewalk2 :
 $\exists m (Localisation(t_m, User, Sidewalk2) \wedge Near(User, Crossing) \wedge LookAt(User, Light))$
3. User can only be on the crossing when the light is green :
 $\forall m (Localisation(t_m, User, Crossing) \rightarrow State(t_m, light, green))$
4. The user should be on Sidewalk1 at the end :
 $Localisation(t_n, User, Sidewalk1)$

This *GUTask* further involves the element of user attention as an active constraint, requiring the user to consciously direct their attention towards the traffic light while approaching the crossing to ensure a safe crossing, which in addition to navigation, involves also an “attention path” of the user’s gaze.

Through generated user logs, further insights into user intention can be drawn, such as whether the user continued to pay attention to the light during the crossing, the variation of walking speed on different navigable spaces (e.g. the user walks faster on the crossing to ensure timely arrival on the other sidewalk).

3.1.3 Evaluation

We assessed our framework through a formative evaluation, which was initially used in education for instructional design (Boston, 2002; Black & Wiliam, 2009) and later extended and popularized in human-computer interactions by Hix and Hartson (1993) to describe evaluations in formative stages of design (Trösterer et al., 2012). As emphasized by Carol et al. (1992), formative evaluations provide observations and recommendations that can be immediately used to improve the design of the product or service, and refine the development specifications.

With this evaluation, our primary aim is to identify gaps between the goals of our framework and our current approach to the system implementation, which is done primarily through feedback from expert interviews, and also collecting specific points to improve on for the implementation before continuing to adapting the system to a professional use case. The case study of the indoor scene (navigation and object interaction) was used as illustrative support.

3.1.3.1 Participants

We recruited 6 expert users (3 male, 3 female) from a convenience sample, which is around the recommended number of participants for this type of formative evaluation (Jonathan Lazar, 2010). Concerning the experience with software available in the market, all six participants reported to have experience with Unity, with some having experience in other software including Blender (5), Unreal Engine (4), Maya (1), Voxel (1), iClone (1), Character Creator (1), 3D Exchange (1), and Tinkercad (1).

3.1.3.2 Procedure

The interviews were organized in four main steps : presentation of the study, profiling of participants, presentation of our workflow and implementation to develop the indoor scenario, and debriefing. We structured our interviews in a similar way proposed by Burmester et al. (2010), by presenting the usage of our application, allowing participants to freely express their positive and negative feedback, and collecting qualitative feedback on specific points to gain concrete advice for future versions.

The presentation of the steps to develop the indoor scenario was at the core of the interview. At first, we have provided an image of a scene composed of 3 rooms : a kitchen, a bedroom and a garage. Participants were then asked to assess and rate the perceived usability for performing the following steps in the workflow in link with embodiment theory :

- S1 - Scene annotation (Ontology) : involves creating new layers, and adding and annotating an object to an existing scene.
- S2 - Creating an interactive scenario (Intersubjectivity) : involves defining the *GUTasks* that the user should carry out in the scenario (i.e. go to the bedroom, then take the lamp placed on the desk, go to the kitchen, and then place the lamp on the table).
- S3 - Running the scenario (Intersubjectivity) : involves the user carrying out the defined *GUTasks* in real-time, with logs of the scene changes and user behavior recorded . This step is performed by the end-user and mostly automated for designers.
- S4 - Analysis of user experience (Intentionality) : involves querying and visualizing the log output of the scenario in order to understand the user experience.

We have created videos showing the real use of the tools with detailed information so that participants could identify usability problems in running scenarios. Participants were prompted to provide detailed comments with respect to : perceived usability, learning curve, efficacy, and flexibility of the system. For steps S1, S2 and S4, we asked participants to provide a score on a 1 (poor) to 5 (very good) Likert scale on their perceived usability, learning curve, efficacy, and flexibility of the workflow.

3.1.3.3 Results

In the analysis of results of the interviews, we look at the two parts : scores on various metrics for each step, and the verbal feedback on pros and cons of our workflow. With the selective study pool, our analysis focuses on the qualitative understanding of how our framework and workflow would improve their design process, and the limitations it imposes.

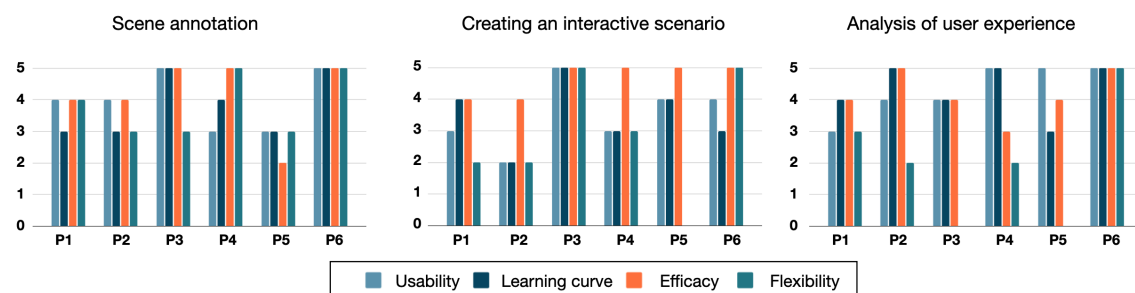


Figure 3.12 – For the three steps in the workflow, the responses of each participant on the usability, learning curve, efficacy, and flexibility of the implementation. The step “Running the Scenario” does not have these metrics, since it is mainly composed of the automatic processes that launch the *GUTasks* and record logs of the experience. The missing responses indicate where participants responded “I don’t know”.

An overview of these scored evaluations received from the expert interviews in steps S1, S2, and S4 is summarized in Figure 3.12, showing the responses of each participant on each metric for the three workflow steps. Overall, the participants found that the workflow implemented in Unity made their design process much more efficient, and on average the responses to the individual steps were all positive, though some more than others. Participants with more experience in 3D development were more critical of the flexibility of the interfaces.

We also analyzed detailed qualitative feedback from the interviews, as summarized in Table 3.1. Overall, participants provided more positive than negative comments for the general workflow. They also provided many suggestions for improving the usability and the user interface in all the workflow steps. On S1, participants (P1, P2, P4, P5) suggested that the interface should “*Centralize different features in a single window*”. On step S2, participants commented that “*The labels of the UI tools could be more easy to understand*”. P3 commented on S3 “*Display tasks with color depending on the guidance type, task goals, and times would help to better see the tasks*”. Another interesting comment on S3 from P4 was “*I don’t need to code just click and select, and progress is automatically saved*”, which is a convenient feature of the implementation. As for the step S4, the participant P5 put himself/herself in the shoes of the end-user and suggested that the designer should “*Ask the user how much guidance they want, and when*”, which definitely would help customize the guidance to be provided by *GUTasks* specified in the GUsT-3D representation, and thus improve the user experience.

The results of the formative evaluation have thus evaluated the usability, learn-ability, efficacy, and flexibility of the implementation of each step of the workflow. The results of such as formative evaluation confirm the needs of Designers for an integrated workflow for the creation and management of interactive user experiences in VR, demonstrates the added-value of the framework, and also provides valuable suggestions for improvements.

TABLE 3.1 – Participants were asked to elaborate the positive and negative points of each step of the workflow and its implementation. Here we summarize the qualitative feedback from expert participants on each step, with the IDs of the participants that provided each point in parenthesis.

Workflow step	Positive feedback	Negative feedback
S1 Scene annotation <i>(Ontology)</i>	<ul style="list-style-type: none"> - Time gain (P1, P2) - Convenient for non-developer (P2, 3) - Allows real-time testing (P2) - Easy to learn (P1, 3, 4, 6) - Any new object can be added (P2, 5) - Properties & pathfinding on objects (P4,5,6) - Ontology can be used with any 3D scene (P6) 	<ul style="list-style-type: none"> - Unclear path indications (P4, 5, 6) - Complex syntax for query language (P2) - Too many separate windows (P1, 2, 4, 5) - No sound annotation or properties (P6) - Can't be used with other software (P6)
S2 Creating an interactive scenario <i>(Intersubjectivity)</i>	<ul style="list-style-type: none"> - Easy to learn (P1, 2, 3, 5, 6) - Time gain (P2, 3, 4) - Convenient for non-developer (P2, 3) - Easy creation of task list (P4) - Good management of user guidance (P5) - Recap of all tasks is easy to understand (P1) - All information in one place (P1) 	<ul style="list-style-type: none"> - Limited to tasks available in the tools (P2, 5) - No parallel tasks (P5) - Names and labels in UI are not clear (P1, 4, 6)
S3 Run task scenario <i>(Intersubjectivity)</i>	<ul style="list-style-type: none"> - Easy to learn (P1, 2) - Provides a color-coded timer (P2) - Provides useful indications to the user (P3, 5) 	<ul style="list-style-type: none"> - Unclear object names (P1, 4) - Text on the screen is hard to read (P3, 6) - Don't know how to interact with objects (P4) - Cannot change UI display (P2) - Current time and time unit is not displayed (P1)
S4 Analysis of user experience <i>(Intentionality)</i>	<ul style="list-style-type: none"> - Scene graph is easy to understand (P4) - Displays information over time (P3, 5, 6) - Information is useful for analyses (P5, 6) - Shows user-scene interactions (P4) - Easy to learn (P2) - Scene graph generation is automatic (P1, 2) 	<ul style="list-style-type: none"> - Limited visualization methods (P3, 4, 6) - Lack of a global view (P4) - Hard to read when too many items (P2, 5) - UI lacks auto-complete/autoplay (P1, 4) - Cannot export scene graph as an image (P3) - Scene graph cannot be customized (P2, 5)

3.1.4 Discussion

In the literature on the subject of semantically enriching VR experiences, many frameworks and ontologies have excellent properties for designing 3D content. Nonetheless, they present a number of limitations when considering embodied experiences. (1) Reasoning about the user experience within the scene is mostly limited to the geometric properties of 3D objects (Brutzman & Daly, 2010; Pittarello & De Faveri, 2006) or interaction in a single usage scenario (Gottschalk et al., 2020). (2) They have a strong focus on the 3D geometry, and those that include interactions are defined at the level of the device and interface (e.g., a mouse, controller, gesture) (Górski, 2017; Raffailac & Huot, 2019; Behr et al., 2004; Navarre et al., 2005; Vanacken et al., 2007) instead of for user tasks (e.g., “take an object”, “move to location”), and hence do not allow systematic task validation. (3) There is no temporal representation of the world, and user state changes. (4) While some works allow the generation of 3D images or scenes (T. Chen & Guy, 2018; Y. Liu et al., 2019; Wang et al., 2019; Speicher et al., 2021; W. Li et al., 2020), these are not semantically annotated. Firstly, they do not have customizable interactive possibilities or user task management, and secondly, they do not allow a faithful re-creation of a specific scene with task and interactivity information included.

This novel framework based on Dourish’s embodiment theory presents multiple capabilities to design, run, and analyze embodied experiences in VR. This involves notably the development of the GUsT-3D representation to establish scene ontology, the *Guided User Tasks (GUTask)* definitions to create and run user tasks, and tools for the query, analysis, and visualization of scene and user behavior evolution. The framework is implemented and integrate well into the Designer’s process to create, manage, and analyze VR embodied experiences of target users.

We believe the scene graph generation using the ontology opens exciting pathways to support model-based evaluations of embodied VR scenarios, which is part of the work presented in this thesis. The next step of this work is the deployment of the framework to a practical use case, to observe how the framework perform in real conditions, and to collect metrics allowing the analysis of embodied interactions in immersive context.

3.2 Framework application for the observation and understanding of multimodal embodied VR experiences

The second major challenge explored, related to the **RQ1B**, is the task of building and executing VR studies enabling the performance and the capture of embodied interactions (e.g., 3D interactions, physical walk). Analyzing 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 (Grabarczyk & Pokropski, 2016); on the individual level, there are non-visible changes in our visual, emotional, motion processing, and related sensations that reflect one’s intentions (Martens et al., 2019). 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 personalized applications, and thus a grand challenge for VR media.

This section presents the practical application of the GUsT-3D framework presented in Section 3.1. This framework was applied to a case study investigating 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 visual scotoma, a region in the center of the visual field where visual information is blocked out); and (3) with physical walking or simulated walking using a headset game pad. 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.

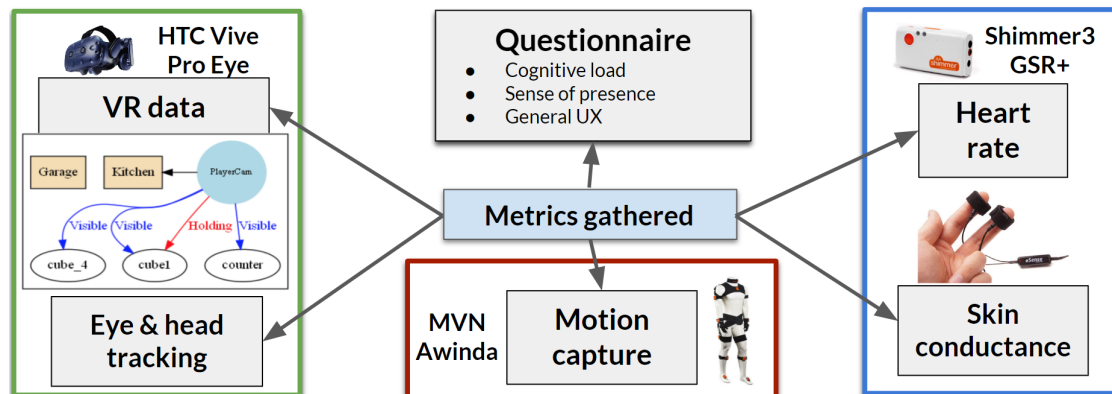


Figure 3.13 – The recorded data includes sensors in the headset and system – gaze and head tracking and system logs –, motion capture data from the XSens MVN Awinda starter, and skin conductance and heart rate from Shimmer3 GSR+ sensors.

We focus here not on the results of the study itself, but on the application of a technical platform using the framework GUsT-3D to enable the study to be realized, from the design of the interactive task model and the set-up and running of the study. This work presents three main contributions :

- realization of a technical platform with multimodal sensors and freedom of navigation
- the definition and encoding of a task model at levels of ontology, intersubjectivity, and intentionality, to enable fine-grained observation
- the execution of a study with 40 participants under this framework to collect embodied interaction related data needed for the next thesis chapter contribution

In this section 3.2, we describe in Section 3.2.1 the technical setup for the embodied experience enabling physical walk, as well as the multimodal data collected. Then, in section 3.2.2 we present the study design using task-based methodology. We explore the analysis of the embodied interactions through the metrics collected during this study in the next thesis chapter.

3.2.1 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 behavior. 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. Analyzing 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 the design of such a study and the rationale for our design choices.

3.2.1.1 Multimodal sensors

The metrics chosen shown in Figure 3.13 were selected in order to be able to rebuild the complete embodied experience by synchronizing 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 and physical state was the user, etc.).

The framework was made in order to incorporate a large variety of metrics, and be as flexible as possible in order to propose a framework adaptable to 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 minimization of latency, in order to facilitate the synchronization 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. (2021). We surveyed the HTC Vive Pro Eye ^{*}, the HP Omnicept Reverb 2 and the Varjo VR2 [†] 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 3.2.1.2. 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 synchronizing with the data captured with other equipment.

Motion capture Motion capture provides very rich information about embodied experiences, as shown in Mousas et al. (2020) 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 [‡] with 19 sensors and the MVN Awinda (jacket + 17 sensors) [§] 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.

Physiology sensors We decided to use sensors including skin conductance and heart rate, metrics popularly used by Memar et al. (2021), and Pidgeon et al. (2022) for the evaluation of the level of user emotion arousal. We surveyed the Empatica E4 [¶] and the GSR Shimmer3+ ^{||} 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 Consensys 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 behavior, 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.

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

†. <https://varjo.com/>

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

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

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

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

3.2.1.2 Navigating in a large space

In this work, we created a setup enabling physical walk locomotion in VR. 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 $10m \times 4m$, delimited in Figure 3.14.

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 module** 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 3.2, including the range of navigation we desired, as well as compatibility with the eye tracking data.

TABLE 3.2 – 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 $4m \times 10m$.
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

To our knowledge, we are the first, as of 2023, 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 (D. Huang et al., 2022; Kumar et al., 2022). 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.

3.2.1.3 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. We used the GUsT-3D framework (Section 3.1 for the task management and metric logging, allowing the conception of a vocabulary to describe 3D environments and

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

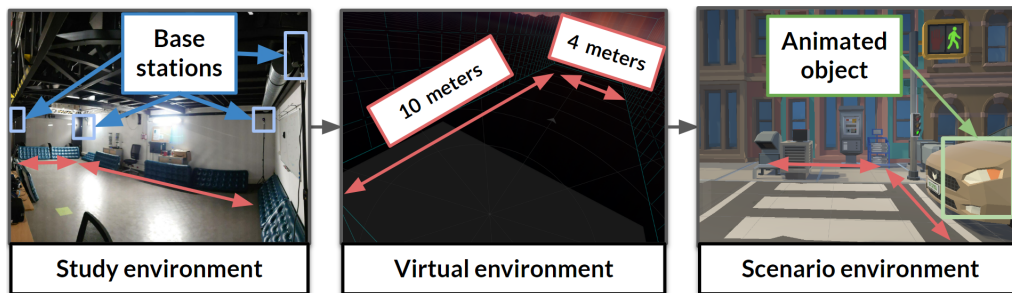


Figure 3.14 – The environment used for the experience measures 4 by 10 meters in navigation zone is delimited by four base stations, one at each corner, and aligned with the virtual environment. Mattresses surround the zone for security.

their interactive properties, as well as the definition, with this vocabulary, of tasks for the user to carry out. The system automatically records 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 visualizations to the designer. This system is effective to offer a control over the complete experience, from the environment design to the experience analysis.

3.2.2 Study

Our conceived framework presents the specificity to observe finely multimodal data, to enable the analysis of embodied interactions in VR. Using this framework, we designed a study to analyze the embodied user experience 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.

3.2.2.1 Task model

The study was built based on a task model shown in Figure 3.15, created to define all scenario tasks, and visualize all objects and interactions the system should be able to support in the study. To allow fine-grained analysis, knowing precisely what tasks are done and in what order allows us to synchronize 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 analyze user behavior 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 3.15 was therefore selected and transcribed as audio instructions played to the user. Tasks below this level are therefore considered as information, the “intentions” that we want to deduct and analyze from our multimodal data. Usage of the task model for in-depth analysis of embodied interactions is described in more detail in the next thesis chapter (Section 4.3).

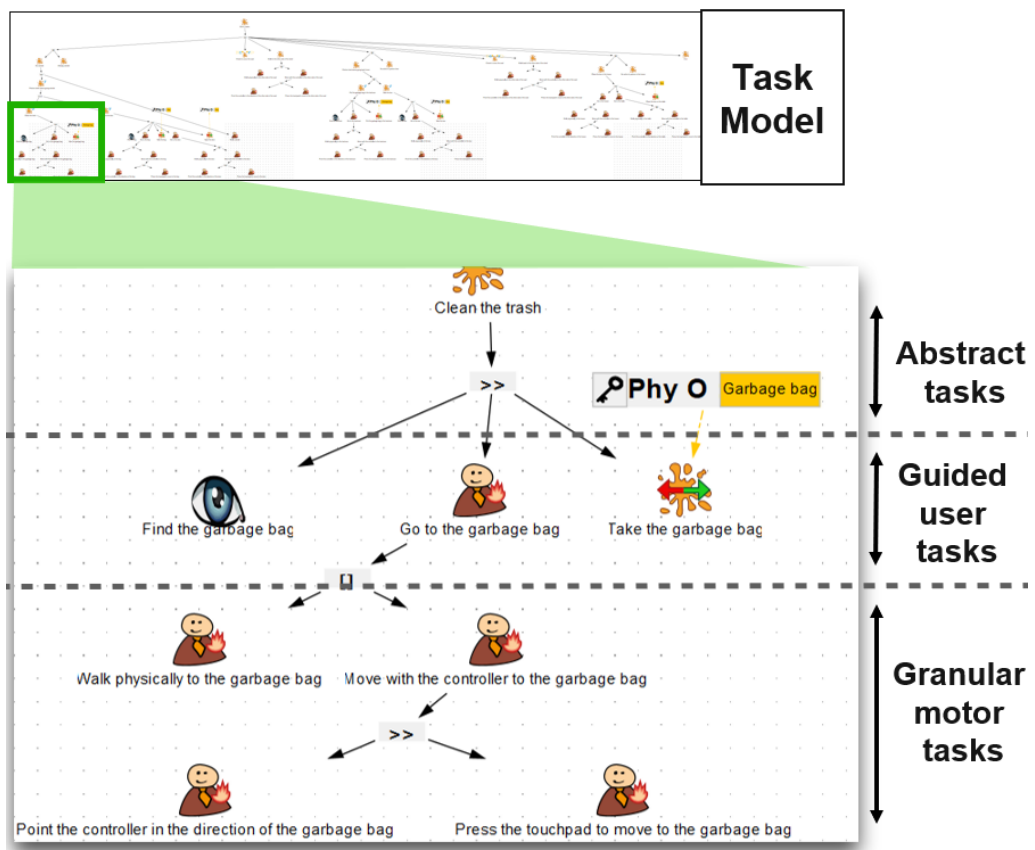


Figure 3.15 – Modeling tasks supported by the system for the study in a task model. The task model defines the precise the order of tasks, and the subdivision of the tasks in various levels of granularity from abstract to low-level motor tasks. In the study, instructions are given at the level of “Guided User Tasks”, and the lower granular motor tasks act as a cursor for intentionality analysis.

3.2.2.2 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 behavior.

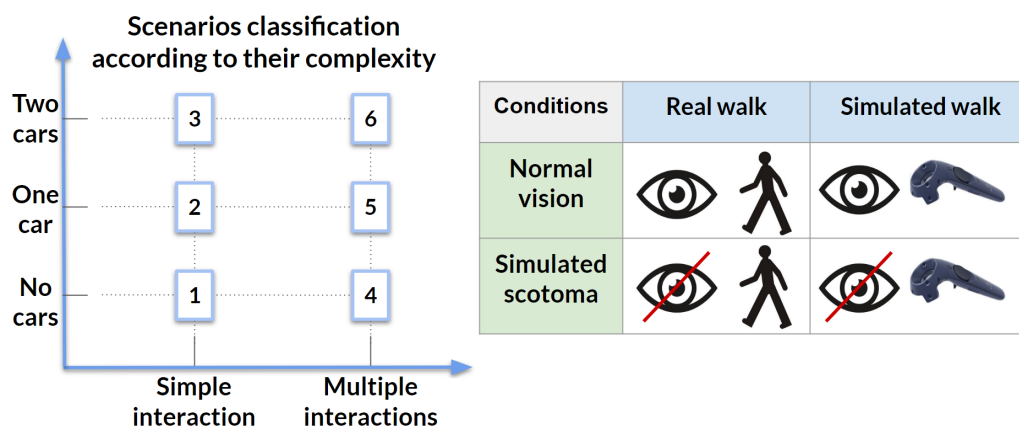


Figure 3.16 – (Right) Our study was comprised of four conditions as a combination of real or simulated walking, and with or without a simulated scotoma as a low vision condition. (Left) Each condition involved six scenarios with varying levels of interaction complexity and cognitive load (number of cars)

In order to build a dataset composed of a large range of user behaviors, six scenarios were designed around two axis of metrics we wanted to observe on the user experience shown in Figure 3.16 :

- **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 3.17.
- **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 3.18 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 3.19 :

- **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.
- **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 center of the vision of the user, as shown in Figure 3.20.



Figure 3.17 – Top-down view of the two different type of environments included in the scenarios. **One lane road** is used for scenario with 1 or less cars, **Two lane road** is used for 2 cars scenarios.

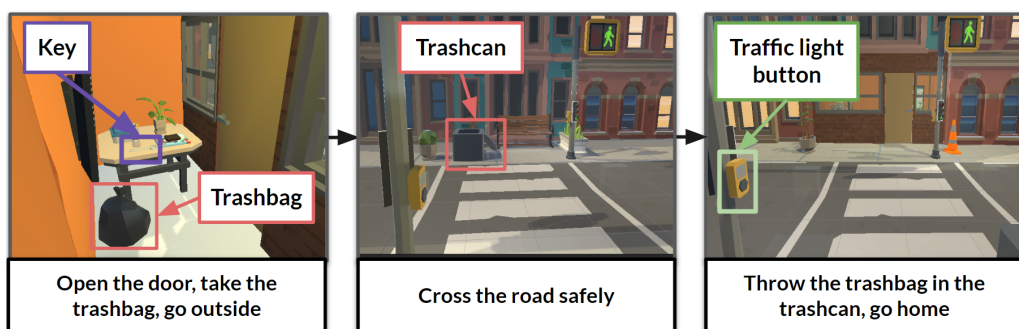


Figure 3.18 – Participant point of view of the scenario #4 from the study featuring multiple interactions (picking up the trash bag and pushing the traffic button) and crossing a single lane street with no cars.

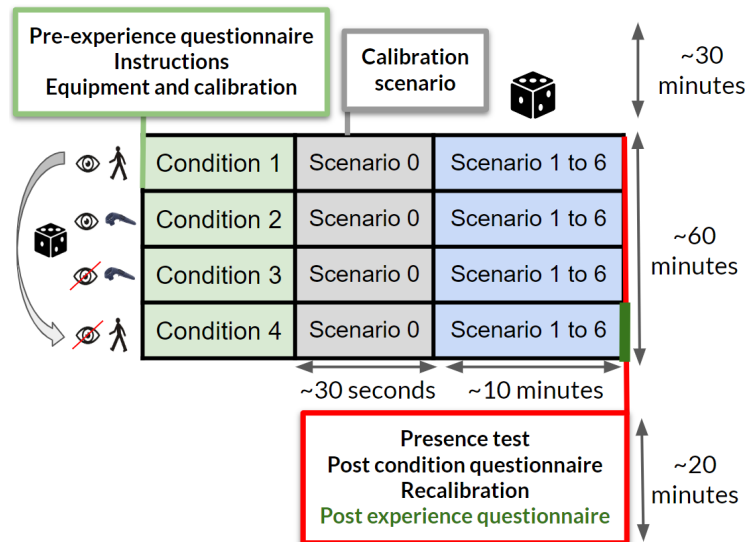


Figure 3.19 – The study protocol consists of three main stages : (1) pre-experience preparation including signing the informed consent, a questionnaire and equipping the headset and sensors (2) the study involving seven scenarios per condition (the first of which is a calibration scenario), two perspective 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 randomized. The entire study lasted roughly two hours.



Figure 3.20 – (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 center 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.

3.2.2.3 Questionnaires

The questionnaire was split into three parts : pre- and post-experience questionnaires, and post-condition survey that were identical following each condition. The questions came globally from existing studies of presence (Witmer & Singer, 1998), emotions (SAM (P. Lang, 1980)), simulator sickness (SSQ (Kennedy et al., 1993), and technology acceptability (UTAUT2 (Venkatesh et al., 2012) with the validated French version (Hayotte et al., 2020)). The full list of items included in the surveys is listed in Table 3.3.

3.2.2.4 Study protocol

Before the study, 10 pilot studies were carried out (four women and six men) from researchers and students in the project, from diverse backgrounds of computer science, cognitive science, neuroscience, and sport science. These pilot studies were repeated over a two month period involving iterative improvements to the study, refining of the protocol and surveys, and finally an all-hands co-design session with project members in December 2022 to validate and approve the final version of the study before the official launch in January 2023.

With all the above elements, the final experimental protocol is presented in Figure 3.19. 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.

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 3.14) to prevent any collision with walls or equipment.

At the beginning of each condition, a training scenario (numbered 0) was presented to the participant to help them discover the environment, familiarize 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 3.20. The calibration was designed after numerous training 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 training 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-randomized 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 (De Beni et al., 2006), 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 deviation between the participant's arm and the correct direction would indicate a high 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.

TABLE 3.3 – We designed surveys for pre-experience, post-experience, and post-condition (following each condition). The participants were allowed to skip questions if they did not wish to respond. The UTAUT2 questionnaire is detailed in Table 3.4.

Questionnaire	Questions	Response options
Pre-experience (T1)		
Study Information	User ID	Number
	Study Language	English / French
Demographics	Gender	Female / Male / Neither / I prefer not to say
	Age group	20-24/.../60+/ prefer not to say
Previous experience with VR and video games	Have you ever used a virtual reality headset?	Never / Rarely / Sometimes / Frequently / Always
	Have you every played 3D video games?	
Motion sickness	Do you experience motion sickness? (e.g., playing video games, transportation)	Open question
	What situations cause motion sickness?	
Technology acceptability (UTAUT2)	See Table 3.4	
Post-condition survey		
Condition information	Setup	RW / SW ; NV / LV
NASA Task Load Index	<ul style="list-style-type: none"> - To what extent was the task mentally demanding? - To what extent was the task physically demanding? - To what extent was the task hurried or rushed? - To what extent did you succeed in carrying out the task requested of you? - To what extent did you have to be invested to attain your level of performance? - To what extent were you anxious, discourage, irritated, stressed, or annoyed? 	10-point scale from (1) strongly disagree to (10) to strongly agree
Emotion	How do you consider the emotions experienced?	5-point scale from (1) calming, (3) neutral, to (5) intense
	How do you evaluate the positivity of the emotions experienced?	5-point scale from (1) negative, (3) neutral, to (5) positive
SSQ	<ul style="list-style-type: none"> - I felt ocular (visual) fatigue - I felt physical fatigue in the shoulders, back, and/or limbs - I have a headache - I had difficulties concentrating - I felt nauseated - I felt general discomfort 	5-point scale from (1) strongly disagree to (5) strongly agree
Perception of the experience	<ul style="list-style-type: none"> - I found the task interesting - I found the task repetitive - The motion capture equipment interfered with the tasks - The physiology sensors interfered with the tasks - The 3D scenes were realistic - I felt that I was really walking in the 3D environment - I felt that I was really interacting with objects 	
Difficulties during the task	Did you encounter any difficulties understanding the task or carrying them out?	Open question
Feedback	Do you have any other feedback on this conditions	
Post-experience (T2)		
Presence questionnaire	Please evaluate your feeling of being present in the virtual environment	7-point scale from (1) not at all present to (7) as if it were a real place
	During the duration of this experience, did you often think you were really in the virtual environment?	7-point scale from (1) never to (7) all the time
	During the experience, which feeling was more dominant : being in the virtual environment of the the experience, or being elsewhere?	7-point scale from (1) being elsewhere to (7) in virtual space
	Concerning your memory of being in the virtual space : to what extent is the structure of the memory similar to that of places you have been physically today?	7-point scale from (1) not at all to (7) exactly like a real memory
Emotional state	Interested, Anxious, Excited, Annoyed, Irritated, Enthusiastic, Alerted, Inspired, Attentive	4-point scale from (1) not at all to (4) very much
Technology Acceptability (UTAUT2)	See Table 3.4	
Feedback	Do you have any other feedback or suggestions?	Open question

TABLE 3.4 – The UTAUT2 technology acceptance questionnaire used in T1 and T2. Each question is responded on a 7-point scale from (1) strongly disagree to (2) strongly agree.

Subscale	Questions
Performance expectancy	<ul style="list-style-type: none"> - I find VR useful in my daily life - Using VR helps me accomplish things more quickly - Using VR increases my productivity
Effort expectancy	<ul style="list-style-type: none"> - Learning how to use VR is easy for me - My interaction with VR is clear and understandable - I find VR easy to use - It is easy for me to become skillful at using VR
Social influence	<ul style="list-style-type: none"> - People who are important to me think that I should use VR - People who influence my behavior think that I should use VR - People whose opinions I value prefer that I use VR
Facilitating conditions	<ul style="list-style-type: none"> - I have the resources necessary to use VR - I have the knowledge necessary to use VR - VR is compatible with other technologies I use
Hedonic motivation	<ul style="list-style-type: none"> - Using VR is fun - Using VR is enjoyable - Using VR is entertaining
Price value	<ul style="list-style-type: none"> - VR is reasonably priced - VR is a good value for money - At the current price, VR provides good value
Habit	<ul style="list-style-type: none"> - The use of VR has become a habit for me - I am addicted to using VR - I must use VR
Behavioral intention	<ul style="list-style-type: none"> - I intend to continue using VR in the future - I will always try to use VR in my daily life - I plan to continue to use VR frequently

3.2.3 Discussion

There is in the literature a lack of frameworks made for the comprehensive understanding of the embodied experience for behavioral studies : a framework containing the complete workflow from the design of the virtual scenario to the data recording, and finally in the analyses. Comparable frameworks could open up in-depth analysis possibilities to address the complex interactions between scene context, task design, and the user's experience. Most systems and usages we surveyed in the related works 2.1.2 (e.g., (Y. Lang et al., 2018; Wu et al., 2021)) focus on a subset of these aspects in a well-defined situation of usage. Due to the highly controlled environment inherent in virtual reality, it is imperative to provide sufficient control over the experience and tools to comprehend the system and tailor it to designer's research hypotheses, while avoiding any unintended external factors that may impact the final outcomes.

In this work, we presented an application of our framework enabling embodied interaction use and observation through a study. A total of forty participants completed the study in all the four conditions (normal vision, simulated low vision, physical walk, simulated walk). All participants performed and enjoyed the physical walk, and all the multimodal metrics were successfully collected.

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 valuable component for the follow-up studies.

The focus of the presented work is on the framework, its deployment in an actual study, and how the technical platform was established, incarnating Dourish's theory. Dataset creation and data analysis will be the focus of the next chapter of the thesis. Questions on how we can analyze granular data collected from sensors will be key to characterizing a comprehensive embodied experience and moving towards a better understanding of user intentions in VR.

3.3 Conclusion

The first contribution of this chapter, addressing RQ1A, is the conception of the GUsT-3D framework for the design of semantically rich experiences through an ontology, to supports designers in the creation, management, and observation of immersive experiences. This system opens ways to enrich the object and interaction properties of embodied VR experiences and create records of the scene graph during an experience to enable granular observations.

The second contribution of this chapter, addressing RQ1B, is the validation of the feasibility of the presented framework and technical platform for embodied VR study. This study enables the usage of physical movement and physical interactions in large immersive environments. Through the design and encoding of a task model to create segmented scenarios and data, it enables the capture of behavioral indices related to embodied interactions on a granular level. This work opens possibilities for in-depth analysis of user experience in the embodied context in VR, a subject investigated in the next chapter of this thesis.

Contributions The publications and publicly available framework presented through this chapter :



F. Robert, H.-Y. Wu, L. Sassatelli, and M. Winckler. "An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual Reality". ACM IMX 2023, Nantes, France. **Best Paper Award** 🏆 [hal-04102737](#)

H.-Y. Wu, **F. Robert**, T. Fafet, B. Graulier, B. Passin-Caneau, L. Sassatelli, M. Winckler. "Designing guided user tasks in VR embodied experiences". ACM EICS 2022, Biot, France. [hal-03635452](#)



GUST-3D Framework : <https://project.inria.fr/creattive3d/gust-3d/>

Analysis of Intentionality from large immersive dataset

The second subject investigated in this thesis concerns the analysis of the embodied interactions using multimodal data, to obtain a better comprehension of the user experience lived and more toward a comprehension of the user intention in an embodied VR context. Contrary to traditional media, the VR system, through embodied interactions using the body as a communication device, proposes multisensory engagement and complex interactive possibilities. This specificity allows, by using the body as a semiotic device, the possibility of obtaining a better representation of the user's lived experience and intention through diverse types of behavioral metrics. Based on Paul Dourish's theory of embodiment, issues with the **intentionality** component result in gaps between the user's intentions and interpretation when living the experience and the designer's comprehension of this lived experience. To bridge this gap between users and designers in embodied experiences, an efficient solution is to capture and analyze multimodal data (e.g., gaze movement, body motion, 3D environment interactions, heart rate) to get a deeper insight into VR embodied experiences by observing the link between behavioral and contextual metrics.

This chapter investigates RQ2 – *How can we analyze intentionality from the user's embodied interactions in immersive experiences using multimodal metrics, and how can this analysis be used to provide feedback to designers?* – through three sub-questions :

- **RQ2A** : How to build an immersive dataset in a way that enables in-depth analysis of embodied interactions ?
- **RQ2B** : How to granularly analyze the user behavior of a lived embodied experience ?
- **RQ2C** : How to use the analysis to support designers in the refinement of the immersive experience ?

We present two types of contributions. The first type of contribution concerns the creation of two datasets, to address the challenge of collecting and observing data related to the embodied interactions relevant to RQ2A. The first dataset, PEM360, results from participants watching 360-degree videos in VR. This dataset is presented along with preliminary analysis on attention and emotion applied to this dataset in Section 4.1. The second dataset, CREATTIVE3D, results from the VR road crossing study presented in the previous chapter (Section 3.2), along with preliminary analysis of multimodal data in Section 4.2. The second type of contribution is our conceived task-based methodology, which addresses the challenge of finely analyzing the embodied interactions in response to RQ2B and RQ2C. For this purpose, this methodology opens the possibility of granularly measuring the link between behavioral and contextual metrics through multimodal data. We applied this methodology to the CREATTIVE3D dataset to characterize the embodied experience, presented in Section 4.3.

ID	Valence	Arousal	Start (s)	End (s)	Duration (s)	YouTubeID
12	7	4.6	5	103	98	T-aOVE22IEw
13	4.92	4.08	4	131	127	GJGxfGEa9Y
17	5.22	5	5	69	64	g7btxyIbQQ0
23	7.2	3.2	8	143	135	CDfsFuDuHds
27	6	1.6	60	180	120	QxxXu_B-ZA
73	6.27	6.18	9	70	61	bUiP-iGN6oI
32	6.57	1.57	40	130	90	-IrUYM-GjU

TABLE 4.1 – Details of selected videos. Videos are accessible at [youtube.com/watch?v=\[YouTubeID\]](https://www.youtube.com/watch?v=[YouTubeID]).

4.1 PEM360 : A dataset with emotion, motion and attention data

The initial pair of dataset-analysis work on PEM360 was conducted early in this thesis. Its implementation and behavioral metrics analysis served as a foundation for the subsequent pair of dataset-analysis work on CREATIVE3D by building and utilizing VR studies, behavioral data collection, and correlation analysis.

The PEM360 dataset is built with the purpose to investigate on ways to optimize storage and distribution of immersive 360 VR contents, by observing and understanding the connection between user attention, emotion and immersive content. PEM360 is composed of user head movements and gaze recordings in 360° videos, along with self-reported emotional ratings of valence and arousal, and continuous physiological measurement of electrodermal activity and heart rate. The stimuli are selected based on high-level and low-level content saliency to enable the spatio temporal analysis of the connection between content, user motion and emotion. This dataset comes along with a set of software tools to pre-process the data of gaze, electrodermal activity and content.

The entire collection of artifacts is presented as Python tools and notebooks to enable reproducibility of the data processing. The dataset and tools are available in a public GitLab repository : <https://gitlab.com/PEM360/PEM360/>.

4.1.1 User experiment and dataset description

We conducted a controlled, indoor laboratory experiment where users watched 360° videos in a VR headset. We collected eye movement (EM), head movement (HM), heart rate (HR) and skin conductance (EDA) data as well as emotion annotations of valence and arousal. The seven videos are selected to enable several levels of content analysis and description, to correlate with user motion and emotion.

4.1.1.1 Equipment

Recordings of head and eye movements have been made with a FOVE headset, equipped with an eye-tracker with a 120Hz acquisition rate, and tethered to a desktop computer. A Unity3D scene was used with a 360° sphere object to display the videos. We use the FOVE Unity plugin to record head and gaze positions.

Recordings of EDA and optical pulse have been made with a Shimmer3 GSR+ sensor with a frequency range of 15.9Hz and 51.2Hz, respectively. All of the measurements were re sampled to 100Hz for analysis. The apparatus is depicted in Fig. 4.1.

4.1.1.2 Participants

The experiment was carried out with a total of 34 users, in which 31 had complete data (10 women, 20 men, 1 non-binary ; 18-29 years old, $M=24$, $SD=3.26$). 19 of them had a normal vision, 9 had corrected to normal vision and 3 did not have a normal vision. Most of them played games but rarely or never in VR, and the majority have seen only one or two 360° videos before the experiment. Participants received monetary compensation for their time. The seven videos were experienced by all 31 users for their entire duration (60 to 135 seconds, see Table 4.1).



Figure 4.1 – Shimmer3 GSR+ used to record EDA and optical pulse. Gray wires connect the EDA sensor, white wire connects the pulse sensor.

4.1.1.3 Procedure

The lab experiment started with a pre-questionnaire assessing the user’s background with VR and checking for visual deficiencies. Eye tracking calibration was done using the FOVE software for each user before beginning the experiment to make sure the eye tracking data is properly recorded. The VR experiment systematically started with a low-arousal (relaxing) video (ID 32) to bring EDA and HR levels to a user-relative baseline. The remaining six VR videos were then experienced in a random order by every user. Users were in standing position during the experience and could freely explore in 360° while holding the back of a chair to maintain balance and orientation. The videos were played without audio. After each viewing, the headset was removed and the SAM scale presented for arousal and valence rating. At least a 1-min break outside of the headset was observed between videos.

4.1.2 Dataset structure

The resulting dataset PEM360 is provided with the structure shown in Fig. 4.2. The `raw_data` folder contains 34 folders, one for each user. User folders contain a Shimmer CSV file containing the EDA and optical pulse data recorded over all the 360° videos experienced by the user, and seven CSV files, one per video, containing the gaze and head motion data recorded during the corresponding video. Entries in the CSV files include system timestamps to synchronize the data modalities for analysis.

Valence and arousal ratings of each user for each video are stored in the root folder under `graded_valence_arousal.csv`. Finally, the root folder PEM360 also contains the Python Jupyter notebook providing the software tools described in following Section 4.1.3, and the entire data processing workflow to reproduce the analysis presented in Sec. 4.1.4.

4.1.3 Pre-processing software

Along with the data, we provide a Jupyter notebook to reproduce the entire processing of head and gaze data, EDA, ratings of valence and arousal, and the code to produce saliency maps from the content. We present here only the most relevant for the thesis subject.

4.1.3.1 Processing gaze data

The HM and EM, 3D positions are logged in Cartesian coordinates $(x, y, z) \in \mathbb{R}^3$. The dataset also propose functions :

- to convert the positions from Cartesian to Eulerian (ϕ, θ, ψ) denoting respectively yaw, pitch and roll,
- to obtain speed and acceleration over yaw and pitch,
- to obtain global speed and acceleration by computing the derivatives of the orthodromic distance,
- to represent rotational motion with quaternions (hence enabling to compute non-linear motion on the sphere as changes in quaternion rotational axis).


```

PEM360/
├── emotional_maps/
│   ├── compute_emotional_map.py
│   ├── requirements.txt
│   └── utils_shimmer.py
├── raw_data/
│   ├── LICENSE
│   ├── user_01/
│   │   ├── shimmer.csv
│   │   ├── video_12.csv
│   │   ├── video_13.csv
│   │   ├── video_17.csv
│   │   ├── video_23.csv
│   │   ├── video_27.csv
│   │   ├── video_32.csv
│   │   └── video_73.csv
│   └── user_02/
├── saliency_maps/
│   ├── compute_gt_saliency.ipynb
│   ├── compute_hl_saliency.py
│   ├── compute_ll_saliency.py
│   ├── nfov.py
│   ├── pySaliencyMap.py
│   ├── pySaliencyMapDefs.py
│   └── requirements.txt
├── LICENSE
├── README.md
├── design_of_experiments.csv
├── graded_valence_arousal.csv
├── notebook.ipynb
├── requirements.txt
└── resample_data.ipynb

```

Figure 4.2 – Folder structure of the dataset with main files.

4.1.3.2 Processing EDA data

The EDA signal is the raw measurement of skin electrical conductance in micro-Siemens (μS). Two main components can be distinguished in an EDA signal (Boucsein, 2012; Braithwaite et al., 2013): the tonic level, also called skin conductance level (SCL), varies slowly and represents slow autonomic changes that may not be associated with stimulus presentation; and the phasic level, which represents faster changes in EDA, and can better reflect the impact of successive stimuli. Raw EDA, phasic and tonic components are shown in Fig. 4.3-top and 4.3-center. We use the Python toolbox Neurokit (Makowski et al., 2021) to process EDA data, which uses the `cvxEDA` method to extract the phasic component. Finally, the physiological arousal to be analyzed in connection with experimental stimuli can be assessed from several metrics on the phasic level, such as peak frequency, duration and amplitude. This is called the skin conductance response (SCR), and can be defined in several ways. In our code, we choose to compute instantaneous SCR as the absolute value of the first-order time derivative of the phasic component, shown in 4.3-bottom. Note however that the code can easily be modified to implement other definitions of SCR from the phasic component. The obtained SCR is therefore a time series for every user-video pair. This enables analysis with SCR averaged over time for each such pair (as often done), or on a time-dependent basis.

4.1.4 Preliminary analysis of the data

In this section we first verify the validity of our data and correspondence with the original dataset and between arousal and EDA. We then exemplify possible analyses of correlation between motion and emotion, and between attention, content saliency and emotion.

4.1.4.1 Data validation

Reliability of the collected ratings We verify the reliability of the collected arousal and valence by assessing the similarity of the user ratings for each video. This is achieved with the intra-class correlation coefficient (ICC), with classes corresponding to the 360° videos. ICC estimates based on mean ratings with a two-way mixed effects model are 0.96 (95% CI 0.87-0.99) for arousal and 0.88 (95% CI 0.72-0.98) for valence. According to Koo and Li’s guidelines (2016), this is excellent and good inter-rater agreement, respectively.

Agreement between collected ratings and original dataset Fig. 4.4 shows the valence and arousal ratings of our users as a boxplot for each video, along with a red dot representing the corresponding average values

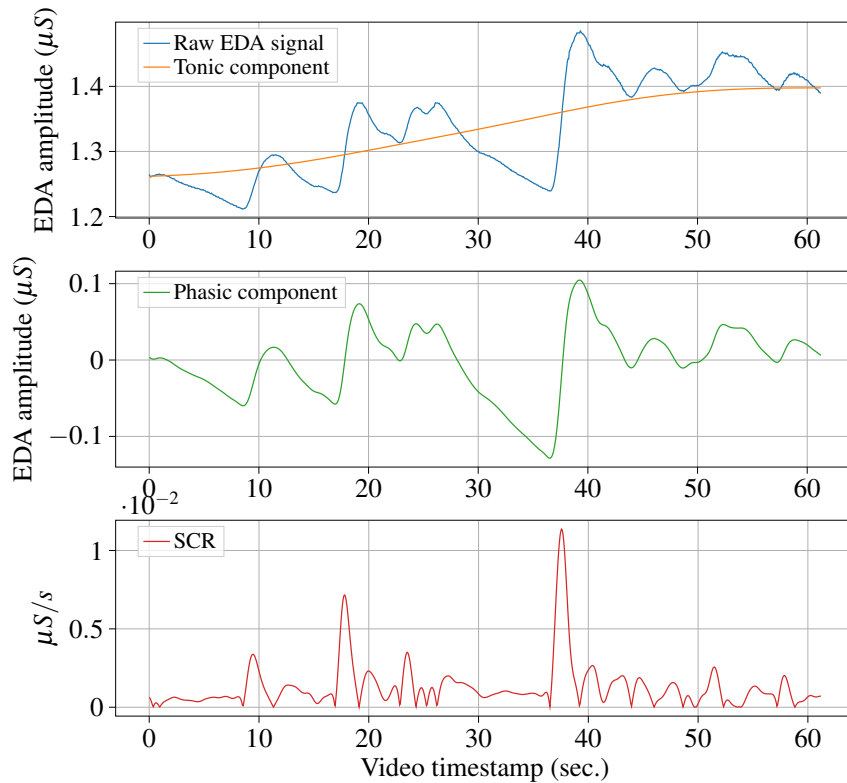


Figure 4.3 – EDA signal recorded for user 03 while watching video 73. The three graphs from the top show the raw EDA data and the tonic component, the phasic component and the SCR (absolute value of phasic first derivative).

available in the original dataset (B. J. Li et al., 2017). We observe the good agreement between both sets, as the latter are all the times but one in the inter-quartile value range of our data. We also compute the median of the root square difference of averages of our valence and arousal ratings with the corresponding averages from (B. J. Li et al., 2017). This median is 1.17 (within a range of 1 to 9), showing the agreement between both.

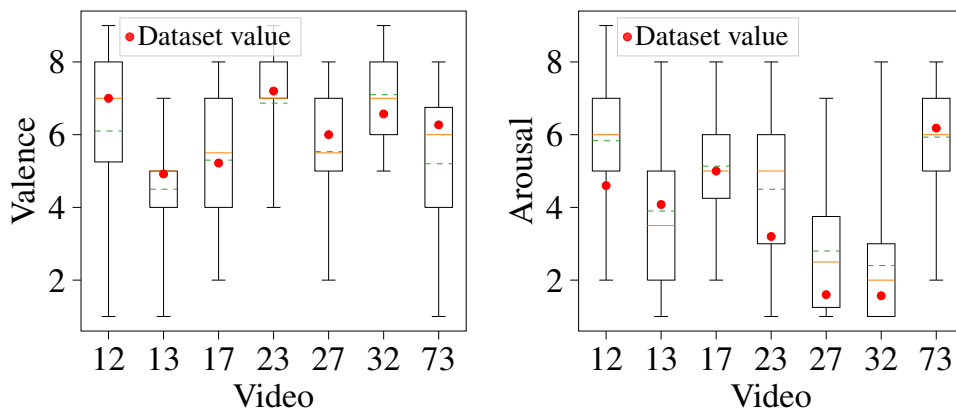


Figure 4.4 – Arousal and valence ratings by users for each videos. The green dotted line corresponds to the mean and the orange solid line to the median.

4.1.4.2 Connecting EDA with graded arousal

We investigate the correspondence between SCR and arousal ratings. We gather the average SCR values $SCR_{u,v}$ for every pair (u, v) of user u and video v , and corresponding graded arousal $GA_{u,v}$. First, we average both variables over all users for every video, and obtain seven sample pairs (GA_v, SCR_v) , shown in Fig. 4.5-left. We verify as did Toet et al. (2020) that the video ranking according to mean graded arousal is similar to the video ranking according to mean SCR. We compute the Spearman correlation coefficient (CC) between GA_v and SCR_v for all videos. The Spearman CC between mean graded arousal and mean SCR is $(0.92, p = 0.003)$. Such correlation is significant ($\alpha = 0.05, \beta = 0.2$) from 7 samples (UCSF, 2021; Jeffrey J. Walline, 2001).

We then consider the 217 sample pairs $(GA_{u,v}, SCR_{u,v})$. It is interesting to observe that the Pearson or Spearman CCs do not show any correlation between these pairs. Looking more closely at the data, we identify that the mean level of SCR per user, $SCR_u = \mathbb{E}_v[SCR_{u,v}]$ (averaged over all videos), varies significantly over the users ($M = 6.0e - 4, S = 6.2e - 4$). With the rationale that the *excitability* of a user is person-dependent and impacts the absolute SCR values, we verify whether the SCR variations relative to this individual's mean are better associated with graded arousal. To do so, we define centered SCR as $cSCR_{u,v} = SCR_{u,v} - SCR_u$, and do the same with graded arousal $cGA_{u,v} = GA_{u,v} - GA_u$. Fig. 4.5-right represents the scatter plot of $cSCR_{u,v}$ against $cGA_{u,v}$. The Spearman CC between both is $(0.25, p < 0.001)$.

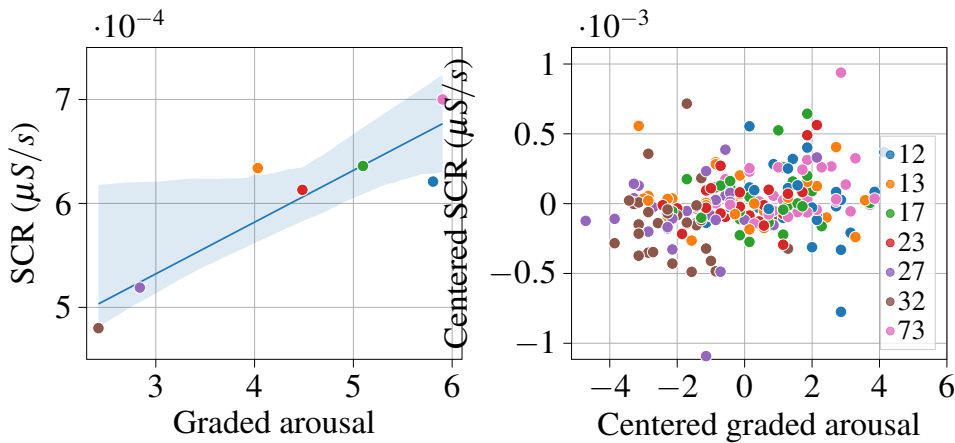


Figure 4.5 – Dots colors code for video ID (legend on the right). Left : Scatter plot of SCR_v against GA_v . The shaded area represents the 95% CI of the linear regressor (solid blue line). Right : Scatter plot of $cSCR_{u,v}$ against $cGA_{u,v}$.

4.1.5 Discussion

In this section, we have presented our PEM360 dataset of in 3DoF context. This immersive dataset was built in a way enabling the observation of the link between user's attention, emotion, and the content shown in VR, to investigate content distribution optimization. We exemplified some analyses to demonstrate the potential insights that can be drawn from PEM360. This work served as a first step for this thesis on embodied interaction investigation, by performing a study VR, collecting multimodal data, and analyzing correlation on this data. The built technical platform and acquired knowledge served as a basis for the work on the CREATTIVE3D dataset presented in the next section.

4.2 CREATTIVE3D : Exploring, walking, and interacting in virtual reality with simulated low vision

The second dataset presented is the CREATTIVE3D dataset, corresponding to the data collected during the virtual road crossing study presented previously in Section 3.2. This dataset is built to explore the understanding of the embodied VR experience in 6DoF conditions, and is for now the largest dataset (in terms of frames, number of subjects, and duration) on human motion in fully-annotated contexts. To our knowledge, it is the only one conducted in dynamic and interactive virtual environments, with the rich multivariate indices of behavior as mentioned above.

The pre-processed and synchronized dataset, along with some tools and usage examples, is available on Zenodo : <https://zenodo.org/records/10406560>.

4.2.1 User experiment and dataset description

During the study, multiple modalities of data were collected. Each participant performed a total of 24 scenarios in six different scenarios with varying interaction complexity and cognitive load, and in four different conditions : physical walk, simulated walk, normal visual, simulated low vision. The full detail on the study performed can be found in Section 3.2. The Table 4.2 summarizes all the data modalities and data characteristics.

We recruited 40 participants (20 women and 20 men) between ages 18 and 34. Participants needed to have normal or corrected to normal binocular vision, no motor sensory difficulties, but no other restrictions. Most of them played games, but rarely or never in VR. In the end, all 40 recruited participants were able to complete the entire 24 scenarios of the study.

TABLE 4.2 – The collected data modalities, equipment used, format, and the logging frequencies.

Type	Equipment	Format	Frequency	Description
System logs	GUsT-3D	.json	10 Hz	objects position, objects state, objects interactive properties, user position, user interaction, current task, objects in user visual field
Physiological sensors	Shimmer GSR+	.csv	EDA 15.9Hz HR 51Hz Resamp. 100Hz	Heart Rate (HR), electrodermal activity (EDA)
Motion capture	XSens Awinda Starter	.csv .bvh	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). The .csv files contain the processed absolute joint coordinates calculated from the .bvh file, as well as the transformation matrix to spatially transform the motion data into coordinates of the 3D scene.
Gaze and head tracking	HTC Vive Pro Eye	.json	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; Head position and rotation vectors
Surveys	LimeSurvey	.json	N/A	Surveys described in Table 3.3 as well as experimenter observation notes

4.2.2 Dataset structure

The resulting dataset CREATTIVE3D is provided with the structure shown in Figure 4.2.2. The data package is contained in a single repository with documentation, tools, a number of usage examples, and data of users 1 to 40. User folders contain 24 folders for each performed road crossing VR scenarios. Each scenario folder contain the raw data recorded during the concerned scenario : body motion, system logs, head motion, gaze motion, heart rate, and skin conductance. Entries in the raw data files include system timestamps to synchronize the data modalities for analysis.

Questionnaire data of each user are stored in the root of the Data folder. Finally, the root folder CREATTIVE3D also contains the Python scripts providing software tools described in following Section 4.2.4, some usage examples, and the 7 scenario scenes used for the study in point cloud format.

```

GUsT-3D Data
├── Readme.md
├── Tools (spatial_sync.py) script to spatially synchronize motion data
├── Usage examples
├── Scenarios: scenes as point cloud
├── Data
│   ├── questionnaire (.lss .csv): survey content and results
│   ├── validation.csv: CSV with all the data issues
│   ├── config.json: order of scenes for each participant
│   └── User 1
│       ├── Condition_RWNV (RW/SW = real/simulated walking; NV/LV = normal/low vision)
│       │   ├── Scene_SI0V
│       │   │   ├── U1_RWNV_SC_dataframe.csv, dataframe125Hz.csv: Example dataframe for the user
│       │   │   ├── U1_RWNV_SC_motion.bvh: Segmented raw motion data
│       │   │   ├── U1_RWNV_SC_motion_hierarchy.csv, pos.csv, rot.csv: Absolute joint coordinates
│       │   │   ├── U1_RWNV_SC_motion_transf.json: Transformation matrix for spatial synchronization
│       │   │   ├── U1_RWNV_SC_shimmer.csv: Shimmer data
│       │   │   ├── U1_RWNV_SC_LogI.json, U1_RWNV_SC_LogP.json: GUsT-3D system logs
│       │   │   ├── U1_RWNV_SC_head.json: Gaze tracking - Head data processed
│       │   │   ├── U1_RWNV_SC_gaze.json: Gaze tracking - Gaze data processed
│       │   │   ├── U1_RWNV_SC_timesync: Gaze tracking - timesync
│       │   │   └── U1_RWNV_SC_por.json: calculated points of regard
│       │   ├── Scene_SI2V / CI0V / CI1V / CI2V
│       │   └── Condition_RWLV / SWNV / SWLV
│       └── User 2, User 3, ...User 40

```

4.2.3 Technical Validation

In order to ensure high data quality, multiple validations were put in place, including ten pilot runs preceding the actual study, detailed list of data issues, conducting data synchronization, and verifying the internal consistency of the questionnaires used. We detail each of these validations below.

Data issues and transparency. Despite the protocol devised and iterative testing, technical issues can and do occur in complex studies such as the one we carried out, due to accidental manipulations on the experimenter's side, hardware crashes, issues in proprietary software, latency, and individual difficulties for participants – most of which is normal and outside of the control of the study design.

With the objective of being full transparent on the data, all the incidents were reported in the production process in a structured way. The data file is provided `validation.csv` giving a comprehensive list of various data issues, and their impact. An empty cell indicates no issues were observed.

Out of the data on 960 scenarios (24 per participant), we observed the following more critical issues that render one datatype unusable for a single scenario of a participant :

- Gaze : missing eye or head data (5 scenarios), inconsistent head coordinates (1 scenario) and timesync (6 scenarios).
- Questionnaire : missing questionnaires in the post-experience survey due to connection problems (8 participants)
- GUsT-3D logs : missing or irregular files (2 scenarios and all data for 2 participants)
- Physiology : oversampling (1 participant), heart rate data not dependable

The missing timesync information can be corrected by applying an average delta of 28000 (28 ms) which will ensure data synchronization with an accuracy of within 6 ms (delta ranged between 22000 and 34000).

Hardware crashes also occurred for seven participants due to the wireless module, impacting maximum one scenario for the participant, mostly at the beginning of the study. The scenario was re-run after a hardware restart and the study continued without further issues.

4.2.4 Pre-processing and analysis software

Along with the data, we provide python scripts to reproduce our work on the creation of multivariate dataset, some preliminary analysis, and the granular visualization of behavioral data according to the context. We present here the most relevant ones for the thesis.

Multivariate dataframe. We provide with the source code the script *ex_dataframe.py* that facilitates the generation of a dataframe with the multivariate gaze, motion, emotion, and user log data for each participant. The resulting dataframe can then be used as input for data visualization, statistical modeling, machine learning, and other applications. A number of these applications are described in more details below.

Fine-grained user understanding. The rich recorded context allows us to have a fine-grained view of the user's current state, both from symbolic and continuous data. We take the example of electrodermal activity (EDA, a.k.a. skin conductance), captured using the Shimmer GSR+ module. Using NeuroKit2 (Makowski et al., 2021) we can process the EDA levels to separate the phasic and tonic components. In Figure 4.8 we show a graph of the evolution of the user's EDA throughout a single scenario. The example script to generate this graph *ex_EDA.py* is included in this package.

4.2.5 Preliminary analysis of the data

We exemplify the potential of analysis 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.

4.2.5.1 Dataset analysis

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. 3.16. The EDA (electrodermal activity) signal is the raw measurement of skin conductance in micro-Siemens (uS).

4.2.5.2 Demographics and questionnaire results

A total of 40 people participated, but at the time of this preliminary analysis, 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. 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 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.

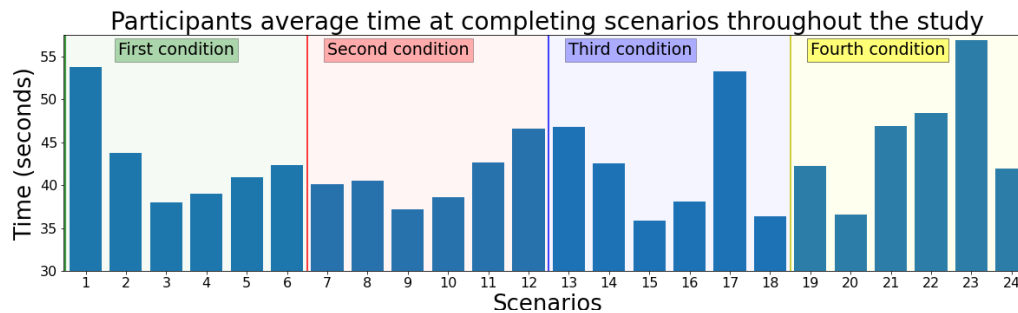


Figure 4.6 – The evolution of the average time spent by all participants over the study across conditions and scenarios in the order executed following the Latin Square assignment. It shows that the participants usually spent more time on the first scenario of each condition than the scenarios that immediately followed, except for the last condition.

4.2.5.3 Analysis of the user experience

Our large variety of data was synchronized by their Unix Timestamp and processed in a Jupyter notebook. The EDA was processed as presented in Section 4.1.3.2. Well known Python libraries were used for data processing, such as Pandas to organize dataframes, NumPy for mathematical functions, and Matplotlib for data visualization. Combining multimodal data (physiological, motion, gaze, questionnaire, and scenario interaction data) though a notebook allows a clear perception and analysis of the user embodied experience.

Learning curve As we can see from Figure 4.6, 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.

Relation between scenario and skin conductance In the top half of Figure 4.7, we can see results on the EDA in relation to the scenarios. Based on the original design, shown in Figure 3.16, 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 4.7, 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 seconds). The impact is however more noticeable on EDA, rising from -0.145 to -0.058 uS in RW, and from 0.052 to 0.146 uS 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 4.7, 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 4.7 top half (complete scenario) compared to bottom left side (road

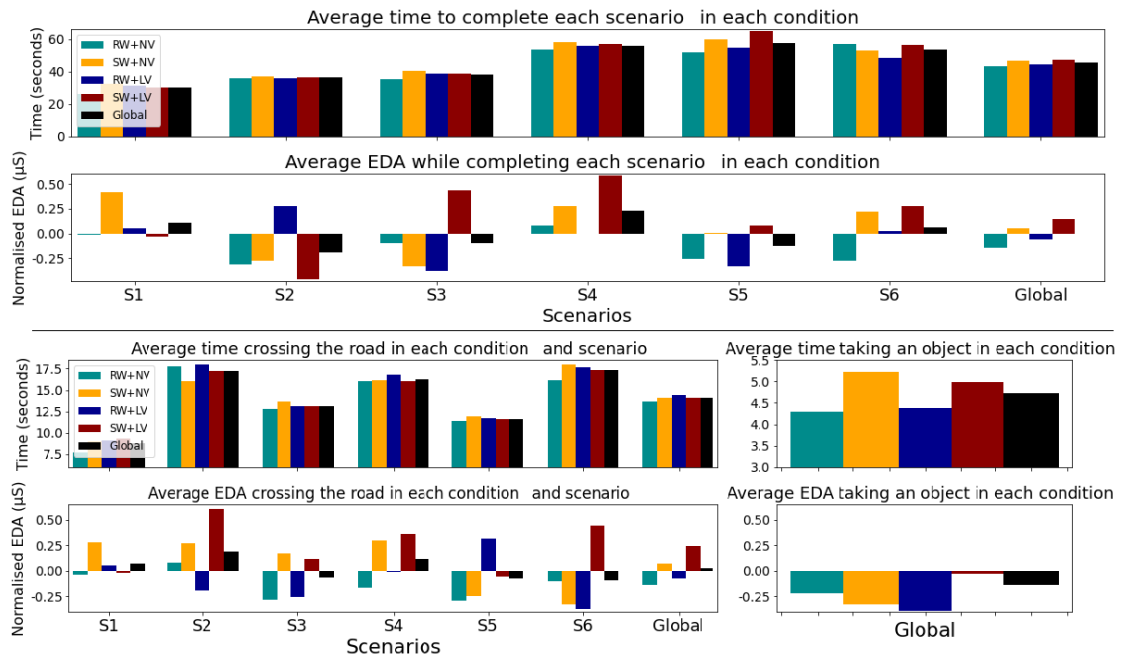


Figure 4.7 – (Top) Average time participants takes to complete scenarios for every condition and the electrodermal activity (EDA) measured during these tasks. (Bottom Left) Average time participants take to cross a road for every condition and the EDA measured during these tasks, (Bottom Right) compared to the average EDA and time to take an object.

crossing task), the difference in EDA rises from 0.291 uS to 0.377 uS. 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 4.7, 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 4.8 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.

4.2.6 Discussion

In immersive experience’s domain, multiple datasets exist in the literature, as presented in Section 2.2.1. Nonetheless, these datasets rarely include the behavioral and contextual metrics related to embodied interactions (i.e., gaze motion, body motion, and virtual interactions) in a controlled 6DoF context. The amount of 6DoF datasets in interactive virtual environments with the rich multivariate behavioral cues needed for a good understanding of embodied experiences is limited, as the majority focus solely on body motion data.

Methodologies to efficiently capture task-segmented and share datasets in virtual reality studies present two advantages : first, they can open many possibilities for the study of behavior in multitask scenario contexts, such as for simulated low vision (Delachambre et al., 2024) or wheelchair training scenarios (Arlati et al., 2020). Second, a collection-processing-analysis framework on multivariate indices would open the possibility of creating fully reproducible datasets in various contexts such that future studies using the same protocol can compare to existing results with sufficient confidence.

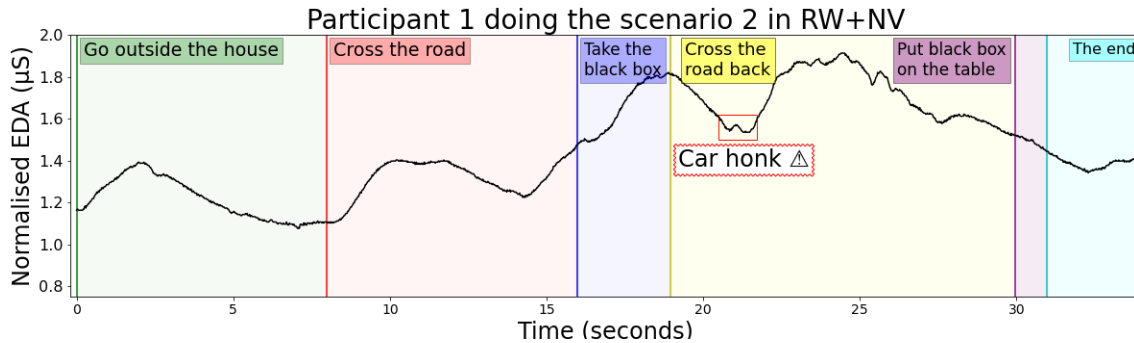


Figure 4.8 – The EDA of participant 1 during the scenario 2 in RW+NV condition, honked by a car while jaywalking during a red traffic light.

We have presented here our CREATTIVE3D dataset of in 6DoF context. This immersive dataset was built in a way enabling the synchronization of all the behavioral data with the contextual data on a granular level, to open in-depth embodied interactions possibilities. We exemplified some analyses to demonstrate the potential insights that can be drawn from CREATTIVE3D. Our task-based methodology is presented in the next section to open-up for more in-depth embodied experience analysis.

4.3 Task-based methodology to characterize immersive user experience with multivariate data

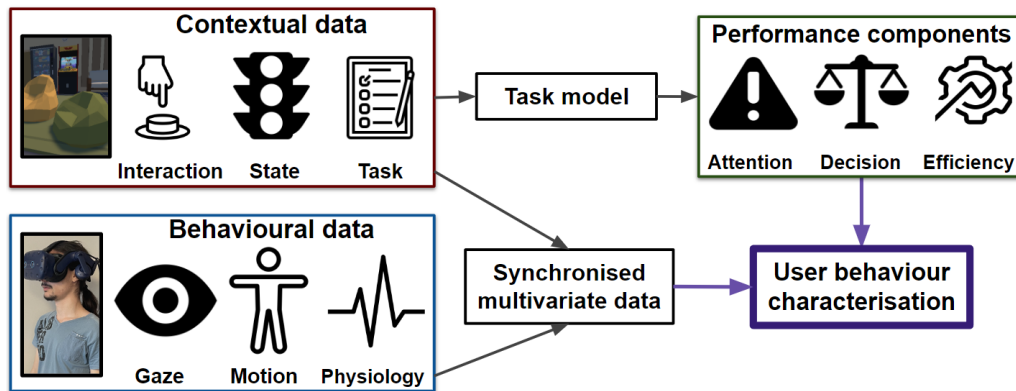


Figure 4.9 – Our task-based methodology comprises three steps : (1) synchronise behavioural data with context data, (2) use task model to define performance baselines, and (3) characterise user behaviour based on their performances using multivariate data.

Related to the **RQ2B** and **RQ2C**, the following work present our conceived task-based methodology applied on the previously presented dataset CREATTIVE3D (Section 4.2) to enable the characterization of the embodied user experience. This work’s contribution is the conception of a methodology using existing HCI domain task-based analysis applied to embodied VR context to analyze embodied interactions in multi-tasks experience in correlation with the observed performances.

When a user misstep occurs when carrying out a task, whether it is overlooking an important indication, forgetting necessary steps, or committing an error resulting in the failure of the task, finding the reason can be extremely difficult : the misstep can be linked to invisible cues such as the person’s physical, attentional, or emotional state, triggered by specific scene elements. In the case of training scenarios that involve a sequence

of varied, non-trivial tasks, carrying out a precise performance evaluation to gain user understanding becomes even more complex. Yet synthetic 3D environments in VR have this capacity, to allow the designers of a scenario to observe and highlight user behavioral indices in their relevant context. Building a proper understanding of granular user performance in the “lived” experience will bring multiple benefits : on the one hand, to better accompany users in their training and propose personalized feedback according to their performance and interaction profiles ; on the other hand, on the system side, design improved affordances and guidance to the observed user behaviors and needs.

We thus present our task-based methodology to help enrich and characterize the experiential aspect of user experiences (UX) that VR systems offer, by investigating the state of the system in relation to the state of the user (Hassenzahl & Tractinsky, 2006). Inspired by performance measures such as GOMS (Sauro, 2009) (Goals, Operators, Methods, and Selection rules) as well as task modeling approaches such as CTT (Santoro, 2005) and HAMSTERS (P. A. Palanque et al., 2011) – popularly used in engineering interactive systems –, our method decomposes high-level VR scenarios and user task into lower-level sub-tasks (or sub-sub-tasks) that can be completed by the user in different ways. The originality of our approach then involves the use of task models to support the analysis of user behavior on various levels of granularity (high and low-level tasks), aligning task execution in the scene context (e.g., current task, state of the scene, user position) with multivariate data describing the embodied UX (i.e. attention, motion, emotion). It is this alignment of user tasks with multivariate data that allows a better understanding of the performance in relation to behaviors for every low-level task (i.e., take an object, go somewhere) within a scenario. To evaluate a user’s performance for each task, we defined three performance components (i.e., efficiency, attention, decision) with set baseline values to determine the success criteria for a given task. Once the tasks and scenarios are settled, our approach allows the comparative analysis of task performance and multivariate data that jointly characterize the immersive UX.

The proposed methodology introduces three primary advances for the analysis of VR training scenarios :

- Definition of user performances following three components (i.e. efficiency, attention, decision) and baseline values for each component to observe the performance missteps on a low-level tasks
- Characterization of behavior using multivariate data based on the performance missteps
- Creation of user profiles based on observed link between behavior and performance missteps, to improve personalized feedback and VR experiences

The remainder of the contribution presentation is organized as follows. In the Section 4.3.1, we present the dataset required specificities to perform the presented analysis. In Section 4.3.2, we detail the task-model method for fine-grained VR experience analysis, and show how we can achieve a fine-grained characterization of user behavior based on task performance. Then in Section 4.3.3 we show the results of our analysis on the variation in metrics for different performance components (i.e., attention, decision, and efficiency), which leads to individualized behavioral profiles in relation to the performance missteps. Finally, we discuss these findings, their implications, and present the next steps of this work in Sections 4.3.4.

4.3.1 Dataset

The dataset CREATTIVE3D fulfills two important criteria for our approach as it includes multivariate data and these metrics can be spatio-temporally aligned with low-level tasks and scene context. These two particularities allows us for each user and each scenario to visualized finely the behavioral metrics in correlation with the context. The Figure 4.10 provides a sample visualization of the multivariate data for one user in one scenario : the raw EDA values (in microsiemens μS), POR (x,y,z) (in meters), and inclination of the center of pressure (COP) motion (x,y,z) (in degrees). This open possibilities for in-depth analysis of behavior and performance, which is the focus of the presented task-based methodology.

We did highlight in the Table 4.3 the most relevant data contained in a synchronized dataframe from CREATTIVE3D dataset, and the variables used for our behavior and performance correlational analysis presented in the following sections. Here are some of the meaningful variables contained in the dataframe :

- *unixTimeStamp* : unix timestamp of the data entry,
- *position* : (x,y,z) position of the player in the 3D environment,

TABLE 4.3 – Data contained in a dataframe of synchronized behavioral and 3D contextual data, decomposed by tasks. Task objects row gives the name of 3D objects relevant for the current task. Performance metrics row list the variables used to identify the missteps during the task.

Task	# of logs	Task objects	Performance metrics
Get the key	1570	key	efficiency (unixTimestamp)
Open the door with the key	480	key, door	efficiency (unixTimestamp)
Get the trashbag	2243	garbage bag	efficiency (unixTimestamp)
Cross the road safely	2929	car, traffic light, traffic light button	efficiency (unixTimestamp) attention (lookedAtItemName) decision (carHonk, trafficLightColor)
Put the trashbag in the trashcan	642	garbage bag, trashcan	efficiency (unixTimestamp)

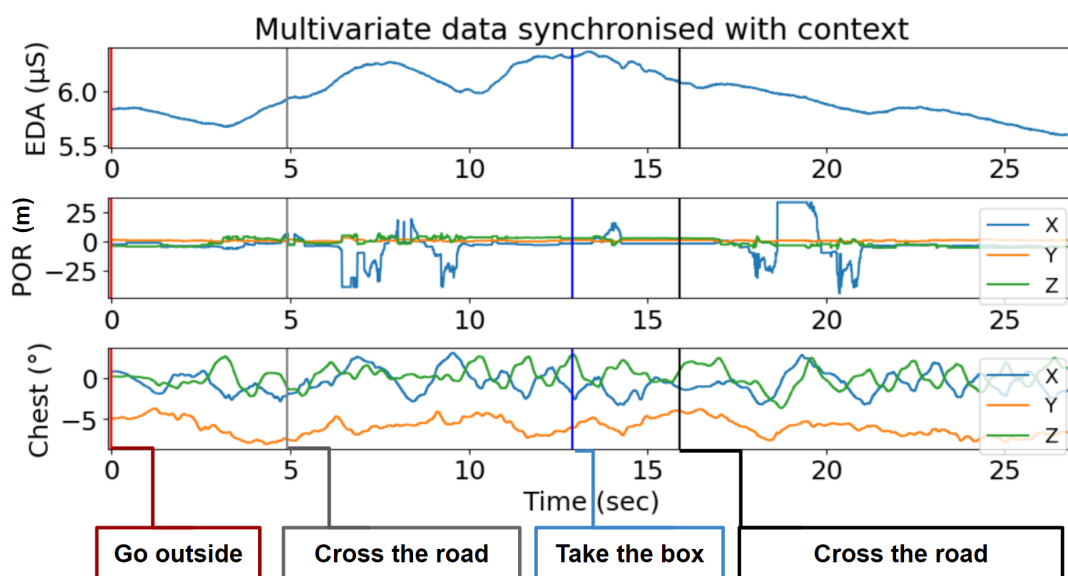


Figure 4.10 – An example of the raw multivariate data (i.e., skin conductance (EDA), point of regard positions with x, y, z axis (POR), chest rotation value) synchronised with the context for one user in a single scenario. During this scenario, the users carried out four tasks : (1) Go outside their house, (2) Cross the road safely, (3) Take the box, (4) Cross the road safely to come back home

- user state variables : *currentTask* (the name of the task the user is performing) and *item* (object the player is holding in VR),
- attention data : *lookedAtItemName* (the object the user is looking at) and *inViewItems* (all objects in the user's field of view),
- scene state variables : *trafficLightColor* (current color of the traffic light) and *honk* (boolean indicating if a car is currently honking),
- *EDA* : raw EDA value in microseimens μS ,
- point of regard : *PorXYZ* as a (x,y,z) position,
- motion capture sensor positions : *MotionPos* including (x,y,z) positions for each of the 17 motion sensors).

4.3.2 Task model

To tackle the challenges of characterizing experiential UX on a granular level for lifelike VR scenarios, we conceived a task-based methodology inspired from existing methodologies in human-machine interaction and robotics to finely observe and characterize user performance using multivariate data. This methodology can be applied widely to studies presenting two characteristics : (1) a clear temporal segmentation and (2) a set of synchronized behavioral and 3D contextual data. The former allows us to build a task model that characterizes the sequence and success criteria for each task, and more importantly, detect the missteps, i.e., when a task is incorrectly carried out. The latter then allows us to perform a fine-grained analysis of behavioral metrics in correlation with the 3D context at the moment a misstep occurs.

In this section we first present the task-model definition from high-level scenario to low-level task. Based on the task model, we then define the three baseline components that will be used to identify user performance missteps : efficiency, attention and decision. Finally, we present and justify the selected behavioural metrics derived from the multivariate data used to characterise the user behaviour in relation to performance missteps.

4.3.2.1 Task model conception

Our first step is to build our task-based approach inspired from task models (Santoro, 2005), describing a scenario as a graph of high to low-level tasks to perform, in a precise order. Each task is composed of precise set of criteria that classifies the execution of the task as a success, or with a misstep.

We constructed our task-model using Hamsters tool* (P. Palanque & Martinie, 2016). Each task has its own separate task-model with the following elements :

- task description from the high abstraction level – for example “open the door” – to the sequence of motor tasks composing this task – “find the key”, “walk to the key”, “press the trigger to grab the key”, “walk to the door”, “interact with the door” – in order to define precisely the ideal sequence of actions the user is expected to perform.
- precise definition of objects included in a task and the interactions expected with them (i.e., touch, grab, look at, place in, place on), which are the criteria to the success of a task.

Our scenarios feature 9 high-level tasks in the following sequence : (1) Open the door, (2) Take the garbage bag, (3) Go outside your home (4) Cross the road safely, (5) Put the garbage bag in the trashcan, (6) Take the box, (7) Cross the road safely, (8) Come back home, (9) Place the box on the table. Each scenario is a subset of tasks from this selection. For example, the simple interaction scenario with one car is composed of the task sequence 3 → 4 → 6 → 7 → 8 → 9. All scenarios are at least composed of one interactive task (e.g., take the box / garbage bag), and one road crossing task.

The Figure 4.11 illustrates a shorter, plausible scenario composed of only three high-level tasks, with a simplified view of the three task models for each high-level task. Each sub-figure highlights the expected interactions and the sequence to perform them in order to succeed at the task. In the “cross the road safely”, the user must, for example, cross the road while the traffic light is green, after checking that the car has stopped. The sequence however can be done in multiple orders creating multiple variation of a same scenario (i.e. State 1 : Look at green light, look at the stopped cars, cross the road ; State 2 : Look at red light, wait, look at green light, look at stopped cars, cross the road). The individual task models are made for all scenario variants. As such, we could for example look only at the road crossing task for one user in a single scenario individually, or we can aggregate all instances of the road crossing task across all users and scenarios in one analysis, and all instances of the opening door task for a different analysis.

The road crossing task is the longest and most complex task, with multivariate metrics to observe and analyse. It also incorporates all of the interactive mechanisms that appear in the other tasks. In the rest of the paper, we therefore use this task as the example to demonstrate our task model approach to UX analysis and performance evaluation.

*. <https://www.irit.fr/recherches/ICS/software/hamsters/>

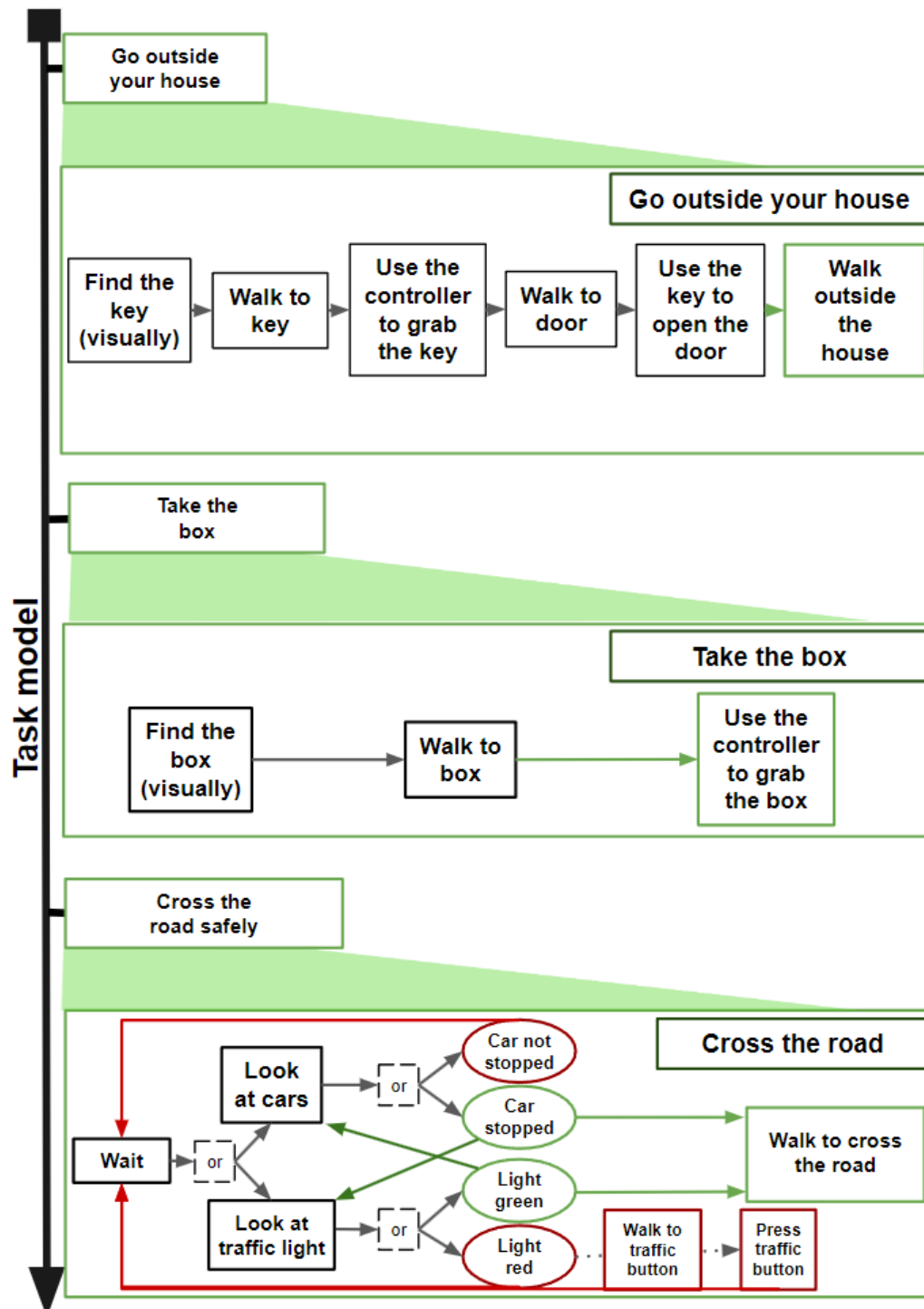


Figure 4.11 – A simplified task model for three high level tasks : (a) “Go outside your house”, (b) “Take the box”, (c) “Cross the road safely”, to properly achieve the task, the user is expected to perform the right interactions at the right time, and in the case of crossing the road, also look at traffic light and cars, and cross when the light is green and the cars are stopped. Rectangles indicate tasks and ovals indicate object states.

TABLE 4.4 – List of actions and the associated expected time to perform them in VR, inspired from GOMS methodology, to define task efficiency baseline value.

Action	Time
Scenario starting delay (S)	2.0 seconds
Audio instruction delay (I)	1.0 seconds
Grab an object (G)	2.6 seconds
Drop an object (D)	2.6 seconds
Press a button (P)	1.45 seconds
Red traffic light duration (T)	8.0 seconds
Walk physically (W)	(Distance / 0.9) seconds

4.3.2.2 Performance baselines

From this task model, we can already begin to observe the different components that could be used to quantify the success at performing the task (e.g., cross in time, cross during the green light, look left and right before crossing, don't get hit by a car). Out of this model, we can define the performances baselines values that an analyst would like to observe and analyse when a user is crossing the road.

We therefore defined three components needed for the successful completion of the “cross the road safely” task : (1) efficiency : the user must finish the task within a given amount of time, (2) attention : the user must pay attention (i.e., look at) certain elements during the task to know when to do an action, and (3) decision : the user must make correct choices in the sequence of executing actions.

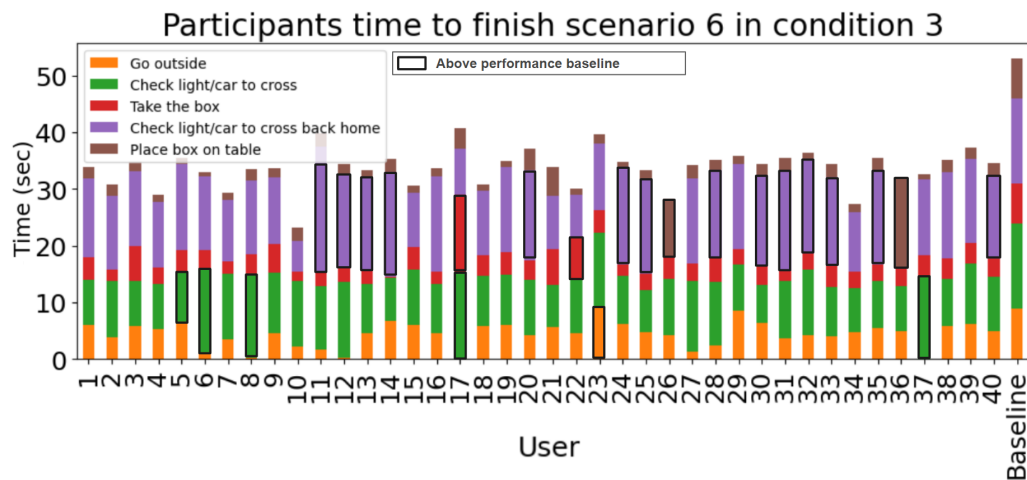


Figure 4.12 – Time taken by all 40 participants (sorted by study completion order) to finish the tasks of the scenario #6 (multiple interactions with two cars). The baseline column on the far right shows the baseline value for each individual task stacked. Tasks where the user surpasses this baseline are outlined in black.

First, we define the efficiency component, that is the time limit to complete the task. We calculated our baseline value on the existing human-computer interaction metric GOMS, which is a well known predictive model to quantify how much time it will take to perform a given task. Based on the extension to VR proposed by Guerra et al. (2022), we defined times to grab object, drop object and press the traffic light button. For the task of physical walking, we estimated the preferred walking speed of users in VR at 0.9m/s based on Wodarski et al. (2020), resulting in the list of actions presented in Table 4.4 to define the efficiency baseline value of a task.

TABLE 4.5 – The performance baseline values defined for the scenario six of the experiment. When crossing the road in this scenario, the user is expected to look at the car and the traffic light, cross only when the traffic light is green and the cars are stopped, in a total of 13.50 seconds.

Task	Efficiency	Attention	Decision
Go outside the house	$S+I+W=2+1+3.5=6.50s$	\	\
Cross the road safely	$I+W+T=1+4.5+8.0=13.50s$	Look at : Cars, Traffic light	Cross : Traffic light green, did not get honked
Take the box	$I+W+G = 1+3.4+2.6 =7.0s$	\	\
Cross the road back home	$I+W+T=1+4.5+8.0=13.50s$	Look at : Cars, Traffic light	Cross : Traffic light green, did not get honked
Place the box on the table	$I+W+D=1+3.4+2.6=7.0s$	\	\

When starting a scenario, we take into account that users will need a short adaptation time to look around at the environment, corresponding to the time S . For every task, users get a short audio instruction with task instructions, which corresponds to the time I . When users want to cross the road, the time they will have to wait for the light to go from red to green corresponds to time T .

Figure 4.12 allows us to visualise the time participants took to finish each task in scenario #6 (multiple interactions with two cars) under one experimental condition. The column *Baseline* in Figure 4.12 shows the efficiency component baseline defined for every task of the scenario. In this scenario, out of 200 tasks performed, we can see that 24 efficiency missteps occurred (i.e., the execution time surpassed the baseline value), highlighted by the black frames.

This definition process is also done for the two other performance components : attention and decision, which are a more straightforward. The attention baseline defines objects in the scene that the user must look at. The decision baseline designates certain states under which objects must be when the user executes a task. The right-most column in Table 4.3, summarises the the data entries in the dataframe that are relevant to each performance component for each task. Using these data entries, Table 4.5 then lists in detail the three performance components for each task the same scenario in Figure 4.12. The efficiency baseline of each task is defined as the sum of the actions comprising the task listed in Table 4.4. For example, during the second task “Cross the road safely”, to fulfill the efficiency component, users must finish the task within 13.50 seconds (i.e., the user can wait 1 second following the audio instruction to start the task, wait up to 8 seconds for the traffic light to turn to green, and finally walk a distance of 5 meters at a speed of 0.9m/s, making a total expected duration of up to 13.50 seconds to complete the task); to fulfill the attention component, users must look at the cars and traffic light prior to going on the road; finally, they must also cross road while the traffic light is green without being honked by a car that wants to pass during their own green light to fulfill the decision component.

After defining these performance components, we can then group the users in each task into those who fulfilled the performance baselines and successfully executed the task under the baseline value, and those who did not, thus incurring a performance misstep. We can then correlate user behavioral metrics to observed performance missteps, allowing a more fine-grained analysis of the behavior in relation to the performance.

In order to carry out our task-based analysis, the field *currentTask* in the synchronized dataframe is used to partition a scenario file in task groups corresponding to the rows in Table 4.3. With this partitioning, we apply our baseline values the relevant variables, to characterize if a user did a misstep for the given task. Again using the example of the “Cross the road safely” task, the missteps for the three performance components are well-defined :

- efficiency : the *unixTimeStamp* value gap between the beginning and the ending of the task is above the baseline value,
- decision : the user crossed the road while *trafficLightColor* value was at “red”, or during the road crossing task, the *honk* value has been equal to true at least once,
- attention : the user crossed the road without *lookedAtItemName* value being equal to “car” at least once during the task.

4.3.2.3 Behavioral metrics

The last step in our approach is to select the metrics with which we wish to characterize the user performance. It is important to note that multiple possibilities can be imagined from such a dataset for each data type. For example, the emotion data is composed of both electrodermal activity and heart rate information. From motion capture data, gait characteristics, stability, limb coordination etc. can be derived. In this work we limit our analysis to only a subset of these metrics to illustrate our approach, using the following principles to help us in our choice :

1. Only one metric is chosen for each data type (i.e. motion, emotion and attention), since the goal is to exemplify our task-based methodology along with its flexibility to be applied to diverse type of data, and not present an overall analysis of all possible metrics.
2. Avoid metrics that are directly related to the person's absolute physical characteristics, such as height or step size, which can be correlated to additional factors such as gender, which could merit new sets of research questions into other aspects of user experience analysis currently outside of the scope of this work.
3. Prioritize metrics that have already been of interest for existing studies in VR, and are also relevant to the performance baselines in our task model, which highlights the added value of our task-based approach for fine-grained analysis.

Emotion we select skin conductance (or electrodermal activity, EDA) which shows the level of arousal of the user, indirectly indicates the level of stress the user is experiencing in the long term and the intensity of their timely physiological responses to stimuli, represented by skin conductance response (SCR) peaks. Skin conductance (EDA) is often used in the literature (Xue, El Ali, et al., 2021) to evaluate the arousal of the user considering multiple context with more or less stress-inducing elements. This metric is particularly interesting in the context of this dataset which includes scenarios with more or less stress inducing elements that could lead to missteps on user performance and observable variation on user behavior.

Attention we choose gaze fixation duration (GFD), which has been used to observe visual recognition processes in Chan et al. work (2023) or gaze behavior for sports in Klostermann et al. (2020), who have identified a potential link between longer duration of gaze fixations for professional players compared to amateurs. We want to investigate notably if GFD is potentially a metric that could help characterize attention and decision performance missteps, such as being correlated to shorter average fixations.

Motion we choose center of pressure inclination (COPI), adopted in existing studies to measure age and height effect on body balance through center of mass (COM) and center of pressure (COP) inclination (S.-C. Huang et al., 2008), or to measure body sway when viewing moving visuals in VR headsets (Kuno et al., 1999). It is an important metric to measure body muscle stress (Zander et al., 2001), respiratory mechanics (Mezidi & Guérin, 2019), and body balance (Eklund, 1972). The stability of users is a major concern for VR experiences that involve physical walking, making this metric interesting in the context of mobility in VR for our selected task, where there is physical walking involved to cross a road. We therefore investigate the link COPI could have with performance.

Setting out from the synchronized dataframe, we process each data type to calculate the above-mentioned metrics. The skin conductance data (EDA) is analyzed with the Python toolkit Neurokit (Makowski et al., 2021), as presented in the previous Section 4.1 where the phasic component is calculated as the first derivative of the normalized EDA, allowing us to observe short term variations of physiological response in relation to the scene context. For gaze fixation, we used the I-VT algorithm (Salvucci & Goldberg, 2000) to obtain the list of fixations and their duration. Center of pressure inclination is computed on raw motion data records, using chest and ankle segments to compute the body angle, with a higher value when the body is more inclined and zero meaning the body is 90°vertical.

This three step process : establishing the task model, defining performance components and baseline values, and finally selecting relevant behavior metrics comprises our methodology which allows us to (1)

TABLE 4.6 – The average values of each metric for tasks with misstep TM_c on a performance component c and those without a misstep. We calculate the p-value significance of each behavioral metric in combination with the performance component. The values in bold indicate significance $p < 0.05$. We can for example observe a strong significance in COPI when comparing the groups with and without TM_e .

Metrics	Efficiency			Attention			Decision		
	COPI	EDA	GFD	COPI	EDA	GFD	COPI	EDA	GFD
TM_c median	3.72°	4.56e-4	153.23ms	3.70°	5.37e-4	153.87ms	3.61°	4.93e-4	153.91
Non TM_c median	3.28°	4.52e-4	148.11ms	3.26°	4.55e-4	148.78ms	3.24°	4.55e-4	147.94
p-value	4e-8	0.582	0.084	6e-14	4e-7	0.010	2e-4	0.159	0.011

finely segment tasks into identifiable actions, (2) define which components of performances are evaluated and how they are quantified for each tasks done by the user, in order to classify the performance result (i.e., success or misstep), and (3) analyze correlations to behavioral metrics.

4.3.3 Results

In the previous section, we presented the multivariate data synchronization with context, the task model based on the multivariate data, establishing baselines to define the components on which performance is evaluated, and selection of metrics for characterizing user behavior. Here we take a deep dive into the analysis of user behavior using multivariate data based on performances, and further investigate on the possibility of identifying profiles of users whose performances are similar on certain observable metrics. We continue using the road crossing task and the three defined performance components and baselines : (1) efficiency : the time to finish the task, (2) attention : the elements looked at during the task and (3) decision : the choices made by the user to finish the task.

The dataset contains the record of 40 participants who each completed 24 scenarios, for a total of 960 scenarios with road crossing tasks to observe. Using the performance component baselines, we can identify the missteps for each performance component. We then analyses the behavioral metrics selected : electrodermal activity (EDA), center of pressure inclination (COPI), and gaze fixation duration (GFD) in order to characterize the behaviors that are in relation to performance missteps.

The following notation is used :

- TM_a , TM_e , and TM_d denote a task with the occurrence of an attention, efficiency, and decision misstep respectively
- the metrics are abbreviated as COPI, EDA, and GFD for center of pressure inclination, electrodermal activity, and gaze fixation duration respectively

4.3.3.1 Behavior at a scenario level

Out of the 960 scenarios with road crossing done by the 40 participants, 630 missteps of either attention, decision or efficiency were observed. The majority of the missteps were TM_a representing 47.9% (302) of all the misstep occurrences, then 27.1% (172) for TM_e , and 24.9% (156) of TM_d . The average missteps for users was 7.6 TM_a , 4.3 TM_e , and 3.9 TM_d .

For the first part of the analysis, we would like to observe potential trends that could signify a link between one performance component and one behavior metric. We take the road crossing task data from the 24 scenarios from all users, each classified according to the three tasks performance components, resulting in a dataframe where each row represents the user, scenario, success or misstep associated to each performance component, and the average value of each behavioral metric. To identify if the distribution for each metric is significant between the TM_c and non- TM_c groups for a given component c , we then compute the null-hypothesis significance testing with p-values (Aseeri & Interrante, 2021 ; Delrieu et al., 2020). As shown in Table 4.6, we can find significance ($p < 0.05$) for TM_a +GFD and TM_d +GFD, and strong significance ($p < 0.001$), for the TM_e +COPI, TM_a +COPI, TM_a +EDA, and TM_d +COPI. The median values of each

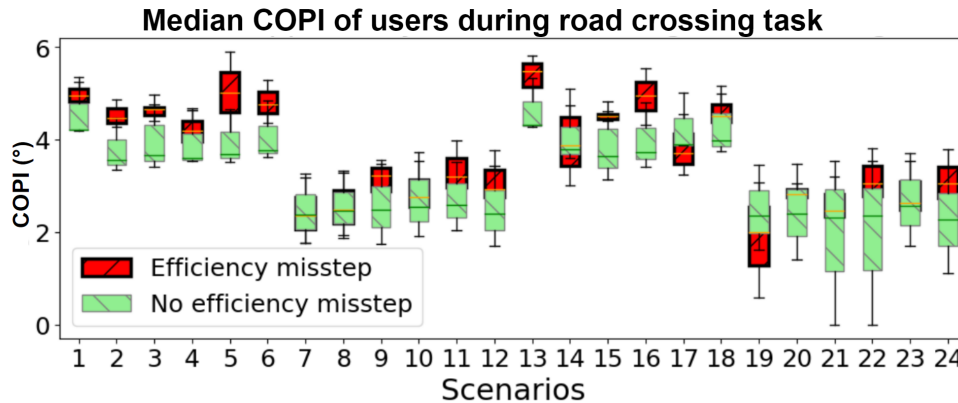


Figure 4.13 – Median COPI of all users for each scenario of the experiment based on the efficiency component of performance. The red boxes represent the COPI values for TM_e , and the green boxes represent the COPI values for non TM_e .

TABLE 4.7 – Individual users whose values for a given behavioral metric are significantly different ($p < 0.05$) between tasks with and without missteps. uID corresponds to the ID of the user, $Misstep$ corresponds to the percentage of tasks that have a misstep for a performance component, and Var corresponds to the variation (with positive/negative sign) observed median of the metric between TM_c and non TM_c for the given performance component c

	Efficiency misstep				Attention misstep				Decision misstep			
	uID	p-value	Misstep	Var	uID	p-value	Misstep	Var	uID	p-value	Misstep	Var
COPI	851	0.034	54%	-29%	321	0.020	29%	+56%	666	0.001	13%	-43%
	963	0.009	42%	-22%	511	0.010	46%	-14%	877	0.004	25%	-40%
					495	0.031	29%	-15%	828	0.026	38%	-19%
					109	0.029	46%	-18%				
EDA	981	0.002	21%	+55%	748	0.037	62%	-18%	666	0.001	13%	-71%
	361	0.006	25%	+47%	091	0.047	50%	-35%	981	0.016	13%	-46%
	658	0.009	17%	+52%	213	0.004	21%	+71%	361	0.021	25%	-50%
GFD	361	0.032	25%	+27%	666	0.020	42%	+28%	446	0.026	29%	+73%
	940	0.029	17%	-29%	682	0.027	56%	+17%	748	0.017	21%	+33%
									446	0.033	21%	-17%

metric are also shown for the different group and component combinations. We can see an increase of COPI, EDA, and GFD for the different missteps compared to the scenarios without missteps. If we look in detail at Figure 4.13, we can visualize the trend of the average COPI of three users. The group of tasks with performance missteps has a globally higher median COPI than those without, the exceptions being scenarios 14, 17 and 19.

4.3.3.2 Behavior on user level

In order to observe on a more fine-grained level if a user presents specific behaviors in relation to their performance, we conducted the same type of analysis for users individually, looking at each road crossing task during the experiment, grouping their scenario data based on success or misstep in a performance component. On these groups of behavioral metrics classified by performance, we computed the p-value in order to identify those users whose behavior is significantly different for a given metric based on their performance. For this analysis, we kept only users with a p-value below 0.05 and at least 10% of tasks with occurrences of missteps in order to have a sufficient amount of data to compare between the performances groups. The users included this analysis are shown in Table 4.7.

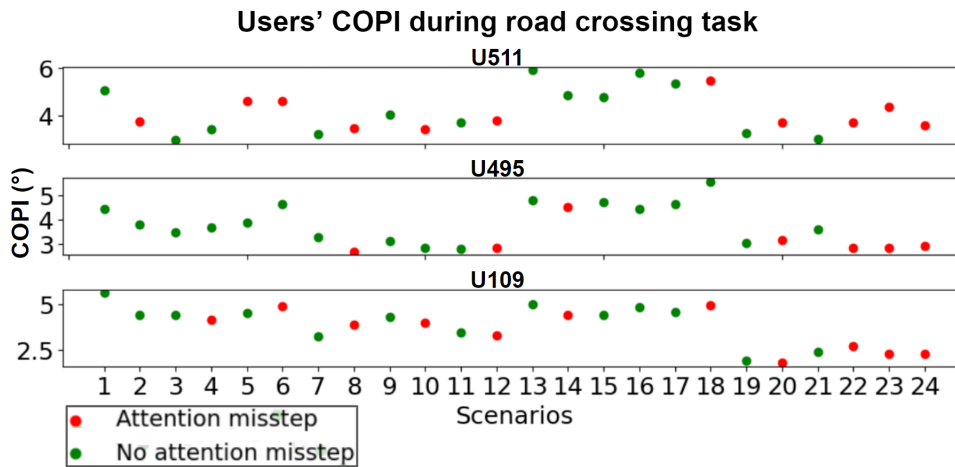


Figure 4.14 – Visualization of the median COPI value of each scenarios with road crossing for the users U511 (top), U495 (middle), U109 (bottom) with significant negative COPI variation in TM_a tasks. Red dots represent the median COPI value during road crossing task for a scenario with attention misstep, green dots represent the median COPI values during road crossing task for a scenario without attention misstep.

These results outline unique behavioral profiles. We found at least two people per combination of performance-metric, and can observe that there is rarely one consistent trend of behavior : significantly higher for one user but the exact opposite for another. Four interesting exceptions exist : a higher EDA for TM_e , a higher GFD for TM_a , and a lower COPI for both TM_e and TM_d . First, the revelation of the relation between the gaze metric (GFD) with attention missteps makes logical sense, as the user does not switch their gaze as frequently to take into account various pieces of information. The relation between emotional metric (EDA) could also be an indicator of the level of stress the user is experiencing. Finally, a lower COPI indicates that the person is more upright, which can be a sign of less engagement or walking slower. The observations can then facilitate the conception of hypotheses to further validate these relations between the performance and behavior. We can visualize more in Figure 4.14 the COPI of users U511, U495 and U109 who have a significant difference in value when committing a TM_a , with a decreasing tendency of their COPI of -14% to -18% compared to non TM_a .

4.3.4 Discussion

Many analyses of immersive experiences using behavioral metrics exist in the literature, showing correlations between diverse parameters such as heart rate and eye behavior related to a precise context (B. J. Li et al., 2017; Y. Lang et al., 2018; Seinfeld et al., 2020). Nonetheless, these works rarely conduct fine-grained analysis of user behavior in relation to scenarios with continuous lifelike training scenarios composed of multiple sub-tasks to pinpoint the elements at a specific time of the study that trigger user behavior. The experimental conditions in such studies are also often difficult to replicate (e.g., framework reproducibility, differences in setup, data availability, population bias), limiting the possible conclusions to the ones reported in the work itself.

While many methodologies exist in HCI for the analysis of user experience, few of them have been designed and applied to embodied VR experiences. Such methodologies can enable granular observations of user performance and behavior in relation to the continuous lifelike VR scenario context. To support designers in the refinement of experiences or the guidance of users, being able to identify the metrics that are most efficient at characterizing user behavior for a given type of task is very valuable. It facilitates the ability to observe with accuracy when and why a user misstep or system limitation happens.

We have presented here a task-based methodology to evaluate and characterize experiential UX, taking inspiration from existing work in human-machine interactions and robotic task modeling, which have not

yet been fully explored and used for analysis of VR experiences. Our methodology proposes new ways to evaluate the UX by looking at the fine-grained behavior synchronized with context on three performance components, and finally allowing the characterization of specific user profiles. By applying the methodology on a selected parameter of performance (i.e., attention, efficiency, decision) with suitably selected behavioral metrics, we can observe whether a metric could potentially be interesting for a better characterization of the UX in a given context, task, or type of performance measure.

While our analysis and results have mainly been focused on a deep dive into the performance of a single road crossing task, it has already allowed us to observe some user behavior profiles, and highlight the metrics are potentially more effective for the characterization of the experience, such as gaze for tasks requiring visual attention, and emotion for task efficiency. We would also like to outline the inclusion of center of pressure inclination (COPI), which is important for tasks that require spatial displacement within a limited amount of time. We believe this is a novel and important contribution for VR that can benefit applications such as training or rehabilitation applications, where the understanding of a user's difficulties and needs when working with a VR system is crucial to improve the both global experience and provide personalisation.

One major strength of the methodology is its flexibility, as it is applicable to various types of studies, allowing the definition of performance parameters according to the task, and the analysis of corresponding behavioral data. In this work, we demonstrated this analysis on a pre-existing dataset, applying new baselines on performance that were not part of the original study design.

4.4 Conclusion

The first contribution of this chapter is the creation of publicly available immersive datasets. These works contribute to the thesis by presenting the datasets PEM360 and CREATTIVE360, enabling the granular analysis of user behavior and embodied interactions in an immersive context, providing answer to the RQ2A.

The second contribution of this chapter is the conception of our task-based methodology to analyze embodied experiences using the datasets created in response to RQ2A. This methodology provides an answer to RQ2B by enabling a better understanding of embodied experiences through the in-depth analysis of embodied interactions by looking at synchronized behavioral and contextual data based on scenario tasks. This work also provides an answer to the RQ2C, creating a link between the user's physical behavior and the task's performance for the creation of user profiles. We believe this work opens new ways to understand the user embodied experience, allowing us to pinpoint a link between user behavior and user difficulty. This represents major information to support designers in the creation of personalized guidance and scenarios according to users' needs.

Contributions The publications and datasets presented through this chapter :



F. Robert, H.-Y. Wu, L. Sassatelli, and M. Winckler. “Task-based methodology to characterise immersive user experience with multivariate data”. IEEE VR 2024, Orlando, Florida. [hal-04446066](#)

H.-Y. Wu, **F. Robert**, F. Franco Gallo, K. Pirkovets, C. Quéré, J. Delachambre, S. Ramanoël, A. Gros, M. Winckler, L. Sassatelli, M. Hayotte, A. Menin, P. Kornprobst. “Exploring, walking, and interacting in virtual reality with simulated low vision : a living contextual dataset”. (Under review for Nature Scientific Data). [hal-04429351](#)

Q. Guimard, **F. Robert**, C. Bauce, A. Ducreux, H.-Y. Wu, L. Sassatelli, M. Winckler, A. Gros. “PEM360 : a dataset of 360° videos with continuous physiological measurements, subjective emotional ratings and motion traces”. MMSys 2022, Athlone, Ireland. [hal-03710323](#)

Q. Guimard, **F. Robert**, C. Bauce, A. Ducreux, H.-Y. Wu, L. Sassatelli, M. Winckler, A. Gros. “On the link between emotion, attention and content in virtual immersive environments”. ICIP 2022, Bordeaux, France. [hal-03825059](#)



CREATTIVE3D dataset : <https://zenodo.org/records/10406560>

PEM360 dataset : <https://gitlab.com/PEM360/PEM360/>

Investigating inter user-designer intersubjectivity communication

5.1 Overview

The third subject investigated in this thesis concerns the conception of methodologies to identify and understand the communication issues happening between designer, system, and user, during the design and usage phases of an embodied experience. The promise of richer and more engaging interactions raises the risk of miscommunication between the experience sought by the designer and the experience lived and interpreted by the user. Based on Paul Dourish's theory of embodiment, the **intersubjectivity** components result in gaps in understanding from the designer to the user when communicating the task goals and constraints of the 3D environment through the system. To bridge this gap between users and designers in the design and use of embodied experiences, we need means to identify the issues causing gaps, what their impact is, and how to prevent them.

This chapter investigates the RQ3 – *How can we specify the intersubjectivity gap between design intention and actual usage in embodied interactive experiences, and how can we use this specification to provide feedback to the designers?* – through two sub-questions :

- **RQ3A** : How to detect and classify the different issues of communication happening during the design and the usage ?
- **RQ3B** : How to prevent or avoid the potential issues to improve the communication between the actors ?

The challenge investigated during this chapter, related to the **RQ3A** and **RQ3B**, concerns the conception of a methodology that can support designers of embodied VR experiences in the comprehension and the improvement of the designer-user communication mediated by VR systems.

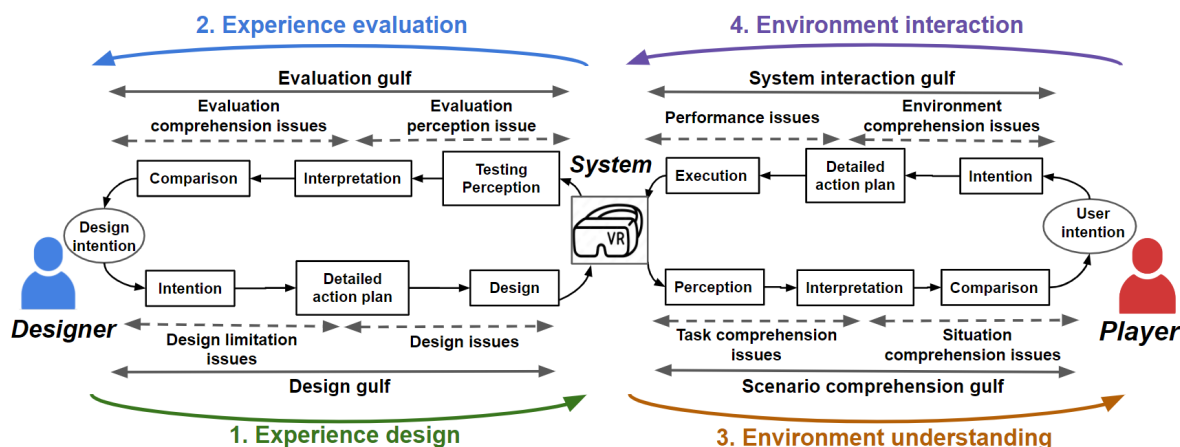


Figure 5.1 – Our model was inspired by Norman to characterize the designer-system-player communication gulfs, allowing the definition of a taxonomy of eight issues sources of gulfs happening during the design, usage, and interpretation phases.

5.2 The triangle of misunderstanding in interactive virtual narratives : gulfs between system, designers and players

Existing works focusing on the user-system-designer communication ultimately aim at reducing the communication gulf between the actors. Bad communication on the scenario's narrative or on how to interact with 3D objects – such as coming from personal biases, system limitations, forgetfulness, or lack of guidelines – can impact the player's immersion or the enjoyment of the experience. On the other hand, an incomplete comprehension of the player's lived experience can impact the designer's knowledge on how to improve an experience. Widely relevant HCI works exist for understanding communication issues, but were designed for traditional media, based on 2D interfaces and affordances. They do not address the specificities of VR applications : the immersion, the embodied interaction possibilities, the physical fatigue etc.

To observe and characterize the nature of the communication gulfs existing between the designer's intention and the player's lived VR narrative experience, we conceived a communication model that supports the identification and mapping of eight classes of issues causing miscommunication or misinterpretation in the design-interpretation process. Our model is inspired by the seven stages of action model built by Norman (2013), which refers to the way a user interacts with a product or a system in seven steps, separated into two phases : the execution phase, during which the user will form an intention, construct a plan, and execute it, and the evaluation phase, during which the user will observe the result of their action, make an interpretation, and compare the consequences with their expectations. As a result, two key gulfs emerge from the usage of a system : the execution gulf (i.e., the gap between the initial intention and the performed actions) and the evaluation gulf (i.e., the gap between what the system shows and what the user expected).

We presents here three contributions :

- A communication model to investigate the designer-player communication gulfs, which highlights the different steps composing the creation process between the designer and the system, and the usage process between the player and the system.
- A qualitative method in the form of a structured interview, to observe and collect communication issues happening during the usage and design of VR narrative experiences. This interview was applied during a two-phase VR study, allowing the collection of 127 design and usage communication issues.
- A taxonomy of communication issues to supports the identification and mapping of all encountered communication issues among eight classes of issues causing miscommunication or misinterpretation, to analyze their individual impact on the proper narrative interpretation of the scenario.

By revisiting Norman’s model in the context of VR narrative experiences, including both the design and the usage process, we can support the identification, and eventually the reduction, of designer-player communication gulfs in VR, by proposing means to treat each class of communication issues.

In the following chapter, we describe in section 5.2.2 our communication model and the structured interview built based on this model to collect and classify issues. Then in section 5.2.4 we present the study design and the technical setup. In section 5.2.5 we present our analysis on the correlation between the issues, alongside some participants post-study insight.

5.2.1 VR communication model

We present our model of VR communication shown in Figure 5.1. The model was inspired by Norman’s seven stages of action, well-known in the HCI research community, to represent the communication between a *player* and a *system*. In our case, we need to add a third actor : the *designer* who uses the system to create the experience for the player, who will then live the experience through the system. Both the designer and player interact with the system following Norman’s seven stages of action : develop and act on their intention to interact with the system, and compare the changes against the initial intention. The system is thus a medium through which the designer communicates with the player, from which also emerge designer-player gulfs, notably between the initial design intention and the final player interpretation. For these reasons, our model is the concatenation of two models of the seven stages of action, with the left side representing the design process and the right side representing the usage process.

Our model specifies four designer-system-player gulfs and eight classes of issues that are sources of these gulfs. The first four issues occur on the designer-system side of Figure 5.1, potentially creating a design gulf (i.e., the gap between what the designer wanted to create, and what they actually create) and an evaluation gulf (i.e., the gap between what the system shows and what the designer understands). During this phase, the designers define their design intention and build a VR scenario to communicate the intention through the system. They then test their scenario, reiterating the design process until they can validate that it corresponds to their initial intention. The four classes of issues are :

(1) Design limitations issues : limitations on the system to implement a desired scenario, potentially limiting the designer’s creativity. For example : a designer who wants to build a scenario with fireworks would rely on the system having an adapted particle system, the lack of which would force the designer to change or suppress this idea.

(2) Design issues : actions that are present in the designer’s detailed plan, but not performed, or performed incorrectly. For example : the designer forgets to create a task, or associates the wrong interaction to an object resulting in a mismatch between the intention and final scenario.

(3) Evaluation perception issues : gaps identified by the designer between what they wanted to create during the design process, and what they experienced while testing the scenario. For example, a designer expects an object to be clearly visible but it is unexpectedly occluded due to the camera perspective.

(4) Evaluation comprehension issues : issues that the designer didn’t notice during the testing. For example if the designer forgot to create a task and didn’t realize it during the test.

The second set of four issues occur on the system-player side in Figure 5.1, creating the scenario comprehension gulf (i.e., the gap between what the system communicates and what the player understands) and the system interaction gulf (i.e., the gap between the expected and the actual interactions of the player with the system). The four classes of issues in this phases are :

(5) Task comprehension issues : The player misunderstands the task to perform when communicated by the system. For example the player takes the wrong object or performs the wrong interaction.

(6) Situation comprehension issues : the lack of comprehension on the player’s side of the purpose of a task. For example : not understanding the goal of the task and/or the higher message of the scenario. As a result, the player may believe the task serves no purpose in the story.

(7) Environment comprehension issues : lack of comprehension by the player due to the manner in which the system communicates (e.g., visuals, audio, text). This can occur for example when the player confuses the target object of the task with similar objects in the scene.

(8) Performance issues : issues in executing the task by the player. For example : a player is unable to find the target object or fumbles with the controllers to perform a desired action.

This specification of issues in eight classes opens up the possibility of investigating in depth the link between issues, gulfs, and narrative interpretation. To detect and classify communication issues among these categories, we built a qualitative methodology presented in the next section.

TABLE 5.1 – The structure interview built for the study, based on our model, to collect the initial intention, the design issues, test issues, and interpretation issues. These questions allows us to characterize the gulf between the designer initial intention and the player narration interpretation, by collecting the though process on both sides and comparing them.

Issues	Questions
Intention evaluation	(1) What scenario do you intend to create ? (2) What reaction, emotion, or experience do you want to generate through this scenario ? (3) What do you want players to experience through your scenario ?
Designer - System issues	Before the design : (3) What do you want players to experience through your scenario ? (4) What do you plan to do in the scenario to generate these reactions, emotions, or experiences ? (5) What sequence of tasks would you like to create to achieve this scenario ? After the design : (6) What scenario did you create ? (7.1) Are there differences between your scenario and your initial intention ? (7.2) What are the reasons for the differences between your scenario and your initial intention ? After the test : (8) What did you observe during your test ? (9) What did you understand about the scenario during the test ? (10) Do you notice any differences between what you tested and what you wanted to create ? (11) If you could change anything, what would you change ?
Player - System issues	To the scenario player : (12) In order, what actions did you perform to achieve the scenario ? (13) Did the observed consequences of your actions correspond to your expectations ? (14) How successful were you in accomplishing what you were asked to do ? (15) What did you understand from the scenario ? To the scenario designer : (16) Was the scenario played as you expected ? (17) The lived experience correspond to your expectations ? (18) In your opinion, what experience did the player live ?
Narrative interpretation	To the scenario player : (19) What did you understand from the scenario ? (20) In your opinion, what did the designer want to communicate through this scenario ? (21) What is the designer's intention : What points correspond or not to your experience ? To the scenario designer : (22) In your opinion, what experience did the player live ? (23) What did the player understand : Does that correspond to what you expected ?

5.2.2 Methodology

We apply the model to a user study in VR where participants had to play both the roles of a designer and a player, paired with a structured interview based on our model. The following sections first detail the design of the structured interview, then describe the player study.

5.2.3 Structured interview

The structured interview is composed of a set of questions for each phase of the design-usage process specified in our model, and was posed after each phase by the experimenter and recorded in a questionnaire. The list of questions are shown in Table 5.1. The reasoning for question in each of the eight categories were :

Design limitation issue : a designer’s conscious gap between the initial design intention (Q1-3) and the scenario they were able to implement using the system (Q6,7).

Design issue : a designer’s unconscious gaps between what they believe they created (Q1-7), and what they actually conceived (Q16-18), identified during the interview with the designer after player testing.

Evaluation perception issue : a designer’s conscious gaps between what they expected to create and communicate (Q5,6,7) and what they observed while testing their own scenario (Q8-11), identified during the designer evaluation phase.

Evaluation comprehension issue : a designer’s unconscious gaps between what they believe their scenario communicates (Q6,8-11), and what the players understood (Q12-18), identified by discussing the elements players didn’t understand with the designer.

Task comprehension issue : a player’s unconscious gap between the action the system asks them to perform (Q5,6), and the action the player understood (Q12-15).

Situation comprehension issue : a player’s unconscious gap between the meaning of the task presented by the system (Q5,6) and the comprehension of the purpose of this task by the player (Q12-15).

Environment comprehension issue : a player’s unconscious gap between the interaction expected by the system to validate a task (Q5,6) and what the player actually performed (Q12-14).

Performance issue : a player’s gap between the expected performance of the designer (Q5,6), and the observed performance of the player during the scenario (Q16,17).

Finally, we included a category on **narrative interpretation**, corresponding to the high-level gap between the player’s understanding of the meaning of the scenario they tested (Q15,19-21) and the scenario the designer actually wanted to communicate (Q1-3,22,23). While all previous presented issues aim at characterizing the eight designer-system and system-player gulfs, the narrative interpretation falls outside of this model. This enables the investigation of the potential correlation between the eight issues encountered and the quality of the narrative interpretation observed.

5.2.4 User study

We present here a qualitative study protocol performed to observe and collect information regarding the eight classes of issues. The study objective was to investigate (1) the impact of an issue on the creation of other issues in a later process, and (2) the impact of a given issue on the final comprehension of the experience by the user. The findings on these objectives are presented in Section 5.2.5

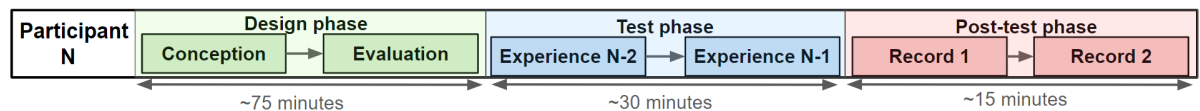


Figure 5.2 – The study protocol consists of three stages : (1) design phase, during which participants create and evaluate their own experience ; (2) test phase, during which each participants test two previously created experiences ; (3) and the post-test phase during which each participant watch record of two persons playing the created scenario. The entire study lasted roughly two hours.

We recruited six participants for the pilot studies (5 women, 1 man) and ten participants for the main study (5 women, 5 men) through 5 university and laboratory mailing lists. Participants needed to have normal or corrected-to-normal vision. Among the ten participants recruited for the main study, five had no prior experience with 3D scenario design, and five had minor or regular experience with it. The study was approved by the instute ethics committee. Upon recruitment, the participants were sent a message with their booked time slot as well as the informed consent form. The study lasted up to two hours. A compensation of twenty euros was given in the form of a check at the end of the post-test phase. On arrival, participants were first invited to sign the informed consent form and answer a short pre-experience survey. Participants were informed of the risks of nausea and motion sickness during the test phase and were encouraged to ask for a pause or request the end of the study if desired, without impact on their compensation. A time slot was defined with the participant at the end of the study to perform the post-test phase of the study.

In this two hours study, participants played the roles of both the designer and the player. Through structured interviews, they explained as designers their intentions and implementation plans, and as players their perceptions and interpretation of the scenario. A series of questions were used to determine if there were issues, which were recorded and classified post-test. The study is composed of three phases : (1) **the design phase** during which participants, as designers, create a VR scenario ; (2) **the test phase** during which participants, as players, test two experiences created by other designers ; and (3) **the post-test phase** at a later date, during which designers watch the recordings of two players who played their scenario. Two additional participants were recruited for only phase two, to ensure that all scenarios had two players. The questions for each phase are shown in Table 5.1 to identify for each phase the issues and their type, as well as the initial designer intention and the narrative interpretation of the scenario by the player.

In Phase 1, the designer phase, designers created their VR experience on a laptop using the GUsT-3D* open toolkit which is built on Unity. This toolkit allows the definition of a vocabulary to name objects and locations, and to designate a series of interactive user tasks such as finding and placing objects in specific locations, going somewhere, or using an object to interact with another. Text, audio, and visual hints can also be added to the tasks. The toolkit has the advantage of allowing the design of interactive VR scenarios without the need for prior training in Unity nor programming. For the VR setup used in phase 2, we used an HTC Vive Pro Eye with a wireless module, two HTC Vive controllers to interact with the environment, four HTC base stations to track a 5 by 5 meter space. This VR setup allows participants to physically walk and interact in the virtual environment freely without cables restraining their movement. Logs are generated from the user's position and interactions at 10Hz.

Design phase : the participant followed a 10 to 15 minutes tutorial on how to build a 3D experience using GUsT-3D comprised of low level tasks (e.g., take an object, go somewhere, place an object in or on something). Designers were then asked to create a scenario composed of at most 10 tasks, a task being the action to "Take an object", "Place an object", "Go somewhere", "Interact with an object" only. Each designer is assigned one of two base 3D environments of 5 by 5 meters to design their scenario : (1) an interior environment (Figure 5.3 top) with an office desk, a shelf, and some objects around the room. Designers assigned to this environment are given the constraints to use the objects "Golf ball", "Golf club", and "An object fallen on the ground" in their scenario, and (2) an exterior environment (Figure 5.3 bottom), comprised of a small entry hall with some furniture, and a crosswalk behind a door leading to the other side of the road, with some additional props. Designers assigned to this environment are given the constraint to use the objects "Key", "Trashcan", and "A sweet object". When played, the tasks appear during the scenario in the form of a text prompt in the field of view of the player, and is automatically updated as tasks are completed. Designers were allowed to add at most two hints, appearing as text below the task prompt, which can either guide the player or give additional narrative information. The designers were allowed to rename the objects and move them in the scene to fit the scenario they want to communicate. As the designers build their scenario, they can test and modify it as much as they wanted on the laptop. Once this phase was finished, designers test their scenario in VR, and but were not allowed to make further modifications. Questions were asked during these phases on the initial design intention, and characterize the issues related to each phase, as shown in Table 5.1 rows "Intention evaluation" and "Designer - System issues". This phase took a maximum of 1 hour and 15 minutes : 1 hour for the experience design and 15 minutes to test the scenario in VR, including the structured interview questions.

Usage phase : during the second phase of the study, participants take the player role and experience two scenarios created by the two preceding designers, one interior scene and one exterior. For example Designer 5 is player 2 for Designer 3 (which we denote as D3U2) and player 1 for Designer 4 (D4U1). The player then responded to questions regarding the issues encountered in this phase, and the interpreted meaning of the scenario in rows "Player - System issues" and "Narrative interpretation" of Table 5.1 respectively. This phase takes up to 30 minutes.

Post-test phase : the third phase was programmed at a later date with the designer after two players played their scenario. During this phase, a recording from the viewpoint of the two players is shown to the designer. Following each recording, questions shown in the Table 5.1 under the rows "Player - System issues"

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

and “Narrative interpretation” were posed to determine if the scenario was interpreted and lived closely to what the designer intended or not. This phase takes up to 15 minutes to watch the two player videos and respond to the questions.

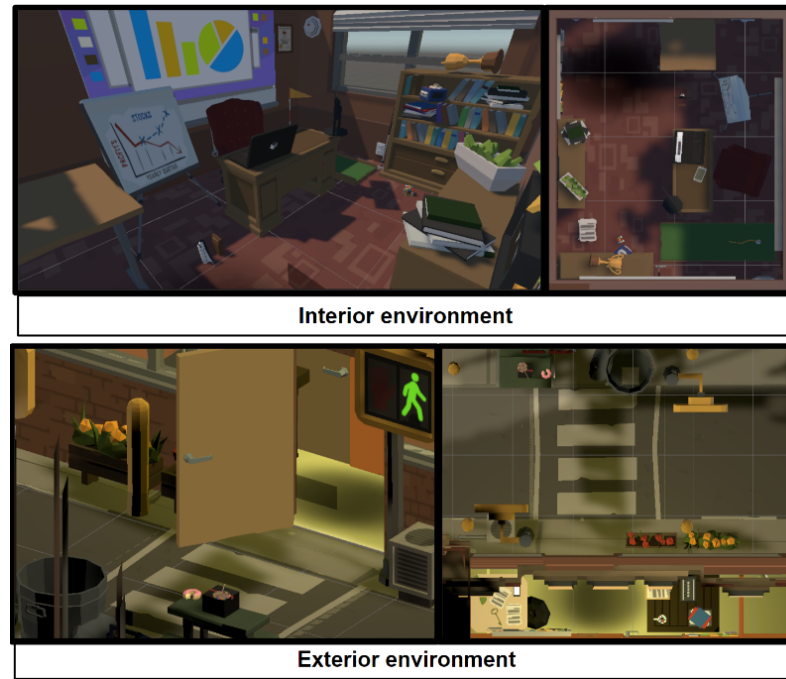


Figure 5.3 – The interior (top) and exterior (bottom) 3D environments were assigned to designers for the creation of their scenarios. Both environments offered a 5-meter by 5-meter design space. Objects can be moved and renamed to fit the designer’s intention. Participants assigned to the interior environments had to include in their scenario the following objects : a golf club, a golf ball, and any fallen objects. Participants assigned to the exterior environment had to include in their scenario the following objects : a key, a trashcan, and any sweet object.

With the data collected from this study, we present in the next section 5.2.5 our observations on the gulfs, showing correlations between issues, between issues and the proper narrative interpretation. We present feedback from the designers observing issues in their design during the post-test, and the impact it got on the communicated and lived experience.

5.2.5 Results

The user study guided through structured interviews allowed us to construct a rich dataset that contains gulfs of communication occurring in interactive narrative experiences mediated by VR. Using our classification methodology, we detected and successfully classified a total of 127 issues throughout the experiments performed with our 10 participants. A total of 74 of them were done on the designer side and 53 on the player side. The design issues and the performance issues were the most frequent ones for the two respective sides. A summary of the encountered issues extracted from the studies is presented in Table 5.3. This data allows us to perform two types of analysis to investigate : (1) the correlation between two different categories of issues, and (2) the impact of a category issue on the proper narrative interpretation by the players.

Here, we present a few illustrative examples. We then go over a quantitative and qualitative analysis of the issues observed, the correlations between issues, and the final comprehension, as well as some relevant feedback collected from the participants.

TABLE 5.2 – The correlation matrix on the eight issues and additionally the narrative interpretation. We highlight (in green) the issues that are potentially correlated with another one in a later part of the design-usage process, or in correlation with the score observed of narrative interpretation of the scenario by the player.

	Designer - System issues				Player - System issues				Narrative interpretation
	System design limitation	Design issues	Evaluation perception issue	Evaluation comprehension issues	Task comprehension issues	Situation comprehension issues	Environment comprehension issues	Performance issues	
System design limitation	1								
Design issues	-0,3	1							
Evaluation perception issue	0,2	-0,2	1						
Evaluation comprehension issues	-0,1	0,2	-0,7	1					
Task comprehension issues	0,2	-0,1	-0,1	-0,1	1				
Situation comprehension issues	0,2	0,2	0,1	0,0	-0,3	1			
Environment comprehension issues	0,2	0,0	0,3	-0,1	0,0	0,2	1		
Performance issues	0,2	0,3	0,4	-0,1	0,0	0,2	0,6	1	
Narrative interpretation	-0,4	0,6	0,1	0,0	-0,1	0,4	0,2	0,5	1
Total	0,3	0,5	0,3	0,0	0,1	0,5	0,6	0,8	0,5

TABLE 5.3 – All the issues observed during the 10 studies, classified under the eight issues specified by our model. User Y of the designed experience X is referred as DXUY, for example user 2 of the experience 7 is D7U2. The narrative interpretation row displays the score given for the accuracy of the narrative interpretation done by the user. A score of 0 means the narrative was interpreted as intended, 1 means partially correct, and 2 means an incorrect interpretation.

	Designer - System issues				Player - System issues				Narrative interpretation
	Design limitation	Design issues	Evaluation perception issue	Evaluation comprehension issues	Task comprehension issues	Situation comprehension issues	Environment comprehension issues	Performance issues	
D1U1	0	2	1	0	1	0	0	1	2
D1U2	0	2	1	0	0	0	1	2	2
D2U1	1	0	1	0	0	1	1	1	0
D2U2	1	0	1	0	0	1	0	0	0
D3U1	1	3	2	0	1	1	1	2	2
D3U2	1	3	2	0	0	1	2	2	2
D4U1	1	2	1	1	0	0	0	1	0
D4U2	1	2	1	1	0	2	1	2	2
D5U1	0	3	0	1	0	1	0	0	2
D5U2	0	3	0	1	0	1	0	1	2
D6U1	0	1	1	1	0	0	1	1	1
D6U2	0	1	1	1	0	0	1	1	1
D7U1	1	0	2	0	1	0	0	1	1
D7U2	1	0	2	0	0	0	1	1	1
D8U1	1	1	1	1	0	1	1	1	2
D8U2	1	1	1	1	0	1	1	2	2
D9U1	1	2	0	1	0	0	1	1	1
D9U2	1	2	0	1	0	0	0	1	1
D10U1	1	1	0	1	1	0	1	1	0
D10U2	1	1	0	1	2	0	1	1	1
Total	14	30	18	12	6	10	14	23	25
Average	0,7	1,5	0,9	0,6	0,3	0,5	0,7	1,15	1,25

5.2.5.1 Example scenarios

In the ten scenarios created by participants, we observed numerous issues with a variable impact on each of the actors : the designer, player, and system. Here, we illustrate three study cases to highlight how we identify and classify some issues encountered.

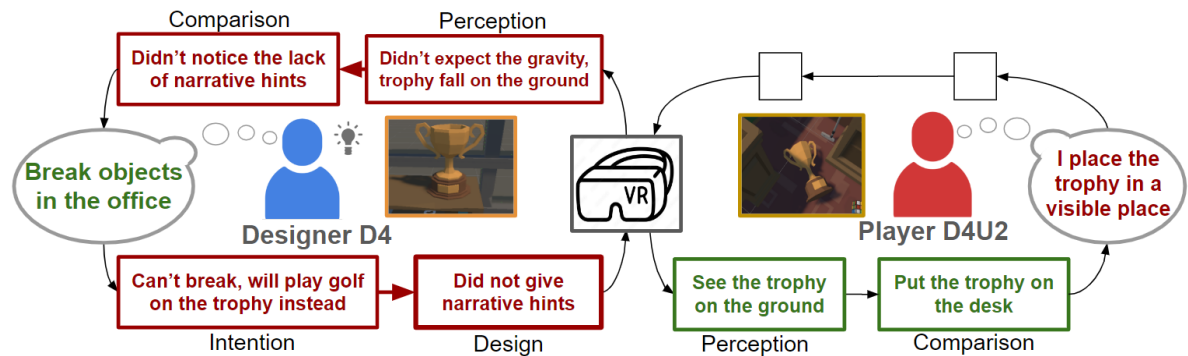


Figure 5.4 – The designer D4 created a scenario, to simulate an office worker in burnout, where the player has to play golf on a precious trophy. The player D4U2 did the tasks properly, however, the player didn't understand the tasks the way the designer expected, interpreting this set of tasks as a desk tidying scenario.

Case 1 : Gulfs in design and evaluation resulting in gulfs in scenario comprehension and narrative comprehension (D4,D8) : The designers of the experiences 4 and 8 both used the interior environment for their scenario. It was experienced by the players D4U1, D4U2, D8U1, and D8U2. A part of the presented scenario and communication issues encountered is shown on Figure 5.4.

Design phase : Both designers had the same idea of creating a scenario to simulate an office worker experiencing burnout. They found this idea when seeing in the 3D scene the whiteboard object, which depicted a red arrow like the plunging of stock prices, along with the unordered objects on the room floor : "I want a scenario about a person working here, alone in the office, in burnout due to the company results. This person wants to break, burn, or throw things in their office's desk to let their anger out" (D4) "The person is depressed, is angry against the world, want to do crazy things" (D8). Both designers encountered system limitations as they couldn't physically burn or break objects. Instead, noticing the mini-golf field, they decided to create tasks to position the "laptop" (D8) or the "golden trophy" (D4), which they considered as "precious objects", on the "golf field" and "office desk" respectively, and then to use the objects "golf club" and "golf ball", to play golf on them. However, no textual information or narrative elements were used to inform the players of the meaning of these interactions. During their self-testing, neither of the designers realized the lack of communication on the meaning of the interactions, leaving the scenario largely open to the player's interpretation. In addition, D4 encountered an evaluation perception issue, not expecting gravity to apply on objects in the scenes on experience start, resulting in the trophy falling on the ground.

Usage phase : All players successfully performed the tasks in the 3D environment. D4U1 interpreted the task of placing the "golden trophy" on the "golf field" to "showcase them". D8U1 and D8U2 expressed confusion on the purpose of the scenario, questioning the presence of a purpose to begin with, as the sequence of tasks didn't make much sense to them. D4U2 was the only player able to correctly interpret the scenario as a burnout when seeing the plunging stocks and described the actions as violent. All three other players, due to the nature of VR allowing free physical and visual exploration of the environment didn't even notice plunging stocks on the whiteboard. D4U1, seeing the trophy on the ground, interpreted this task as tidying up the office.

Gulf evaluation : In the post-test, both D4 and D8 said they were aware that their scenario was hard to understand. D4 described their scenario as "a bit far-fetched" and was surprised to learn that D4U1 successfully interpreted their intention. After revealing the designer's intention to the players, D8U1 responded : "the scenario is not really well-communicated, I don't think I could have guessed it at all". In this case study, we thus observed a design issue (lack of context and information to communicate the

meaning of interactions), compounded by an evaluation comprehension issue (designers didn't realize the lack of communication on interaction meaning during the test phase). These two communication issues resulted in situation comprehension issues for 3 players (D4U2, D8U1, D8U2), who did not understand the meaning of the scenario at all. These communication issues resulted in a gulf between the designer intentions and player interpretation and in an incorrect narrative interpretation by D4U2, D8U1, and D8U2.

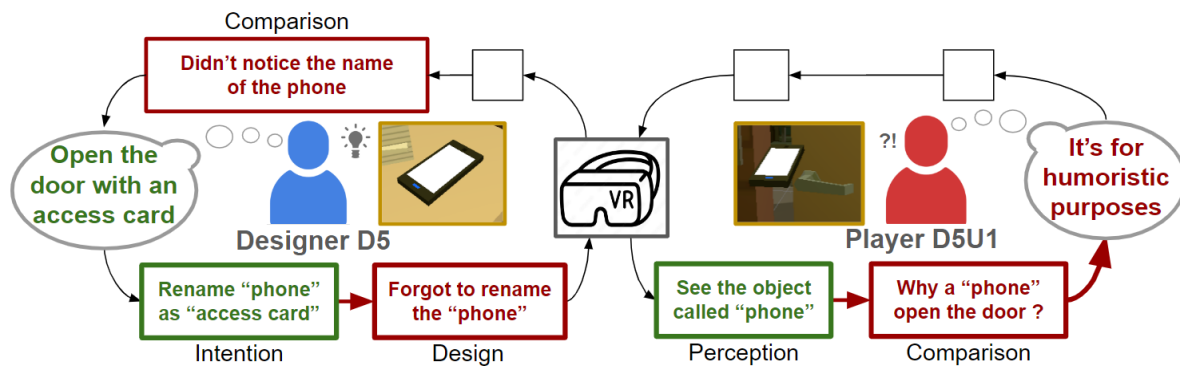


Figure 5.5 – D5 created a scenario, where players perform a number of interactions, one of which is to open a door with an access card. However, the designer forgot to rename the target object (the phone) as an access card. D5U1 completed the tasks, but did not understand the logic of opening a door with a phone, interpreting this interacting as humorous.

Case 2 : Design and evaluation gulfs resulting in system interaction gulfs (D5) : D5 used the exterior environment. It was experienced by the players D5U1, D5U2. A part of the presented scenario and communication issues encountered are shown on Figure 5.5.

Design phase : D5 wanted to create a simple scenario where the player have to go outside from home to collect some sweets for an afternoon snack. They created tasks to “open the door of the house using the access card”, “go outside”, and then tasks to bring home multiple “sweets” objects. The designer intended to use a phone-looking object as the the card needed to open the door by renaming it “access card”. However, during the design phase, D5 forgot to change the name of the object, resulting in a design issue. During the test phase, D5 didn't notice that the name of the object didn't matching their initial design intention, resulting in an evaluation comprehension issue. Because of this, the final scenario asks the player to take the phone and open the door with it.

Usage phase : Both the players D5U1 and D5U2 performed the scenario's tasks properly. Both players, however, didn't understand the reason the task asked them to use the phone as a key to open a door, nor why the phone did actually open the door. The rest of the scenario was performed correctly.

Gulf evaluation : In the post-test, D5 realized the confusion with an object's name not matching the tasks performed. As a result, both D5U1 and D5U2 encountered situation comprehension issues. D5U1 said that presenting an interaction that makes no sense was probably for humoristic purposes. On the other hand D5U2 said that either the interaction made no sense or they had missed the meaning. This created a communication gulf between the designer, who wanted to simulate a normal interaction with an access card, and the players, who both interpreted this interaction as an interaction that makes no sense. During the post-test, looking at the records, D5 first assumed that players understood the interaction's meaning, only to discover at the reveal of the interpretations the communication gulf around the phone interaction. D5 then said, “I can understand the gulf; it's true that opening a door with a phone isn't common; that's fun, but I didn't expect them to go as far as believing it's an absurd scenario for humoristic purposes.”

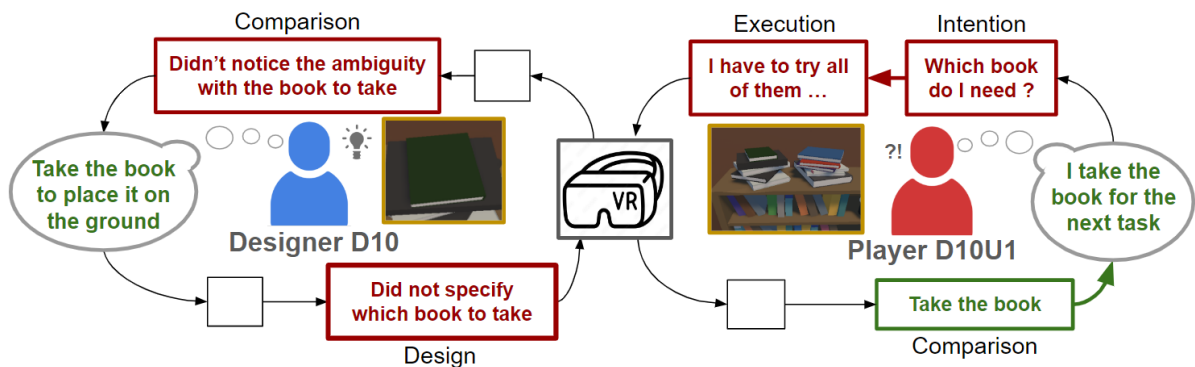


Figure 5.6 – The designer D10 created a scenario where players have to take a book and place it on the ground. However, the designer did not specify which book exactly to take and didn't notice this issue in the test phase. The player D10U1 did understand the meaning of the task, however, due to the lack of detail on the book to take in a scene full of books, the player took some time to find the right one.

Case 3 : Multi-interaction between all four gulfs (D10) : D10 used the exterior environment, experienced by the players D10U1, D10U2. A part of the presented scenario and communication issues encountered is shown in Figure 5.6.

Design phase : D10, motivated by the curiosity to experiment with how physics behaves in VR, decided to create a scenario based solely on 3D interactions, where players had to find, take, and place objects. Players were first asked to pile up multiple objects found in the 3D environment. The first group of tasks involve placing the “cube” on the “pile of paper”, then place the “ball” on the “cube” previously placed “pile of paper”. Then the second group of tasks involves placing the “plant” and the “book” and place them on the “ground”. Finally, a task was created to ask players to place the “golf club” in balance on top of the “book” and the “plant” previously placed on the “ground”. Multiple design issues occurred. For example ; in the task to take the “book”, there were over ten book objects in the scene. This design issue was linked to an evaluation issue, as D10 didn't notice the issue, since they knew exactly which book to interact with.

Usage phase : Both players D10U1 and D10U2 successfully piled up the objects of the first group of tasks in the scenario. On the second group of tasks, they both encountered similar task comprehension issues : the task did not specify which book was concerned. Both players encountered a performance issue, trying multiple books without being sure which one was the right one for the task. Afterward, both players had difficulty knowing exactly how to place the “golf club”, given the rich 3D interaction possibilities VR offers, and the designer's instruction wasn't explicit enough.

Gulf evaluation : During the post-test phase, D10 immediately realized the issue with the books : “*I should have specified which book, something like : the blue book next to the plant*”. Multiple comparable issues were encountered during this scenario, linked to a lack of information given by the designer on how or which objects to interact with, notably on how to place the “book”, and how to place the “golf club”. However, while these issues impacted the proper 3D interaction and task execution, they didn't have a strong impact on the narrative interpretation. The narrative of the scenario of experimenting with piling objects was partially understood by D10U2 and fully understood by D10U1.

These case studies were selected to illustrate the difficulty of achieving a full understanding of the initial designer intentions, but found that a partial understanding is mostly feasible. The overall narrative comprehension is shown in Table 5.3, where 0 means “Fully understood”, 1 mean “Partially understood” and 2 mean “Not understood at all”. We also observed issues that fit under multiple classes, potentially cascading into other issues in later stages of the model and widening the gulf between the designer and the player.

5.2.6 Correlation between issues

While the prior case studies present a number of complicated causal relations between issues and communication gulfs, we now investigate in more detail the correlations between multiple types of issues, as well as their impact on narrative interpretation. For this purpose, we build a correlation matrix as shown in Table 5.2 to highlight potential correlations between two issues, where -1 mean a negative correlation, 0 an absence of correlation, and 1 a correlation. This includes the score on the interpretation of the narrative given by players to estimate if an issue have an impact on the scenario comprehension. Here we discuss the results obtained for each issue and their potential impact :

Design limitation issue : two minor negative correlations (-0.3 and -0.4) exist between design issues and the narrative interpretation, indicating that setting design limitations could reduce design issues and better narrative interpretation. We believe this can be linked to the fact that designers who encountered limitation issues often tried to simplify their intentions, lowering the chance of design issues, making the scenario easier to communicate and interpret. For example, D7 wanted to make players run around and eat objects in the scene, but due to the limitations, they changed the scenario to find hidden objects that looked edible.

Design issues : there is a minor correlation of 0.3 with performance issues, and a notable correlation of 0.6 with the narrative interpretation. This can be because a design issue is unconscious to the designer and affects the communicated intention through the system. For example, when D5 forgot to rename the “phone” as an “access card”, the system communicates an erroneous message to the player, changing the interpretation of the players in an unexpected way.

Evaluation perception issue we observed two minor correlations of 0.3 to 0.4 between environment comprehension issues and performance issues. This can be explained by the fact that an evaluation perception issue corresponds to a gulf between the designer intention and what the system actually displays. For example, D3 noticed that a target object was not visible enough when testing. This can result in an environment comprehension issue that makes the task harder for the players (performance issues) requiring more effort and time than originally expected. We also observe a negative correlation of -0.7 with evaluation comprehension issues, meaning that when the designer is conscious of a gulf between what they wanted and what the system displayed, the unconscious gulf between their intentions and the actual scenario communicated was reduced. For example, D1 noticed during testing that placing an object (a “lollipop”) on the ground in VR can be dangerous as it asks people to bend down with a VR headset on. Due to this issue, the designer foresaw that the task might be done slowly or incorrectly.

Evaluation comprehension issues & Task comprehension issue : No notable correlations were observed with these two issues in relation to other issues later in the process.

Situation comprehension issue A minor correlation of 0.4 is present between situation comprehension issues and narrative interpretation. This can be because this issue corresponds to the proper understanding of the purpose of the tasks. For example, D4U1 and D4U2 encountered a situation comprehension issue when they didn’t understand why they opened a door with a phone, and for this reason, they misinterpreted the overall narrative the scenario was communicating.

Environment comprehension issues we observe a notable correlation of 0.6 with performance issues, which is expected. When a player doesn’t properly understand what the task is, they will often solve the problem through trial and error. For example D6U1 and D6U2 had difficulty finding a “penguin” (a spaceship toy renamed “penguin”), resulting in slow performance.

Performance issue there is a correlation of 0.5 between performance issues and narrative interpretation. Two explanations are possible. First, players failing to perform the tasks properly could need more concentration to succeed, spending less attention understanding the task purpose. Another possibility is that this issue is the last one in the design-usage process, thus being the accumulated outcome of multiple previous issues. We observe a strong correlation of 0.8 between the total number of issues and the performance issue type, meaning that more issues often means more performance issues.

Finally the last row of Table 5.2 shows a 0.5 correlation between the total amount of issues observed and the narrative interpretation, meaning that a higher number of issues also correlates with a worse interpretation by the player. Situation comprehension, environment comprehension, and performance issues all show a correlation above 0.5 with the total, which identify these issues as key impacts on good comprehension and

interaction, and directly linked to narrative interpretation. On the design phase, only design issues displays a correlation above 0.5, making it the main factor impacting narrative interpretation in the design phase.

These analyses allow us to have a view on the impact of the different issues, and which ones have the strongest influence on proper narrative interpretation. Based on our analysis, we believe the design issues, the situation comprehension issues, environment comprehension issues and the performance issues are the most meaningful issues to investigate if we want to reduce the gulf designer-player, as they are the issues that seem to be the ones that cause the biggest numbers of communication issues, as they show strong correlations with other issues and with a low narrative interpretation score.

5.2.7 Post-test insight

This study and analysis allowed us to highlight the issues, conscious or not, that designers and players encountered. All ten designers participated in the post-test to observe the issues in their design and suggested improvements to make an experience that better matched or communicated their intention. Out of the 20 players, three believed that the explanation behind the issues was the players themselves and not what the system communicated, three felt there were no issues, and the remaining fourteen found in addition one or more improvements inspired from others' designs that they could make to their own scenarios to better communicate their initial intention. We detail here a few examples of issues that seven designers discovered during the post-test in their created scenarios, as well as their ideas to prevent these issues.

D2 and D6 were very satisfied with their scenarios after seeing player replays. D2 observed the players and saw them do exactly what was intended, and felt nothing needed to be modified. On the other hand, D6 was stressed when they realized that the design and evaluation issues encountered during the design phase created an experience potentially too hard or frustrating compared to what was intended due to a lack of instructions. However, seeing a very good comprehension and enjoyment of the scenario by the players, where the lack of instructions created an escape-game like experience, they concluded the scenario is better the way it is, even if it wasn't what they intended originally.

Multiple designers noticed issues created due a lack of clarity in the vocabulary used. D3, D9 and D10 realized some object names and models were confusing, and most of their players encountered environment comprehension issues (i.e., confusing brown table with black, newspaper holder with newspaper, and Rubik's cube with cube). Both D3 and D9 believe they should have added more instructions or renamed the objects, while D10 believes the observed issue is a rare case which shouldn't require changes. D10 and D4 encountered comparable issues : D4 agreed that the phone should have been renamed as access card to more clearly communicate the meaning of the task. D10 noticed players had trouble identifying the correct "book" amongst multiple similar objects and suggested that they could be more precise, such as "small blue book".

D4, D8 and D9 noticed a bad narrative interpretation as well as multiple situation comprehension issues by their players, due to a lack of instructions in their scenarios. Multiple tasks required players to perform interactions (e.g., place the golden trophy on the table, go to the entrance with the golf club) without giving context as to why these tasks are relevant. All three believe this is the main reason why their scenarios were misunderstood by their players.

From the 10 participants and analysis of over 100 issues on the design and the test of experiences, this study highlights multiple important issues to the designers. The way designers proposed potential solutions to reduce the designer-player gulf in their VR experiences based on the structured interview responses was also very insightful, such as by adding more instructions to better communicate the scenario intention, renaming objects in the environment to avoid confusion, or changing the task order to prevent potential incomprehension by the player.

5.2.8 Discussion

We have presented a communication model revisiting Norman's seven stages of action, and a structured interview to evaluate and characterize the different issues occurring during the design-usage process of a VR narrative experience, leading to gulfs of comprehension. The challenge of reducing the gulf between the designer's intended way to use a system and the player's actual interaction and interpretation can be significant for research in highly immersive media such as VR : an experience not lived the designer's intended way could alter the way the data is collected, resulting in a less successful investigation of the research questions. This is especially true in immersive environments, where observing and understanding the experience can be more complex due to the free physical and visual exploration possibilities offered by the 3D environment, as well as the possibility to use the body to physically move and interact with objects.

The presented model defines a taxonomy of issues related to the design and the usage of a VR experience. This taxonomy was efficient for the context of our study, allowing us to classify all 127 issues encountered into the eight categories in order to observe the link between issue, gulf, and narrative interpretation. By analyzing the results obtained from the study through the structured interviews, we were able to identify correlations between multiple issues and observe how compounding issues result in a wider designer-player gulf. More specifically, we see a meaningful impact on the narrative interpretation for the design issues and situation comprehension issues. This means that during the majority of our studies, unconscious issues performed by the designers and an issue on the player side to understand the purpose of the given task were the main reasons for inefficient designer-player communication.

A limitation linked to our model is the lack of representation of choices : when designing or experiencing a scenario in VR, the designer can consciously decide to change the initial intention, or the player can decide to not follow the presented scenario. While this is a case we did not encounter in our study, this could cause problems in the later parts of the analysis. This would result, for example, in a gulf between what the designer initially intended to create and what they actually created, but without any issues collected and classified to explain this gulf. A way to improve this could be to explore how to include the reiterating nature of the communication model in the taxonomy : in the design of a system, the system intention is usually not fixed and can evolve over time.

Another limitation lies in the analysis presented in this paper, which is primarily focused on the proper comprehension of the designer's intention by the player who lives the experience, presented as the narrative interpretation. The interaction performed could be further investigated by examining if the player did interact with the system the way the designer expected them to. This is a case encountered with D10, who wanted players to place a plant on the ground, but both D10S1 and D10S2 dropped the plant on the ground instead, not satisfying the designer. Our communication model allows the observation and classification of these elements, however deeper investigations are needed to precisely characterize the articulatory distance between the designer's intended usage and the real usage, in the same way that we investigated the semantic distance by defining a score of narrative interpretation.

We believe the methodology proposed, by allowing the detection and classification of communication issues, could support future work in investigating ways to detect and prevent issues. On the designer side, usability heuristics could be efficient to provide guidelines to designers to help avoid design issues (Sutcliffe & Gault, 2004). Cognitive walkthroughs could be efficient for the prevention of evaluation perception issues (Doumanis & Economou, 2019 ; Q. Liu & Steed, 2021) to simulate experience usage. On the player side, the usage of system dynamics modeling coupled with guidance methodologies (Nielsen et al., 2016 ; Sassatelli et al., 2018) could be very efficient ways to address environment comprehension and performance issues by reorienting players when they deviate from the expected experience. Surveys such as the System Usability Scale (SUS), NASA-TLX, or more recently VR System Usability Questionnaire (VRSUQ) (Y. M. Kim & Rhiu, 2024), could be used to address the environment comprehension and performance issues by gaining insight on the usability viewed by the player. The analysis of logs such as task time, or complex data like eye tracking and body movement (David-John et al., 2021a) could also be used to address environment comprehension and performance issues, to finely observe the player experience and understand the source of issues encountered. In both the design and the usage, think-aloud protocol (X. Zhang & Simeone, 2022)

can be an efficient solution used to encourage designers and players to explicitly spell out any intention or observation when designing or interacting, to pinpoint the exact time and context of potential issues.

5.3 Conclusion

We presented in this chapter our communication model, which aims at describing the design-usage process of an embodied VR experience and the issues in communication and comprehension potentially hurting the good communication of the narrative. To investigate the RQ3A, we presented our taxonomy as well as our structured interview, which was revealed to be suitable for our user study, allowing us to identify 127 issues and categorize all of them among the eight defined categories.

This work explores the RQ3B through two contributions : (1) enabling the observation of the issues and their impacts on the scenario ; and (2) supporting designers in the refinement of their experience. The most meaningful observed categories of issues were design issues, situation comprehension issues, environment comprehension issues, and performance issues, which need to be addressed in priority to reduce the designer-user communication gulf. Through the post-study phase, we received a majority of positive feedback from the designers who got to identify the issues in their design, helping them to determine means to improve the built experience.

Contributions The publication presented through this chapter :



F. Robert, H.-Y. Wu, L. Sassatelli, and M. Winckler. “The triangle of communication in interactive virtual narratives : gulfs between designers, system and players”. (Under review for SIGCHI 2025).

CHAPTER 6

Conclusion

This thesis explored the three components of the theory of embodiment proposed by Dourish (2004), with the purpose of reducing the comprehension gap existing between designers and users of embodied experiences in virtual reality. We proposed frameworks, datasets, methodologies and models to support designers on the creation, observation and comprehension of embodied experiences. In this last section, we reflect on our findings and the next contributions these three thesis chapters lead towards.

6.1 Contributions

To help designers of embodied experiences to understand and refine the UX, we took a user-centered approach to design systems that sought to reduce the designer-user communication gap. We performed this work through three research questions :

- **RQ1** : How can we support designers in the creation and observation of embodied interactions in rich, annotated 3D contexts ?
- **RQ2** : How can we analyze intentionality from the user’s embodied interactions in immersive experiences using multimodal metrics, and how can this analysis be used to provide feedback to designers ?
- **RQ3** : How can we specify the intersubjectivity gap between design intention and actual usage in embodied interactive experiences, and how can we use this specification to provide feedback to the designers ?

We present here an overview of the contributions of each chapter, and how they are connected to each other.

The first chapter explores the **creation, management, and observation of rich interactive embodied VR experiences**. This chapter started with the introduction of our conceived framework GUsT-3D. The contribution and novelty of this framework lies in the usage of a modular vocabulary that supports designers in annotating the interaction and navigation properties of 3D environments. This foundation enables interactive scenario creation and observation possibilities, through the generation of scene graphs with rich attributes, the creation of guided tasks, and the logging of interactive and contextual data at a granular level. To validate the efficiency of our framework, **we conducted an embodied VR study using GUsT-3D, with the collection of multimodal data**. During this user study, participants engaged in physical interaction and navigation to interact with the system and perform virtual road crossing tasks. The main contribution of this work lies in the technical basis to enable the usage of embodied interactions, physical navigation, and the collection of multimodal data. This work serves as the foundation for the second chapter’s work by allowing the creation of our CREATTIVE3D large immersive dataset opening ways for granular embodied experience analysis.

The second chapter explores the **analysis and understanding of embodied experiences in VR through multimodal data**, with immersive dataset creation and analysis methodology conception. This chapter started with the introduction of our CREATTIVE3D and PEM360 immersive datasets. We stand apart from other immersive datasets, as these datasets were built with the purpose to enable in-depth analysis of the embodied experiences. The main contribution lie in the significant number of multimodal data (emotion, motion, attention, interactions, and context) collected, temporally and spatially synchronized on a granular task level in an immersive context. To work on these data, we conceived our **task-based methodology to analyze and characterize embodied experiences**. We applied this methodology to the dataset CREATTIVE3D to analyze on a granular level the correlation between physical behavior (body motion, skin conductance,

gaze movement) and scenario performance (efficiency, attention, and decision). The use of task-based methodology applied to interactive virtual experiences is currently limited in the literature. We believe this methodology will open new ways to characterize embodied experiences to create a user's profile and support designers in the refinement of personalized guidance and scenarios.

The last chapter explores the **comprehension and improvement of the inter designer-user intersubjective communication mediated by an embodied VR experience**. The main contribution and novelty of this work is our conceived communication model to understand and improve designer-user communication mediated by an embodied VR experience, a subject still little explored in the VR literature. This communication model supports the specification of a taxonomy of communication issues to identify and classify them during the design and usage process of embodied experiences. We built a methodology using the taxonomy and applied it in a study incorporating the complete design-usage workflow of an embodied VR experience. The methodology proved to be efficient in classifying all encountered communication issues and in supporting designers to understand communication issues and how to avoid them to build a better designer-user communication.

6.2 Limitations

There are a number of limitations on the different contributions. Most of the limitations can represent short term future work in direct link with the previously presented contributions.

User's intentionality evaluation The intention of the user when interacting with a system is a component that can support a more complete comprehension of the user's embodied experience, by observing the link between intention and interaction. While this thesis aims for a comprehension of intentionality, it primarily concentrates on behavioral and contextual metrics to analyze and understand embodied interactions. We believe we could go further on intention analysis, achieved through a survey design to collect the user's explicit intention during the usage of an embodied experience. By combining this explicit intention collection with multimodal data analysis, we can open a deeper understanding of the user's intention and how this intention can be observed and understood through metrics.

Reproducible framework for embodied studies VR technology evolves rapidly, and multiple techniques exist for the collection of behavioral data, such as gaze or motion data. This creates multiple issues when trying to share datasets, technical setups, or study protocols. We believe the domain could gain a lot by adopting common techniques and practices that are reproducible by designers to create datasets following the same established procedure, opening up more in-depth analysis possibilities on large datasets. We explored the idea with the CREATTIVE3D dataset, proposing a fully reproducible study that could be expanded with more studies and introducing the concept of a living contextual dataset. Likewise, we could contribute more in this sense on diverse aspects (i.e., study protocol, eye data collection, 3D scene sharing) to create an environment where the research is fully reproducible and thus encourage more collaborative works. Research domains using VR for user studies would greatly benefit from shared collection techniques and formats to improve the reproducibility of a study and the re-usability of the generated data.

Cyclical communication model The presented communication model breaks down the workflow of an embodied experience design and usage in 14 steps to specify a taxonomy of communication issues. This communication model is able to support the investigation of the cyclical nature of the design and usage processes, a point that wasn't taken into account in the proposed methodology and study. After creating an experience, the designer can evaluate it and refine it as many times as they want to. The designer intention and the user intention can both evolve with time, and this model can allow us further exploration on that topic to potentially create a richer taxonomy including that dimension of the designer-user communication.

Communication issues survey conception The specification and identification of communication issues, as shown in our contribution, supports designers in the refinement of experiences, but it can take a lot of time. Supported by our communication model, we presented a methodology to detect and classify designer-user communication issues. However, this methodology isn't applicable to any existing VR study as is, and reproducing it requires the composition of a task model and the creation of a custom-made structured interview. To expand the usage of this methodology more broadly to detect and classify communication issues, the next step is to design and validate structured interview templates or questionnaires based on this model of communication, for users and designers of embodied VR systems. The use of structured interviews to inquire about the design, evaluation, usage, and observation phases could help pinpoint the communication issues to support designers of embodied experiences.

6.3 Future work

There are a number of subjects that could be further explored to expand on each presented contribution. For example, how can we further improve the ontology to more accurately describe embodied experiences? Can guidance systems support designers in building better intersubjective communication? Can visualization systems improve the designer's comprehension of the user's intentionality? As the subject of supporting designers in understanding embodied experiences is very broad, we see numerous perspectives to further develop it. Most of the perspectives presented here were at some point discussed during the thesis as potential directions to take during the thesis. In this section, we provide a wider perspective on the overall contributions of this thesis, the areas that have not been explored deeply, and the new path to explore opened by this thesis.

6.3.1 Rich experience management and creation

The first identified perspective, mostly in link with the contribution of the first chapter of the thesis, is the exploration of ways to further support designers in the management and creation of rich interactive embodied experiences. While our framework already explores the basis of rich interactive embodied experience creation using an ontology, the development of the ontology and of the scenario could be developed in many ways.

Complex scenario design and observation For applications and researches valuing life-like narration, such as training and simulation, supporting the design of complex embodied scenarios is a very valuable component. Using GUsT-3D's framework as a basis, already allowing the creation and granular observation of embodied experiences, multiple directions could be considered to support more complex scenario design, such as the usage of probabilistic models to represent uncertainty (i.e., Markov Decision Processes) or scenarios comprised of tasks that can be performed in parallel. This work, however, represents multiple challenges to overcome, to observe granularly the new complexity that has been introduced: how to represent a task model, a scene graph, or an interactive experience, including parameters such as uncertainty or parallel tasks. This perspective is particularly interesting, as more complex scenario creation and observation could support more complex research using the controlled environment offered by VR.

Rich scene graph properties for embodied interaction description Defining a structured vocabulary and relationships that accurately describe how objects in an environment can be perceived and interacted with by a user is an efficient way to improve rich embodied interaction possibilities and observation. The presented GUsT-3D's scene graph already includes interaction properties on a basic level (grab, push, and drop). However many object properties (i.e., temperature, fragility, weight) and interactive properties (i.e., balance, throwability, attachability, squeezeability) could be added to better describe an embodied interactive environment. More detailed interaction properties can also help provide more accurate and granular logs of the user's interactions with the environment. Specifying precise properties could open perspectives to create systems to predict interactions in embodied VR experiences to expand the scenario and guidance possibilities, using system dynamics model or machine learning.

6.3.2 Granular experience visualization

The second identified perspective is the exploration of ways to further support the designer in the in-depth visualization of the embodied experience. While this thesis already proposes valuable user's embodied experience visualization through multimodal data analysis, we believe the research could go much further on experience-based visualization.

Query system for granular behavior visualization Experience visualization is the cornerstone of experience understanding and refinement. We believe the proposed frameworks and datasets open doors to the conception of visualization tools that can automate the granular analysis and visualization of multimodal data. We presented in our first contribution through GUsT-3D the Log and Query component to visualize during the post-scenario phase the logs related to the experience. In our second contribution, we presented our task-based analysis methodology to characterize on a granular task-level the user embodied experience using multimodal data. To take these ideas one step further, we could design an extension of the Log and Query component, using the defined GUTask as a task model, to perform an analysis of the behavioral metrics synchronized with the current context (e.g., tasks, interactions) on a task level. By automating a part of our task-based analysis methodology directly in GUsT-3D, we could offer designers access to in-depth visualization to obtain a granular understanding of user behavior in correlation with the context without requiring the performance of complex, handcrafted data analysis. This would open the door to further work on the possibilities opened by such tools, the impact it can have on embodied experience comprehension by the designer, and how it can help with experience refinement.

6.3.3 Deeper embodied experience analysis

The third identified perspective is the exploration of ways to improve the granular analysis of the UX in the VR embodied context. While this thesis already proposes a multimodal dataset along with methodology to analyze and characterize the user experience, we believe the research could go much further to support designers on user behavior analysis and comprehension.

Efficient metrics for embodied experience analysis While behavioral metrics are an important insight for embodied experience comprehension to observe the body used as a communication tool, we don't know exactly which metrics are the most efficient ones to characterize an experience. We used in this thesis multiple types of metrics : body motion, gaze movement, skin conductance, heart rate, etc. Using our task-based analysis methodology, we identified for our analysis that COPI (Center of Pressure Inclination) was much more relevant than the two other metrics used as examples (gaze fixation duration and skin conductance). We could engage in deeper exploration of diverse tasks and conditions to determine which metrics are most instructive when examining embodied experiences. As these sensors can be consuming in multiple ways (e.g., cost, time, user's cognitive load), it would be very valuable for future research performing embodied experience analysis to know more accurately what metrics are efficient to characterize embodied experience.

6.3.4 Designer-user communication improvement

The last identified perspective, mostly in link with the last contribution of this thesis, is the exploration of more ways to understand and improve the designer-user communication in the embodied context. While this thesis proposes a methodology to classify and understand communication issues during design and usage, we believe the research could go further to support designers on designer-user communication improvement.

Communication issues detection As highlighted in our work, many communication issues can arise during the design and usage of an embodied experience. Being able to identify the issues and how to avoid them greatly support designer-user communication improvement. In our work, we identified eight categories of communication issues and their individual impact. We believe the next important step is to explore diverse methodologies to detect and prevent communication issues on the design side, such as the design

of guidelines and surveys to support designers in understanding their missteps. On the usage side, multiple options could be explored to resolve some communication issues, such as using personalized guidance to reorient users deviating from the specified tasks. For this purpose, we could use system dynamics modeling to support designers in the creation of scenarios that adapt to user behavior over the time of the scenario. Techniques to help designers on detecting and preventing communication issues would greatly benefit VR research by allowing a reduction of the designer-user communication gulf.

Designer-User communication quality evaluation As presented in our initial state of the art on designer-user communication, two methodologies are broadly adopted in HCI to reduce designer-user communication gulfs. The first one is the construction and usage of an ontology to specify and understand issues, the second one is the creation of a methodology to evaluate the quality of communication (C. S. D. Souza & Leitão, 2009). Thus, the next valuable step of the presented communication model would be to explore ways to evaluate the quality of the communication performed through the embodied VR experience. Used in conjunction with our proposed model, we could open opportunities for deeper comprehension of the communications issues and potential ways to prevent them.

6.4 Summary

The common feature of all of these future directions of embodied experience comprehension is the need for multimodal immersive data collection and visualization, to supports designers in experience refinement. A better UX comprehension in the VR embodied context is linked to a better analysis of the lived experience : personalized guidance asks for precise task definition and observation, and an in-depth experience visualization asks for granular multimodal analysis. Ultimately, the purpose is to provide techniques for the creation of better embodied experiences, which can greatly benefit a wide variety of research and application domains (e.g., sport, health, games), where the simulation of life-like narration and interaction is a very valuable feature.

Contributions

Publication in international conferences

1. **F. Robert**, H.-Y. Wu, L. Sassatelli, and M. Winckler. “Task-based methodology to characterise immersive user experience with multivariate data”. IEEE VR 2024, Orlando, Florida. [hal-04446066](#)
2. **F. Robert**, H.-Y. Wu, L. Sassatelli, and M. Winckler. “An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual Reality”. ACM IMX 2023, Nantes, France. **Best Paper Award**. [hal-04102737](#)
3. Q. Guimard, **F. Robert**, C. Bauce, A. Ducreux, H.-Y. Wu, L. Sassatelli, M. Winckler, A. Gros. “PEM360 : a dataset of 360° videos with continuous physiological measurements, subjective emotional ratings and motion traces”. MMSys 2022, Athlone, Ireland. [hal-03710323](#)
4. Q. Guimard, **F. Robert**, C. Bauce, A. Ducreux, H.-Y. Wu, L. Sassatelli, M. Winckler, A. Gros. “On the link between emotion, attention and content in virtual immersive environments”. MMSys 2022, Athlone, Ireland. [hal-03825059](#)
5. H.-Y. Wu, **F. Robert**, T. Fafet, B. Graulier, B. Passin-Caneau, L. Sassatelli, M. Winckler. “Designing guided user tasks in VR embodied experiences”. ACM EICS 2022, Biot, France. [hal-03635452](#)

Currently under review

1. **F. Robert**, H.-Y. Wu, L. Sassatelli, and M. Winckler. “The triangle of communication in interactive virtual narratives : gulfs between designers, system and players”. (Under review for SIGCHI 2025).
2. H.-Y. Wu, **F. Robert**, F. Franco Gallo, K. Pirkovets, C. Quéré, J. Delachambre, S. Ramanoël, A. Gros, M. Winckler, L. Sassatelli, M. Hayotte, A. Menin, P. Kornprobst. “Exploring, walking, and interacting in virtual reality with simulated low vision : a living contextual dataset”. (Under review for Nature Scientific Data.) [hal-04429351](#)

Publicly available framework

1. GUsT-3D : <https://project.inria.fr/creative3d/gust-3d/>

Publicly available datasets

1. CREATTIVE3D : <https://zenodo.org/records/10406560>
2. PEM360 : <https://gitlab.com/PEM360/PEM360/>

References

- Alcañiz, M., Rey, B., Tembl, J., & Parkhutik, V. (2009). A neuroscience approach to virtual reality experience using transcranial doppler monitoring. Presence : Teleoperators and Virtual Environments, *18*(2), 97–111.
- Araújo, J. P., Li, J., Vetrivel, K., Agarwal, R., Wu, J., Gopinath, D., . . . Liu, K. (2023). Circle : Capture in rich contextual environments. In Proceedings of the ieee/cvf conference on computer vision and pattern recognition (pp. 21211–21221).
- Arlati, S., Colombo, V., Ferrigno, G., Sacchetti, R., & Sacco, M. (2020). Virtual reality-based wheelchair simulators : A scoping review. Assistive Technology.
- Aseeri, S., & Interrante, V. (2021). The influence of avatar representation on interpersonal communication in virtual social environments. IEEE transactions on visualization and computer graphics, *27*(5), 2608–2617.
- Baños, R. M., Botella, C., Rubió, I., Quero, S., García-Palacios, A., & Raya, M. L. A. (2008). Presence and emotions in virtual environments : The influence of stereoscopy. Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society, *11* 1, 1-8.
- Behr, J., Dähne, P., & Roth, M. (2004). Utilizing x3d for immersive environments. In Proceedings of the ninth international conference on 3d web technology (p. 71–78). New York, NY, USA : Association for Computing Machinery. doi: 10.1145/985040.985051
- Bender, S. M., & Sung, B. (2021). Fright, attention, and joy while killing zombies in Virtual Reality : A psychophysiological analysis of VR user experience. Psychology & Marketing, *38*(6), 937–947. (_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mar.21444>) doi: 10.1002/mar.21444
- Benthall, J. (1973). Body as a medium of expression (Vol. 5) (N° 1). BLACKWELL PUBL LTD 108 COWLEY RD, OXFORD, OXON, ENGLAND OX4 1JF.
- Bernhaupt, R., Martinie, C., Palanque, P. A., & Wallner, G. (2020). A generic visualization approach supporting task-based evaluation of usability and user experience. In R. Bernhaupt, C. Ardito, & S. Sauer (Eds.), Human-centered software engineering - 8th IFIP WG 13.2 international working conference, HCSE 2020, eindhoven, the netherlands, november 30 - december 2, 2020, proceedings (Vol. 12481, pp. 24–44). Springer.
- Beßler, D., Porzel, R., Pomarlan, M., Beetz, M., Malaka, R., & Bateman, J. (2020). A formal model of affordances for flexible robotic task execution. In Ecai 2020 (pp. 2425–2432). Santiago de Compostela, Spain : IOS Press.
- Biocca, F. (1997). The cyborg’s dilemma : Progressive embodiment in virtual environments. Journal of computer-mediated communication, *3*(2), JCMC324.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability (formerly : Journal of Personnel Evaluation in Education), *21*(1), 5–31.
- Blumenthal, S., & Bruyninckx, H. (2014). Towards a domain specific language for a scene graph based robotic world model.
- Boldt, M., Bonfert, M., Lehne, I., Cahnbley, M., Korschinq, K., Bikas, L., . . . Smeddinck, J. (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) (p. 143-150). Tuebingen/Reutlingen, Germany : IEEE. doi: 10.1109/VR.2018.8446177
- Bolton, M. L. (2017, décembre). A task-based taxonomy of erroneous human behavior. International Journal of Human-Computer Studies, *108*, 105–121. doi: 10.1016/j.ijhcs.2017.06.006

- Borras, J., Boix-Granell, A., Foix, S., & Torras, C. (2023). A virtual reality framework for fast dataset creation applied to cloth manipulation with automatic semantic labelling. In 2023 IEEE International Conference on Robotics and Automation (ICRA) (pp. 11605–11611).
- Boston, C. (2002). The concept of formative assessment. Practical Assessment, Research, and Evaluation, 8(1), 9.
- Boucsein, W. (2012). Electrodermal activity, 2nd ed. New York, NY, US : Springer Science + Business Media. (Pages : xviii, 618) doi: 10.1007/978-1-4614-1126-0
- Bouville, R., Gouranton, V., Boggini, T., Nouviale, F., & Arnaldi, B. (2015, mars). #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) (pp. 33–40). Arles, France : IEEE. doi: 10.1109/SEARIS.2015.7854099
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion : The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), 49–59. doi: [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Braithwaite, J. J., Watson, D. P. Z., Jones, R. S. G., & Rowe, M. A. (2013). Guide for Analysing Electrodermal Activity & Skin Conductance Responses for Psychological Experiments. CTIT technical reports series.
- Brutzman, D., & Daly, L. (2010). X3d : extensible 3d graphics for web authors. San Francisco, CA, USA : Elsevier.
- Buffa, M., & Lafon, J.-C. (2000). 3d virtual warehouse on the web. In 2000 IEEE Conference on Information Visualization. An International Conference on Computer Visualization and Graphics (pp. 479–484). London, UK : IEEE.
- Burmester, M., Mast, M., Jäger, K., & Homans, H. (2010). Valence method for formative evaluation of user experience. In Proceedings of the 8th ACM conference on designing interactive systems (pp. 364–367). Aarhus, Denmark : ACM.
- Carroll, J., SINGLEY, M., & Rosson, M. B. (1992, septembre). Integrating theory development with design evaluation. Behaviour & Information Technology - Behaviour & IT, 11, 247–255. doi: 10.1080/01449299208924345
- Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., & Göbel, S. (2021). Cybersickness in current-generation virtual reality head-mounted displays : systematic review and outlook. Virtual Reality, 25(4), 1153–1170.
- Cavalcanti, J., Valls, V., Contero, M., externe, L. v. u. s., fenêtre, c.-c. s. d. u. n., Fonseca, D., . . . fenêtre, c.-c. s. d. u. n. (2021). Gamification and Hazard Communication in Virtual Reality : A Qualitative Study. Sensors, 21(14), 4663. (Num Pages : 4663 Place : Basel, Switzerland Publisher : MDPI AG) doi: 10.3390/s21144663
- Chan, S., Zhang, H., & Nanayakkara, S. (2023). Eye movement analysis of human visual recognition processes with camera eye tracker : Higher mean and variance of fixation duration for familiar images. In Extended abstracts of the 2023 CHI conference on human factors in computing systems (pp. 1–8).
- Chen, H., Wang, Y., & Liang, W. (2022). Vcoach : Enabling personalized boxing training in virtual reality. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 546–547).
- Chen, T., & Guy, S. (2018). GIGL : A domain specific language for procedural content generation with grammatical representations. In Proceedings of the AAAI conference on artificial intelligence and interactive digital entertainment (Vol. 14(1)). Edmonton, Canada : AAAI.
- Cheng, L.-P., Ofek, E., Holz, C., & Wilson, A. D. (2019, mars). 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) (pp. 359–366). Osaka, Japan : IEEE. doi: 10.1109/VR.2019.8798074

- Chevallier, P., Trinh, T.-H., Barange, M., De Loor, P., Devillers, F., Soler, J., & Querrec, R. (2012). Semantic modeling of virtual environments using mascaret. In 2012 5th workshop on software engineering and architectures for realtime interactive systems (searis) (pp. 1–8). Costa Mesa, CA, USA : IEEE.
- Christensen, J. V., Mathiesen, M., Poulsen, J. H., Ustrup, E. E., & Kraus, M. (2018). Player experience in a vr and non-vr multiplayer game. In Proceedings of the virtual reality international conference-laval virtual (pp. 1–4).
- Clifford, R. M., Jung, S., Hoermann, S., Billinghamurst, M., & Lindeman, R. W. (2019, mars). 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) (pp. 181–189). Osaka, Japan : IEEE. (ISSN : 2642-5254) doi: 10.1109/VR.2019.8797889
- Crilly, N., Good, D., Matravers, D., & Clarkson, P. J. (2008, septembre). Design as communication : exploring the validity and utility of relating intention to interpretation. Design Studies, 29(5), 425–457. doi: 10.1016/j.destud.2008.05.002
- David-John, B., Peacock, C., Zhang, T., Murdison, T. S., Benko, H., & Jonker, T. R. (2021a). Towards gaze-based prediction of the intent to interact in virtual reality. In Acm symposium on eye tracking research and applications (pp. 1–7).
- David-John, B., Peacock, C., Zhang, T., Murdison, T. S., Benko, H., & Jonker, T. R. (2021b, mai). Towards gaze-based prediction of the intent to interact in virtual reality. In ACM Symposium on Eye Tracking Research and Applications (pp. 1–7). Virtual Event Germany : ACM. doi: 10.1145/3448018.3458008
- De Beni, R., Pazzaglia, F., & Gardini, S. (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), 807–821.
- Delachambre, J., Wu, H.-Y., Vizcay, S., Di Meo, M., Lagniez, F., Morfin-Bourlat, C., ... Kornprobst, P. (2024). Amd journee : A patient co-designed vr experience to raise awareness towards the impact of amd on social interactions. In Proceedings of the 2024 acm international conference on interactive media experiences (pp. 17–29).
- Delaux, A., de Saint Aubert, J.-B., Ramanoël, S., Bécu, M., Gehrke, L., Klug, M., ... Arleo, A. (2021). Mobile brain/body imaging of landmark-based navigation with high-density eeg. European Journal of Neuroscience, 54(12), 8256–8282.
- Delrieu, T., Weistroffer, V., & Gazeau, J. P. (2020). Precise and realistic grasping and manipulation in virtual reality without force feedback. In 2020 IEEE conference on virtual reality and 3d user interfaces (vr) (pp. 266–274).
- Dewez, D., Hoyet, L., Lécuyer, A., & Argelaguet Sanz, F. (2021, mai). 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 (pp. 1–14). New York, NY, USA : Association for Computing Machinery. doi: 10.1145/3411764.3445379
- Dey, A., Chatburn, A., & Billinghamurst, M. (2019). Exploration of an eeg-based cognitively adaptive training system in virtual reality. In 2019 IEEE conference on virtual reality and 3d user interfaces (vr) (pp. 220–226).
- Diaper, D., & Stanton, N. (2003). The handbook of task analysis for human-computer interaction. USA : CRC press.
- Dickinson, P., Cardwell, A., Parke, A., Gerling, K., & Murray, J. (2021, mars). Diegetic Tool Management in a Virtual Reality Training Simulation. In 2021 IEEE Virtual Reality and 3D User Interfaces (VR) (pp. 131–139). Lisboa, Portugal : IEEE. (ISSN : 2642-5254) doi: 10.1109/VR50410.2021.00034
- Doumanis, I., & Economou, D. (2019). Using cognitive walkthrough and hybrid prototyping to gather user requirements in early design virtual reality prototypes. In Immersive learning research network : 5th international conference, ilrn 2019, london, uk, june 23–27, 2019, proceedings 5 (pp. 234–246).

- Dourish, P. (2004). *Where the action is : the foundations of embodied interaction*. USA : MIT press.
- Eklund, G. (1972). General features of vibration-induced effects on balance. *Upsala journal of medical sciences*, *77*(2), 112–124.
- Felnhofner, A., Kothgassner, O. D., Schmidt, M., Heinzle, A.-K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015, oct). Is virtual reality emotionally arousing ? investigating five emotion inducing virtual park scenarios. *Int. J. Hum.-Comput. Stud.*, *82*(C), 48–56. doi: 10.1016/j.ijhcs.2015.05.004
- Ferreira, J. M., Acuña, S. T., Dieste, O., Vegas, S., Santos, A., Rodríguez, F., & Juristo, N. (2020). Impact of usability mechanisms : An experiment on efficiency, effectiveness and user satisfaction. *Information and Software Technology*, *117*, 106195.
- Flotyński, J., & Sobociński, P. (2018). Logging interactions in explorable immersive vr/ar applications. In *2018 international conference on 3d immersion (ic3d)* (pp. 1–8). Brussels, Belgium : IEEE.
- Flotyński, J., & Walczak, K. (2017). Ontology-based representation and modelling of synthetic 3d content : A state-of-the-art review. In *Computer graphics forum* (Vol. 36(8), pp. 329–353). Poznań, Poland : Wiley Online Library.
- Fremont, D. J., Dreossi, T., Ghosh, S., Yue, X., Sangiovanni-Vincentelli, A. L., & Seshia, S. A. (2019). Scenic : a language for scenario specification and scene generation. In *Proceedings of the 40th acm sigplan conference on programming language design and implementation* (pp. 63–78). New York, NY, USA : Association for Computing Machinery.
- Gibson, W. H., Megaw, E. D., Young, M. S., & Lowe, E. (2006, janvier). A taxonomy of human communication errors and application to railway track maintenance. *Cognition, Technology & Work*, *8*(1), 57–66. doi: 10.1007/s10111-005-0020-x
- Górski, F. (2017). Building virtual reality applications for engineering with knowledge-based approach. *Management and Production Engineering Review*, *8*(4), 63–73. doi: 10.1515/mper-2017-0037
- Gottschalk, S., Yigitbas, E., Schmidt, E., & Engels, G. (2020). Model-based product configuration in augmented reality applications. In *International Conference on Human-Centred Software Engineering* (pp. 84–104). Eindhoven, The Netherlands : Springer.
- Grabarczyk, P., & Pokropski, M. (2016). Perception of affordances and experience of presence in virtual reality. *Avant. The Journal of the Philosophical-Interdisciplinary Vanguard*, *7*(2), 0.
- Guerra, E., Kurz, B., & Bräucker, J. (2022). An extension to the keystroke-level model for extended reality interactions. *Mensch und Computer 2022-Workshopband*.
- Hardie, P., Darley, A., Carroll, L., Redmond, C., Campbell, A., & Jarvis, S. (2020). Nursing & midwifery students' experience of immersive virtual reality storytelling : an evaluative study. *BMC nursing*, *19*(1), 78.
- Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, *25*(2), 91–97.
- Hawes, D., & Arya, A. (2021, mars). VR-based Student Priming to Reduce Anxiety and Increase Cognitive Bandwidth. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (pp. 245–254). Lisboa, Portugal : IEEE. (ISSN : 2642-5254) doi: 10.1109/VR50410.2021.00046
- Hayotte, M., Thérouanne, P., Gray, L., Corrion, K., & d'Arripe Longueville, F. (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), e16520.
- Hirzle, T. (2023). *Digital eye strain in virtual reality head-mounted displays : Properties, causes, solutions, and perspective* (Thèse de doctorat non publiée). Universität Ulm.
- Hix, D., & Hartson, H. R. (1993). *Developing user interfaces : ensuring usability through product & process*. New York : J. Wiley.
- Huang, D., Wang, X., Liu, J., Li, J., & Tang, W. (2022). Virtual reality safety training using deep eeg-net and physiology data. *The visual computer*, *38*(4), 1195–1207.

- Huang, S.-C., Lu, T.-W., Chen, H.-L., Wang, T.-M., & Chou, L.-S. (2008). Age and height effects on the center of mass and center of pressure inclination angles during obstacle-crossing. *Medical engineering & physics*, *30*(8), 968–975.
- Hughes, L. A., & Kanakri, S. (2020). Measuring behavior of low-vision populations using virtual reality. *Measuring Behavior 2020-21 Volume*.
- Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C. (2014, jul). Human3.6m : Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(7), 1325-1339.
- Javerliat, C., Villenave, S., Raimbaud, P., & Lavoué, G. (2024). Plume : Record, replay, analyze and share user behavior in 6dof xr experiences. *IEEE Transactions on Visualization and Computer Graphics*.
- Jeffrey J. Walline. (2001). Designing Clinical Research : an Epidemiologic Approach, 2nd Ed. *Optometry and Vision Science*, *78*(8).
- Jicol, C., Wan, C. H., Doling, B., Illingworth, C. H., Yoon, J., Headey, C., . . . O'Neill, E. (2021, mai). Effects of Emotion and Agency on Presence in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Yokohama Japan : ACM. doi: 10.1145/3411764.3445588
- Jonathan Lazar, H. H., Jinjuan Feng. (2010). *Research Methods in Human-Computer Interaction - 2nd Edition*. United States : Wiley.
- Keighrey, C., Flynn, R., Murray, S., & Murray, N. (2021). A physiology-based qoe comparison of interactive augmented reality, virtual reality and tablet-based applications. *IEEE Transactions on Multimedia*, *23*, 333-341. doi: 10.1109/TMM.2020.2982046
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire : An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203–220.
- Kim, U.-H., Park, J.-M., Song, T.-j., & Kim, J.-H. (2020). 3-d scene graph : A sparse and semantic representation of physical environments for intelligent agents. *IEEE Transactions on Cybernetics*, *50*(12), 4921-4933. doi: 10.1109/TCYB.2019.2931042
- Kim, Y. M., & Rhiu, I. (2024). Development of a virtual reality system usability questionnaire (vrsuq). *Applied Ergonomics*, *119*, 104319.
- Klostermann, A., & Moeinirad, S. (2020). Fewer fixations of longer duration? expert gaze behavior revisited. *German journal of exercise and sport research*, *50*(1), 146–161.
- Koo, T. K., & Li, M. Y. (2016, juin). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, *15*(2), 155–163. (Edition : 2016/03/31 Publisher : Elsevier) doi: 10.1016/j.jcm.2016.02.012
- Kroemer, O., Van Hoof, H., Neumann, G., & Peters, J. (2014). Learning to predict phases of manipulation tasks as hidden states. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 4009–4014).
- Kumar, M., Delaney, C., & Krusienski, D. (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* (pp. 396–401). New York, NY, United States : ACM.
- Kuno, S., Kawakita, T., Kawakami, O., Miyake, Y., & Watanabe, S. (1999). Postural adjustment response to depth direction moving patterns produced by virtual reality graphics. *The Japanese journal of physiology*, *49*(5), 417–424.
- Lang, P. (1980). Behavioral treatment and bio-behavioral assessment : Computer applications. *Technology in mental health care delivery systems*, *0*(0), 119–137.
- Lang, Y., Wei, L., Xu, F., Zhao, Y., & Yu, L.-F. (2018, mars). Synthesizing Personalized Training Programs for Improving Driving Habits via Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 297–304). Reutlingen : IEEE. doi: 10.1109/VR.2018.8448290

- Larusdottir, M., Thue, D., & Vilhjálmsón, H. H. (2019). User evaluations of virtually experiencing mount everest. In Human-centered software engineering : 7th ifip wg 13.2 international working conference, hcse 2018, sophia antipolis, france, september 3–5, 2018, revised selected papers 7 (pp. 294–307).
- Laubheimer, P. (2015). Preventing user errors : avoiding unconscious slips. Nielsen Norman Group.
- Lee, G., Lozano-Pérez, T., & Kaelbling, L. P. (2015). Hierarchical planning for multi-contact non-prehensile manipulation. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (p. 264–271). doi: 10.1109/IROS.2015.7353384
- Li, B. J., Bailenson, J. N., Pines, A., Greenleaf, W. J., & Williams, L. M. (2017, décembre). A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures. Frontiers in Psychology, 8, 2116. doi: 10.3389/fpsyg.2017.02116
- Li, S., & Zhang, X. (2017, août). Implicit Intention Communication in Human–Robot Interaction Through Visual Behavior Studies. IEEE Transactions on Human-Machine Systems, 47(4), 437–448. doi: 10.1109/THMS.2017.2647882
- Li, W., Talavera, J., Samayoa, A. G., Lien, J.-M., & Yu, L.-F. (2020, mars). Automatic Synthesis of Virtual Wheelchair Training Scenarios. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 539–547). (ISSN : 2642-5254) doi: 10.1109/VR46266.2020.00075
- Liu, Q., & Steed, A. (2021). Social virtual reality platform comparison and evaluation using a guided group walkthrough method. Frontiers in Virtual Reality, 2, 668181.
- Liu, Y., Wu, J., Wu, Z., Ritchie, D., Freeman, W. T., & Tenenbaum, J. B. (2019). Learning to describe scenes with programs. In International conference on learning representations. New Orleans, USA : ICLR.
- Lundin, J., Lundström, A., Gulliksen, J., Blendulf, J., Ejeby, K., Nyman, H., ... Hedman-Lagerlöf, E. (2022). Using 360-degree videos for virtual reality exposure in cbt for panic disorder with agoraphobia : a feasibility study. Behavioural and Cognitive Psychotherapy, 50(2), 158–170.
- Luong, T., Lecuyer, A., Martin, N., & Argelaguet, F. (2021). A survey on affective and cognitive vr. IEEE transactions on visualization and computer graphics, 28(12), 5154–5171.
- Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., ... Chen, S. H. A. (2021, feb). NeuroKit2 : A python toolbox for neurophysiological signal processing. Behavior Research Methods, 53(4), 1689–1696. doi: 10.3758/s13428-020-01516-y
- Mantiuk, R. K., Denes, G., Chapiro, A., Kaplanyan, A., Rufo, G., Bachy, R., ... Patney, A. (2021). Fovvideovdp : A visible difference predictor for wide field-of-view video. ACM Transactions on Graphics (TOG), 40(4), 1–19.
- Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., & Tunbridge, E. M. (2019). It feels real : physiological responses to a stressful virtual reality environment and its impact on working memory. Journal of Psychopharmacology, 33(10), 1264–1273.
- Matthews, M., Chowdhary, G., & Kision, E. (2017, août). Intent Communication between Autonomous Vehicles and Pedestrians. arXiv. (arXiv :1708.07123 [cs])
- Meijer, E., Beckman, B., & Bierman, G. (2006). Linq : reconciling object, relations and xml in the .net framework. In Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data (pp. 706–706). NY, New York : ACM.
- Memar, M., & Mokaribolhassan, A. (2021, janvier). Stress level classification using statistical analysis of skin conductance signal while driving. SN Applied Sciences, 3(1), 64. doi: 10.1007/s42452-020-04134-7
- Menin, A., Torchelsen, R., & Nedel, L. (2022). The effects of vr in training simulators : Exploring perception and knowledge gain. Computers & Graphics, 102, 402–412.
- Mevlevioğlu, D., Murphy, D., & Tabirca, S. (2021). Real-time Anxiety Prediction in Virtual Reality Exposure Therapy. New York City, 0(0), 5.
- Mezidi, M., & Guérin, C. (2019). Effect of body position and inclination in supine and prone position on respiratory mechanics in acute respiratory distress syndrome. Intensive Care Medicine, 45(2), 292–294.

- Mousas, C., Kao, D., Koilias, A., & Rekabdar, B. (2020). Real and virtual environment mismatching induces arousal and alters movement behavior. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 626–635). Atlanta, GA, USA : IEEE.
- Navarre, D., Palanque, P. A., Bastide, R., Schyn, A., Winckler, M., Nedel, L. P., & Freitas, C. M. D. S. (2005). A formal description of multimodal interaction techniques for immersive virtual reality applications. In M. F. Costabile & F. Paternò (Eds.), Human-computer interaction - INTERACT 2005, IFIP TC13 international conference (Vol. 3585, pp. 170–183). Rome, Italy : Springer. doi: 10.1007/11555261_17
- Nielsen, L. T., Møller, M. B., Hartmeyer, S. D., Ljung, T. C., Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Missing the point : an exploration of how to guide users' attention during cinematic virtual reality. In Proceedings of the 22nd ACM conference on virtual reality software and technology (pp. 229–232).
- Nordin Forsberg, B., Lundström, A., & Gulliksen, J. (2023). Vr for hr—a case study of human resource development professionals using virtual reality for social skills training in the workplace. In Ifip conference on human-computer interaction (pp. 231–251).
- Norman, D. (2008, novembre). Design as Communication.
- Norman, D. (2010, septembre). Systems Thinking : A Product Is More Than the Product.
- Norman, D. (2013). The design of everyday things. MIT Press.
- Oh, J., & Kim, G. J. (2019). The Effect of Onomatopoeia to Enhancing User Experience in Virtual Reality. In J. Y. Chen & G. Fragomeni (Eds.), Virtual, Augmented and Mixed Reality. Multimodal Interaction (pp. 143–152). Cham : Springer International Publishing. doi: 10.1007/978-3-030-21607-8_11
- Paek, T. (2003). Toward a taxonomy of communication errors. In Isca tutorial and research workshop on error handling in spoken dialogue systems.
- Palanque, P., & Martinie, C. (2016). Designing and assessing interactive systems using task models. In Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems (pp. 976–979).
- Palanque, P. A., Barboni, E., Martinie, C., Navarre, D., & Winckler, M. (2011). A model-based approach for supporting engineering usability evaluation of interaction techniques. In F. Paternò, K. Luyten, & F. Maurer (Eds.), Proceedings of the 3rd ACM SIGCHI symposium on engineering interactive computing system, EICS 2011, pisa, italy, june 13-16, 2011 (pp. 21–30). ACM. doi: 10.1145/1996461.1996490
- Pallavicini, F., Pepe, A., & Minissi, M. E. (2019). Gaming in virtual reality : What changes in terms of usability, emotional response and sense of presence compared to non-immersive video games ? Simulation & Gaming, 50(2), 136-159. Consulté sur <https://doi.org/10.1177/1046878119831420> doi: 10.1177/1046878119831420
- Peck, T. C., Doan, M., Bourne, K. A., & Good, J. J. (2018, apr). The effect of gender body-swap illusions on working memory and stereotype threat. IEEE Transactions on Visualization & Computer Graphics, 24(04), 1604-1612. doi: 10.1109/TVCG.2018.2793598
- Pidgeon, M., Kanwal, N., & Murray, N. (2022). Federated learning to understand human emotions via smart clothing : Research proposal. In Proceedings of the 13th ACM multimedia systems conference (p. 408–411). New York, NY, USA : Association for Computing Machinery. doi: 10.1145/3524273.3533936
- Pittarello, F., & De Faveri, A. (2006). Semantic description of 3D environments : a proposal based on web standards. In Proceedings of the eleventh international conference on 3D web technology (pp. 85–95). New York, NY, USA : ACM.
- Pöhlmann, K. M. T., Wilson, G., Li, G., McGill, M., & Brewster, S. A. (2024). From slow-mo to ludicrous speed : Comfortably manipulating the perception of linear in-car vr motion through vehicular translational gain and attenuation. In Proceedings of the CHI conference on human factors in computing systems (pp. 1–20).
- Qiu, Z., McGill, M., Pöhlmann, K. M. T., & Brewster, S. A. (2023). Using vr while travelling : Manipulating the orientation of planar 2d content in vr as an implicit visual cue for mitigating passenger motion sickness.

- Raffaillac, T., & Huot, S. (2019). Polyphony : Programming Interfaces and Interactions with the Entity-Component-System Model. *Proceedings of the ACM on Human-Computer Interaction*, *3*(EICS), 1–22.
- Rebenitsch, L., & Owen, C. (2021). Estimating cybersickness from virtual reality applications. *Virtual Reality*, *25*(1), 165–174.
- Ricca, A., Chellali, A., & Otrnane, S. (2021, mars). The influence of hand visualization in tool-based motor-skills training, a longitudinal study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (pp. 103–112). Lisboa, Portugal : IEEE. (ISSN : 2642-5254) doi: 10.1109/VR50410.2021.00031
- Rice, A. D., & Lartigue, J. W. (2014). Touch-level model (tlm) evolving klm-goms for touchscreen and mobile devices. In *Proceedings of the 2014 acm southeast regional conference* (pp. 1–6).
- Rossol, N., Cheng, I., Bischof, W. F., & Basu, A. (2011). A framework for adaptive training and games in virtual reality rehabilitation environments. In *proceedings of the 10th international conference on virtual reality continuum and its applications in industry* (pp. 343–346).
- Sagnier, C., Loup-Escande, E., Lourdeaux, D., Thouvenin, I., & Valléry, G. (2020). User acceptance of virtual reality : an extended technology acceptance model. *International Journal of Human-Computer Interaction*, *36*(11), 993–1007.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 71–78).
- Santoro, C. (2005). *A task model-based approach for design and evaluation of innovative user interfaces*. Presses univ. de Louvain.
- Sassatelli, L., Pinna-Déry, A.-M., Winckler, M., Dambra, S., Samela, G., Pighetti, R., & Aparicio-Pardo, R. (2018). Snap-changes : a dynamic editing strategy for directing viewer’s attention in streaming virtual reality videos. In *Proceedings of the 2018 international conference on advanced visual interfaces* (pp. 1–5).
- Sassatelli, L., Winckler, M., Fisichella, T., Dezarnaud, A., Lemaire, J., Aparicio-Pardo, R., & Trevisan, D. (2020). New interactive strategies for virtual reality streaming in degraded context of use. *Computers & Graphics*, *86*, 27–41.
- Saunier, J., Barange, M., Blandin, B., Querrec, R., & Taoum, J. (2016). Designing adaptable virtual reality learning environments. In *Proceedings of the 2016 virtual reality international conference* (pp. 1–4). New York, NY, USA : ACM.
- Sauro, J. (2009). Estimating productivity : Composite operators for keystroke level modeling. In *Human-computer interaction. new trends : 13th international conference, hci international 2009, san diego, ca, usa, july 19-24, 2009, proceedings, part i 13* (pp. 352–361).
- Schroeder, V. M., & Biggs, L. A. T. (2023). Fatigue and eye strain associated with virtual reality and augmented reality systems. *Adapting Virtual Reality and Augmented Reality Systems for Naval Aviation Training*, 104.
- Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. (2001). Research on presence in virtual reality : A survey. *Cyberpsychology & behavior*, *4*(2), 183–201.
- Seinfeld, S., Feuchtner, T., Pinzek, J., & Müller, J. (2020). *Impact of Information Placement and User Representations in VR on Performance and Embodiment (Vol. PP)*. remote : IEEE. (Journal Abbreviation : IEEE transactions on visualization and computer graphics Publication Title : IEEE transactions on visualization and computer graphics) doi: 10.1109/TVCG.2020.3021342
- Siu, K.-C., Best, B. J., Kim, J. W., Oleynikov, D., & Ritter, F. E. (2016). Adaptive virtual reality training to optimize military medical skills acquisition and retention. *Military medicine*, *181*(suppl_5), 214–220.
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British journal of psychology*, *109*(3), 431–433.
- Souza, C. S. D., & Leitão, C. F. (2009). *Semiotic Engineering Methods for Scientific Research in HCI*. Morgan & Claypool Publishers.
- Souza, V., Maciel, A., Nedel, L., & Kopper, R. (2021). Measuring presence in virtual environments : A survey. *ACM Computing Surveys (CSUR)*, *54*(8), 1–37.

- Speicher, M., Lewis, K., & Nebeling, M. (2021, juin). Designers, the Stage Is Yours ! Medium-Fidelity Prototyping of Augmented & Virtual Reality Interfaces with 360theater. Proceedings of the ACM on Human-Computer Interaction, 5, 205 :1–205 :25. doi: 10.1145/3461727
- Sutcliffe, A., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. Interacting with computers, 16(4), 831–849.
- Sveistrup, H. (2004). Motor rehabilitation using virtual reality. Journal of neuroengineering and rehabilitation, 1, 1–8.
- Tang, W., Wu, S., Vigier, T., & Da Silva, M. P. (2020, mai). Influence of Emotions on Eye Behavior in Omnidirectional Content. In 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX) (pp. 1–6). Athlone, Ireland : IEEE. doi: 10.1109/QoMEX48832.2020.9123126
- Tian, L., & Oviatt, S. (2021, juin). A Taxonomy of Social Errors in Human-Robot Interaction. ACM Transactions on Human-Robot Interaction, 10(2), 1–32.
- Toet, A., Heijn, F., Brouwer, A.-M., Mioch, T., & van Erp, J. B. F. (2020, septembre). An Immersive Self-Report Tool for the Affective Appraisal of 360° VR Videos. Frontiers in Virtual Reality, 1, 552587. doi: 10.3389/frvir.2020.552587
- Trösterer, S., Beck, E., Dalpiaz, F., Paja, E., Giorgini, P., & Tscheligi, M. (2012). Formative user-centered evaluation of security modeling : Results from a case study. International Journal of Secure Software Engineering (IJSSE), 3(1), 1–19.
- UCSF. (2021). Sample size calculators for designing clinical research.
- Ugwitz, P., Šašinková, A., Šašinka, Č., Stachoň, Z., & Juřík, V. (2021). Toggle toolkit : A tool for conducting experiments in unity virtual environments. Behavior research methods, 0(0), 1–11.
- Vanackén, L., Raymaekers, C., & Coninx, K. (2007). Introducing semantic information during conceptual modelling of interaction for virtual environments. In Proceedings of the 2007 workshop on multimodal interfaces in semantic interaction (pp. 17–24). New York, NY, USA : ACM.
- Van Gemert, T., Chew, S., Kalaitzoglou, Y., & Bergström, J. (2024). Doorways do not always cause forgetting : Studying the effect of locomotion technique and doorway visualization in virtual reality. In Proceedings of the chi conference on human factors in computing systems (pp. 1–13).
- Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology : extending the unified theory of acceptance and use of technology. MIS quarterly, 0(0), 157–178.
- Vergara, D., Rubio, M. P., & Lorenzo, M. (2017). On the design of virtual reality learning environments in engineering. Multimodal technologies and interaction, 1(2), 11.
- Vesga, J. B., Xu, X., & He, H. (2021, mars). The Effects of Cognitive Load on Engagement in a Virtual Reality Learning Environment. In 2021 IEEE Virtual Reality and 3D User Interfaces (VR) (pp. 645–652). Lisboa, Portugal : IEEE. (ISSN : 2642-5254) doi: 10.1109/VR50410.2021.00090
- Vignais, N., Bideau, B., Craig, C., Brault, S., Multon, F., & Kulpa, R. (2009, janvier). Virtual Environments for Sport Analysis : Perception-Action Coupling in Handball Goalkeeping. International Journal of Virtual Reality, 8(4), 43–48. doi: 10.20870/IJVR.2009.8.4.2748
- Vigo, M., Santoro, C., & Paternò, F. (2017). The usability of task modeling tools. In 2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 95–99).
- Villenave, S., Cabezas, J., Baert, P., Dupont, F., & Lavoué, G. (2022). Xrecho : A unity plug-in to record and visualize user behavior during xr sessions. In Proceedings of the 13th ACM Multimedia Systems Conference (pp. 341–346). Athlone, Ireland : ACM.
- Voigt-Antons, J.-N., Lehtonen, E., Palacios, A. P., Ali, D., Kojic, T., & Möller, S. (2020). Comparing Emotional States Induced by 360° Videos Via Head-Mounted Display and Computer Screen. In 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX) (pp. 1–6). doi: 10.1109/QoMEX48832.2020.9123125

- Vuarnesson, L., Zamplaras, D., Laroche, J., Dumit, J., Lutes, C., Bachrach, A., & Garnier, F. (2021). Shared Diminished Reality : A New VR Framework for the Study of Embodied Intersubjectivity. Frontiers in Virtual Reality, 2.
- Wagner, D., & Schmalstieg, D. (2006). Handheld augmented reality displays. In Ieee virtual reality conference (vr 2006) (pp. 321–321).
- Wang, K., Lin, Y.-A., Weissmann, B., Savva, M., Chang, A. X., & Ritchie, D. (2019). Planit : Planning and instantiating indoor scenes with relation graph and spatial prior networks. ACM Transactions on Graphics (TOG), 38(4), 1–15. (Publisher : ACM New York, NY, USA)
- Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related : a review. Frontiers in psychology, 10, 415654.
- Wen, J., & Gheisari, M. (2020). Using virtual reality to facilitate communication in the aec domain : a systematic review. Construction Innovation, 20(3), 509–542.
- Weser, V. U., Sieberer, J., Berry, J., & Tuakli-Wosornu, Y. (2024). Navigation in immersive virtual reality : a comparison of 1 : 1 walking to 1 : 1 wheeling. Virtual Reality, 28(1).
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments : A presence questionnaire. Presence, 7(3), 225–240.
- Wodarski, P., Jurkoć, J., Polechoński, J., Bieniek, A., Chrzan, M., Michnik, R., & Gzik, M. (2020). Assessment of gait stability and preferred walking speed in virtual reality. Acta of bioengineering and biomechanics, 22(1), 127–134.
- Wolfartsberger, J., & Niedermayr, D. (2020, mars). Authoring-by-Doing : Animating Work Instructions for Industrial Virtual Reality Learning Environments. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 173–176). doi: 10.1109/VRW50115.2020.00038
- Wu, E., Piekenbrock, M., Nakumura, T., & Koike, H. (2021, mai). SPinPong - Virtual Reality Table Tennis Skill Acquisition using Visual, Haptic and Temporal Cues. IEEE Transactions on Visualization and Computer Graphics, 27(5), 2566–2576. (Conference Name : IEEE Transactions on Visualization and Computer Graphics) doi: 10.1109/TVCG.2021.3067761
- Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., ... others (2021). A review on virtual reality skill training applications. Frontiers in Virtual Reality, 2, 645153.
- Xue, T., Ali, A. E., Ding, G., & Cesar, P. (2021, mai). Investigating the Relationship between Momentary Emotion Self-reports and Head and Eye Movements in HMD-based 360° VR Video Watching. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–8). Yokohama Japan : ACM. doi: 10.1145/3411763.3451627
- Xue, T., El Ali, A., Zhang, T., Ding, G., & Cesar, P. (2021). CEAP-360VR : A Continuous Physiological and Behavioral Emotion Annotation Dataset for 360 VR Videos. IEEE Transactions on Multimedia, 1–1. doi: 10.1109/TMM.2021.3124080
- Yang, B. (2023). Virtual reality and augmented reality for immersive learning : A framework of education environment design. International Journal of Emerging Technologies in Learning (IJET), 18(20), 23–36.
- Yu, E., Chevalier, F., Singh, K., & Bousseau, A. (2024). 3d-layers : Bringing layer-based color editing to vr painting. ACM Transactions on Graphics (TOG), 43(4), 1–15.
- Zander, T., Rohlmann, A., Calisse, J., & Bergmann, G. (2001). Estimation of muscle forces in the lumbar spine during upper-body inclination. Clinical Biomechanics, 16, S73–S80.
- Zhang, S., Ma, Q., Zhang, Y., Qian, Z., Kwon, T., Pollefeys, M., ... Tang, S. (2022). Egobody : Human body shape and motion of interacting people from head-mounted devices. In European conference on computer vision (pp. 180–200).
- Zhang, X., & Simeone, A. L. (2022). Using the think aloud protocol in an immersive virtual reality evaluation of a virtual twin. In Proceedings of the 2022 acm symposium on spatial user interaction (pp. 1–8).

Zhao, J., & Klippel, A. (2019). Scale-unexplored opportunities for immersive technologies in place-based learning. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 155–162).

Zhao, Y., Sarkar, A., Elmhadi, L., Karray, M. H., Fillatreau, P., & Archimède, B. (2024). An ontology of 3d environment where a simulated manipulation task takes place (envon). *Semantic Web*(Preprint), 1–28.

Zheng, Y., Yang, Y., Mo, K., Li, J., Yu, T., Liu, Y., . . . Guibas, L. J. (2022). Gimo : Gaze-informed human motion prediction in context. *arXiv preprint arXiv :2204.09443*.

Zhou, X., Teng, F., Du, X., Li, J., Jin, M., & Xue, C. (2022, jul). H-goms : A model for evaluating a virtual-hand interaction system in virtual environments. *Virtual Real.*, *27*(2), 497–522. doi: 10.1007/s10055-022-00674-y

List of Figures

1.1	The seven stage of action theory by D. Norman, composed of seven steps : (1) defining the goal to achieve, (2) defining the intention to interact with the system to achieve the goal, (3) defining the detailed plan to achieve the interaction, (4) executing the detailed plan, (5) perceiving the state of the system after executing the detailed plan, (6) interpreting the observed state of the system, (7) comparing the observed changes to the initial goal. . . .	3
1.2	We build systems to support designers of embodied experiences across three components. First, we propose systems to help with the design and observation of embodied experiences. Using this first system, we then construct frameworks to collect and analyze data to enable a deeper comprehension of embodied interactions. Finally, we introduce a methodology to understand designer-user communication issues and ways to guide designers and users in the design and usage phases of an embodied experience to avoid said issues. All systems are conceived using a user-centered approach, with user studies involving designers and users of embodied experiences to evaluate the system’s usability.	3
1.3	This model represents the designer-system-user relationship when creating and using embodied experiences in VR. The left side represents the design process between the designer and the system, while the right side represents the usage process between the user and the system. We specify three components to address to avoid the creation of communication gulfs between the actors : the ontology, intersubjectivity, and intentionality.	5
3.1	3-Component framework for designing <i>GUTasks</i> in 3D virtual environments based on embodiment theory : the GUsT-3D vocabulary which is the full representation of the scene <i>ontology</i> , the definition and management of <i>GUTasks</i> communicated between the Designer and User for <i>intersubjectivity</i> , and the logging and query of user experience to allow Designers to analyze user <i>intentionality</i>	22
3.2	4-Step workflow to create <i>Guided User Tasks</i> in 3D environments involves : design and annotation of the 3D scene, defining <i>GUTasks</i> , running the task scenario, and post-scenario analysis of the user experience. Each step of the workflow contains a number of sub-tasks that the designer may need to do.	23
3.3	A partial scene graph generated from the scene to show the relations between various entities in the scene, including navigable spaces and entryways, movable props, objects that serve as support or containers, and other elements such as ground and wall entities. The hard edges denote location relations, whereas dashed edges of different colours indicate interactive and navigation relations between entities.	25
3.4	The syntax of the query is composed of keywords from the vocabulary defined in GUsT-3D including a minimum of three keywords – the reference point (usually the user camera), query head, and target – and potentially two additional keywords that can further limit the search to entities that fulfil specific location or interaction constraints. The result of a runtime query can be seen in Figure 3.8.	28
3.5	The combined framework and task model to indicate how each element of the framework intervenes on the workflow of creating a 3D embodied experience. The arrows indicate the information flow on the three embodied experiences aspects : ontology (red), intersubjectivity (blue), and intentionality (green)	28
3.6	Schema of the layers definition interface, which allows the scene designer to add and edit the units, constraints, and layers either through point-and-click or importing a configuration file. The actual interface we designed based on the schema is shown on the two sides. . . .	29

3.7	This interface allows designers to define <i>GUTasks</i> , add and edit task constraints, view the list of existing tasks. The tasks can be easily loaded and exported in the GUT-3D representation for reuse in various scenarios.	30
3.8	Here we show the window to construct a simple query using the syntax from Figure 3.4 that counts the number of props within a near distance (3 meters) of the player camera. The two views show the highlighting in the scene from the player viewpoint (left) and the Unity scene view from above (right).	30
3.9	Window for querying the logs of a previously run scenario, including loading a log file, a slider to show all the state changes recorded to an entity, and the descriptive text on the state at a time point. A generated scene graph of the user’s interaction state at a specific time point is shown on the right, indicating, for example, visibility of entities and objects being held.	30
3.10	Given the <i>GUTask</i> of placing a stack of books on the bookshelf in the garage, this figure depicts the interaction and navigation paths of the user for two constraints : (1) locating and retrieving the stack of books, and (2) navigating to the garage, and placing the stack of books on the bookshelf. We use the same color and shape representation of entities as Figure 3.3. It is worth noting that the interaction with the stack of books in (1) requires evaluating an attention constraint – that the object has to be in the user’s visual field.	31
3.11	The second case study features a basic road crossing scene in an open space. The scene has four navigable spaces : a road and a crossing that overlap, and two sidewalks. From left to right, this figure shows the top view, the navigable spaces, and a viewpoint from the user camera.	32
3.12	For the three steps in the workflow, the responses of each participant on the usability, learning curve, efficacy, and flexibility of the implementation. The step “Running the Scenario” does not have these metrics, since it is mainly composed of the automatic processes that launch the <i>GUTasks</i> and record logs of the experience. The missing responses indicate where participants responded “I don’t know”.	34
3.13	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.	37
3.14	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.	40
3.15	Modeling 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.	41
3.16	(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)	42
3.17	Top-down view of the two different type of environments	43
3.18	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.	43
3.19	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 randomized. The entire study lasted roughly two hours.	44

3.20 (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 center 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. 44

4.1 Shimmer3 GSR+ used to record EDA and optical pulse. Gray wires connect the EDA sensor, white wire connects the pulse sensor. 53

4.2 Folder structure of the dataset with main files. 54

4.3 EDA signal recorded for user 03 while watching video 73. The three graphs from the top show the raw EDA data and the tonic component, the phasic component and the SCR (absolute value of phasic first derivative). 55

4.4 Arousal and valence ratings by users for each videos. The green dotted line corresponds to the mean and the orange solid line to the median. 55

4.5 Dots colors code for video ID (legend on the right). Left : Scatter plot of SCR_v against GA_v . The shaded area represents the 95% CI of the linear regressor (solid blue line). Right : Scatter plot of $cSCR_{u,v}$ against $cGA_{u,v}$ 56

4.6 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. 60

4.7 (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. 61

4.8 The EDA of participant 1 during the scenario 2 in RW+NV condition, honked by a car while jaywalking during a red traffic light. 62

4.9 Our task-based methodology comprises three steps : (1) synchronise behavioural data with context data, (2) use task model to define performance baselines, and (3) characterise user behaviour based on their performances using multivariate data. 62

4.10 An example of the raw multivariate data (i.e., skin conductance (EDA), point of regard positions with x, y, z axis (POR), chest rotation value) synchronised with the context for one user in a single scenario. During this scenario, the users carried out four tasks : (1) Go outside their house, (2) Cross the road safely, (3) Take the box, (4) Cross the road safely to come back home 64

4.11 A simplified task model for three high level tasks : (a) “Go outside your house”, (b) “Take the box”, (c) “Cross the road safely”, to properly achieve the task, the user is expected to perform the right interactions at the right time, and in the case of crossing the road, also look at traffic light and cars, and cross when the light is green and the cars are stopped. Rectangles indicate tasks and ovals indicate object states. 66

4.12 Time taken by all 40 participants (sorted by study completion order) to finish the tasks of the scenario #6 (multiple interactions with two cars). The baseline column on the far right shows the baseline value for each individual task stacked. Tasks where the user surpasses this baseline are outlined in black. 67

4.13 Median COPI of all users for each scenario of the experiment based on the efficiency component of performance. The red boxes represent the COPI values for TM_e , and the green boxes represent the COPI values for non TM_e 71

4.14	Visualization of the median COPI value of each scenarios with road crossing for the users U511 (top), U495 (middle), U109 (bottom) with significant negative COPI variation in TM_a tasks. Red dots represent the median COPI value during road crossing task for a scenario with attention misstep, green dots represent the median COPI values during road crossing task for a scenario without attention misstep.	72
5.1	Our model was inspired by Norman to characterize the designer-system-player communication gulfs, allowing the definition of a taxonomy of eight issues sources of gulfs happening during the design, usage, and interpretation phases.	76
5.2	The study protocol consists of three stages : (1) design phase, during which participants create and evaluate their own experience ; (2) test phase, during which each participants test two previously created experiences ; (3) and the post-test phase during which each participant watch record of two persons playing the created scenario. The entire study lasted roughly two hours.	79
5.3	The interior (top) and exterior (bottom) 3D environments were assigned to designers for the creation of their scenarios. Both environments offered a 5-meter by 5-meter design space. Objects can be moved and renamed to fit the designer's intention. Participants assigned to the interior environments had to include in their scenario the following objects : a golf club, a golf ball, and any fallen objects. Participants assigned to the exterior environment had to include in their scenario the following objects : a key, a trashcan, and any sweet object.	81
5.4	The designer D4 created a scenario, to simulate an office worker in burnout, where the player has to play golf on a precious trophy. The player D4U2 did the tasks properly, however, the player didn't understand the tasks the way the designer expected, interpreting this set of tasks as a desk tidying scenario.	83
5.5	D5 created a scenario, where players perform a number of interactions, one of which is to open a door with an access card. However, the designer forgot to rename the target object (the phone) as an access card. D5U1 completed the tasks, but did not understand the logic of opening a door with a phone, interpreting this interacting as humorous.	84
5.6	The designer D10 created a scenario where players have to take a book and place it on the ground. However, the designer did not specify which book exactly to take and didn't notice this issue in the test phase. The player D10U1 did understand the meaning of the task, however, due to the lack of detail on the book to take in a scene full of books, the player took some time to find the right one.	85

Analyzing and understanding embodied interactions in virtual reality systems

Florent ROBERT

Summary

Embodied interactions allow the creation of realistic and immersive 3D storytelling scenarios by giving users the possibility to engage with virtual environments through the use of their physical bodies, including gestures, posture, and locomotion. Nonetheless, new methodologies to capture and analyze experiences are required for the case of embodied VR experiences, because of the use of the body as a communication device, stimulating the brain and enabling reactions differently than traditional media. Communication problems are also likely to arise between the experience sought by the designer and the experience lived and interpreted by the user, due to the variety of scenarios and interaction modalities that make up a realistic embodied scenario. To address these challenges, this thesis contributes to the observation, analysis, and comprehension of the lived embodied experiences through the conception of frameworks, models, and methodologies, a subject that has not yet been fully explored in the literature applied to VR.

Paul Dourish's theory of embodiment highlights three key elements to take into account for an efficient comprehension of embodied interaction. Inspired by this theory, this thesis is composed of three components : (1) The ontology component, presenting an embodied experience creation framework, to address the creation, the management, and the observation of experiences using embodied interactions in immersive environments ; (2) the intentionality component, presenting an embodied interaction analysis methodology, to address the comprehension of the embodied interactions through the analysis of the correlations between the user behavioral data and the system contextual data ; and (3) the intersubjectivity component, presenting a designer-system-user communication model in embodied context, to address the communication issues existing between a designer building an embodied experience and a user living this experience. Each component represents a chapter of this thesis, taking a user-centered approach to support designers in their understanding of the embodied experience lived in an immersive context by proposing tools inspired by existing works in human-computer interactions (HCI). To validate our tools, three user studies were conducted. The first study evaluates the system implementation of the proposed creation framework, with participants who possess expertise in 3D experience development. The second study enables embodied interaction observation and analysis by collecting participant behavioral metrics (e.g., body motion, skin conductance) while they perform embodied interactions in VR. The third investigates the communication gulf existing between the designer and the user of an embodied experience by observing and classifying the issues happening during the design and the usage phases of the experience. This thesis can support research that aims at a better understanding of human behavior or the improvement of user experience (e.g., health, HCI) by submitting tools that can support designers in acquiring a better understanding of how embodied experiences are lived in VR.

Keywords : Embodied interactions, Virtual reality, User behavior, Ontology, Intersubjectivity, Intentionality