UNIVERSITÉ DE RENNES 1

# THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1

ECOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies de l'Information et de la Communication*
Spécialité : *Informatique*

Par
**Florian Berton**

# Immersive Virtual Crowds: Evaluation of Pedestrian Behaviours in Virtual Reality

**Thèse présentée et soutenue à Irisa, le 14/12/2020**
**Unité de recherche : Inria**
**Thèse N° : (8)**

**Rapporteurs avant soutenance** :

| | |
|---|---|
| PELECHANO Nuria | Professeure associée Université Polytechnique de Catalogne (UPC) |
| BOULIC Ronan | Chercheur Senior à l'Ecole Polytechnique Fédérale de Lausanne (EPFL) |

**Composition du Jury :**

| | | |
|---|---|---|
| Président : | MARCHAND Eric | Professeur à l'Université Rennes 1 |
| Examinateurs : | O'SULLIVAN Carol | Professeur au Trinity College Dublin |
| | WARREN William | Professeur à l'Université Brown |
| | PELECHANO Nuria | Professeur associée Université Polytechnique de Catalogne (UPC) |
| | BOULIC Ronan | Chercheur Senior à l'Ecole Polytechnique Fédérale de Lausanne (EPFL) |
| Dir. de thèse : | PETTRE Julien | Directeur de recherche à l'INRIA Rennes |
| Co-encadrants de thèse : | OLIVIER Anne-Hélène | Maître de conférence de l'Université Rennes 2 |
| | HOYET Ludovic | Chargé de recherche à l'INRIA Rennes |

# Contents

**Contents** iii

**1 Introduction** 1
- 1.1 Context . . . 1
- 1.2 Our Approach . . . 3
- 1.3 Contributions . . . 4
- 1.4 Overview . . . 6

**2 State of the art** 7
- 2.1 Crowd Simulation . . . 8
  - 2.1.1 Approaches . . . 9
  - 2.1.2 Local Interaction Models . . . 11
- 2.2 Pedestrian Interactions in Real Environment . . . 15
  - 2.2.1 Trajectory Analysis . . . 15
  - 2.2.2 Vision and Locomotion . . . 21
- 2.3 Virtual Reality . . . 29
  - 2.3.1 Technology . . . 29
  - 2.3.2 Evaluation of Virtual Reality . . . 31
  - 2.3.3 Experiment . . . 34
- 2.4 Conclusion . . . 40

**3 Experimental Tools** 43
- 3.1 CrowdMP . . . 44
  - 3.1.1 Experimental Procedures . . . 45
  - 3.1.2 CrowdMP Architecture . . . 46
  - 3.1.3 Virtual Reality Displays . . . 49
- 3.2 Tobii Tools . . . 50
  - 3.2.1 TobiiController . . . 50
  - 3.2.2 Tobii Python Tool . . . 53

**4 Influence of Virtual Reality Setup on Gaze Activity during Collision Avoidance** **55**
4.1 Methods . . . 57
4.1.1 Participants . . . 58
4.1.2 Task . . . 58
4.1.3 Environment . . . 58
4.1.4 Experimental Design . . . 59
4.2 Analysis . . . 61
4.2.1 Collected Data . . . 61
4.2.2 Time Variables . . . 62
4.2.3 Collision Avoidance Trials . . . 62
4.2.4 Kinematics of the Collision Avoidance Task . . . 63
4.2.5 Gaze Behaviour . . . 64
4.2.6 Statistics . . . 65
4.3 Results . . . 66
4.3.1 Locomotion . . . 66
4.3.2 Gaze Behaviour . . . 67
4.4 Discussion . . . 70
4.4.1 Collision Avoidance Behaviour . . . 70
4.4.2 Fixations and Gaze Allocation . . . 71
4.4.3 Gaze and Head Angles . . . 72
4.4.4 Limitations . . . 73
4.5 Conclusion . . . 74

**5 Navigation in Crowds: Influence of Virtual Reality and Crowd Density** **77**
5.1 Introduction . . . 78
5.2 Overview . . . 79
5.2.1 Apparatus & Task . . . 79
5.2.2 Participants . . . 80
5.2.3 Analysis . . . 81
5.3 Real vs. Virtual Validation . . . 84
5.3.1 Procedure . . . 85
5.3.2 Analysis & Results . . . 86
5.3.3 Discussion . . . 89
5.4 Effect of Crowd Density on Eye-gaze Behaviour . . . 90
5.4.1 Procedure . . . 90
5.4.2 Analysis & Results . . . 91

5.4.3 Discussion . . . 93
5.5 General Discussion . . . 94
5.5.1 Crowd Simulation . . . 94
5.5.2 Limitations . . . 94
5.6 Conclusion . . . 96

**6 Impact of Haptic Rendering of Collisions During Navigation through a Virtual Crowd 97**
6.1 Virtual Realty and Collision Rendering . . . 99
6.2 Experimental Overview . . . 101
6.2.1 Materials & Methods . . . 101
6.2.2 Environment & Task . . . 104
6.2.3 Protocol . . . 105
6.2.4 Participants . . . 106
6.2.5 Hypotheses . . . 106
6.3 Analysis . . . 107
6.3.1 Collected Data . . . 107
6.3.2 Trajectories . . . 107
6.3.3 Body Motions . . . 108
6.3.4 Collisions . . . 110
6.3.5 Presence and Embodiment . . . 110
6.3.6 Statistical Analyses . . . 113
6.4 Results . . . 113
6.4.1 Trajectory Analysis . . . 113
6.4.2 Body Motion . . . 114
6.4.3 Collisions . . . 115
6.4.4 Presence and Embodiment . . . 117
6.5 Discussion . . . 119
6.5.1 Trajectories . . . 119
6.5.2 Avoidance Behaviour . . . 119
6.5.3 Haptic Rendering After-effects . . . 120
6.5.4 Embodiment & Presence . . . 121
6.5.5 Limitations . . . 121
6.6 Conclusion . . . 122

**7 Conclusion and Perspectives 125**
7.1 Virtual Reality as a Relevant Tool to Study Human Behaviour . . . 126
7.1.1 Experiments Lessons . . . 126

7.1.2 Recommendations to Design Virtual Experiments . . . . . 128
7.2 Future Works . . . . . 131
7.2.1 Short Term Perspectives . . . . . 131
7.2.2 Long Term Perspectives . . . . . 132

**Acronyms** **137**

**List of Figures** **139**

**List of Tables** **147**

**Bibliography** **149**

Chapter

# 1 Introduction

## Contents

1.1 Context . . . . . 1
1.2 Our Approach . . . . . 3
1.3 Contributions . . . . . 4
1.4 Overview . . . . . 6

## 1.1 Context

During this thesis, we were interested in acquiring new insights about collective behaviour in the case of a moving crowd. To this end, we have worked on the design and the evaluation of an experimental platform in virtual reality to study pedestrians' behaviour, which is at the core of collective behaviour. Our research applies to the field of crowd simulation, which aims at computing the movement of a crowd of individuals to reproduce their real behaviour. Such simulations are of interest for several fields of applications, such as entertainment (Figure 1.1-Left) or crowd management (Figure 1.1-Right), each of them having their own performance standards. Indeed, agents in a virtual crowd in a video game may be required to interact with the player. It is therefore necessary to design a simulation that enables real-time interactivity, thus inducing high performance in computation time. This is not the case for cinema, where one may be asked to simulate a

crowd of several thousand agents with highly visually realistic movements and the possibility to edit all these animations. Performance is also of course implies in such cases as it directly affects the budget, but visual quality remains the prime element. In the case of evacuation scenario in the frame of event planning, or architecture, it is important to generate realistic predictive simulations that can be adapted to any situation, which requires evaluating all possible effects of the parameters on the crowd simulator model. This category of model is used to choose the best locations for exits or to evaluate various safety criteria such as the evacuation time or the minimal physical distancing maintained between people in the context of Covid (Figure 1.1-Right).

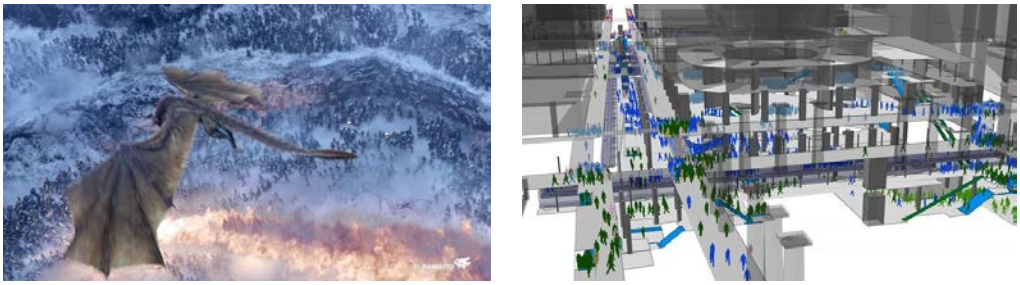


**Figure 1.1** – Left: battle scene from the Game of Thrones series using Goalem Software's crowd simulator, Right: MassMotion crowd simulator developed by Oasys and used to test the ability of the Arup Liverpool office to comply with physical distancing measures during Covid.

To satisfy the requirements of these applications and comply with these performance standards, several simulation methods have been developed. They can be divided into three main categories:

1. *Flow-based* methods aim at simulating very large crowds of people by considering the crowd as a whole.

2. *Data-driven* methods aim at mimicking crowd movements from real data.

3. *Agent-based* methods simulate the behaviour of each virtual agent within the crowd. Global crowd motion then emerges from the combination of local interactions.

In this thesis, we aim at improving *agent-based methods* by providing new methods to analyse pedestrians' behaviour at the local scale. Local interactions when walking within a crowd are diverse and include following tasks, walking in a group, reaching someone as well as avoiding a collision. Our work specifically focused on the collision avoidance task during goal directed locomotion, where an

individual is moving towards a goal while avoiding any collision with other pedestrians. Several models of such an interaction have been proposed in the literature and are based on Physics principles (e.g., repulsive forces influenced by the interpersonal distance between walkers) or considered relative speeds of walkers. The complexity in such modeling is to understand which agent within a crowd influences the motion of the walker. This is named *"interaction neighbourhood"*. It goes without saying that when we walk in the street, our movements and actions do not take into account all the people in the street but a subset around us. The definition of such an interaction neighbourhood has received little attention in the literature and is a challenging task. Indeed, understanding and modelling this neighbourhood is equivalent to inverting the injective process by which multiple sources of interactions combine to influence a single trajectory of lower dimension. However, we believe that the improvement of current agent-based crowd simulation models relies on a better definition of this neighbourhood. It requires then a deep analysis of how a human behaves at the local scale when navigating in populated environments.

## 1.2 Our Approach

This thesis aims to improve current crowd simulation models by understanding how pedestrian motion is controlled. In that context, our approach relies on the ecological theory of visual perception developed by Gibson [1958]. According to Gibson, *"We must perceive in order to move, but we must also move in order to perceive"*. Gibson considers the system agent-environment, where interactions can be described as a perception-action loop, illustrated Figure 1.2. In this approach, perception is considered as direct, meaning that the agent has an immediate perception of high level variables, directly available in the sensory flow. The agent, therefore, perceives the environment through their perceptual systems (vision, touch, hearing ...) and control their action accordingly which will in turn modify the perceived environment. This approach has been used as the theoretical basis for other researchers, such as Warren [1998, 2006] who introduces in this perception-action loop the notion of behaviour law to describe the interactions between the environment and the agent.

Following this theory, we believe that in order to understand pedestrians interactions in a complex environment, including their interaction neighbourhood, we have to consider both motion, i.e., the pedestrian kinematics, and perception. In particular, we would like to analyse pedestrians' gaze activity, since it has been shown that vision is the main perceptual system used during locomotion to

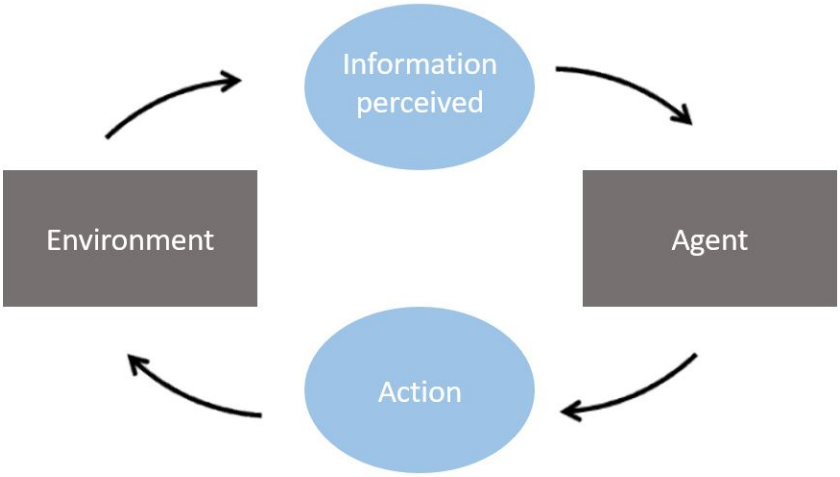



**Figure 1.2** – The Perception and Action loop proposed by Gibson [1958].

take information about the physical characteristics of the environment but also about the (relative) position and motion of the observer and the elements of the environment [Warren, 1998, Patla, 1997].

To this end, we designed our experiments in Virtual Reality (VR), as it provides strong control over the experimental conditions which is very challenging in a real environment full of people. Interestingly, it enables to reproduce the exact same visual stimuli for several trials of the same participant and across participants, which is an essential element for experimental studies. This control is crucial in our approach, as a difference in the visual stimulus can have a significant impact on gaze activity. Furthermore, VR also provides direct access to the environment perceived by the participant, so that we know exactly what they looked at based on their gaze's coordinates and without making additional computations. However, since previous works have shown that modification of perception and action can occur in VR, it is necessary to evaluate the effect of VR on gaze activity for the study of interaction between pedestrians. Indeed, differences in human behaviour between a real and virtual environment must be identified in order to be able to transfer results observed in virtual reality to situations in real environments.

## 1.3 Contributions

This thesis proposed three main contributions based on three experimental studies that rely on the development of two technical platforms presented in Chapter 3.

In our first contribution, we focused our research on the impact of virtual reality on gaze activity in a pairwise collision avoidance task. For this purpose, we carried out an experiment in a real and a virtual environment with four different kinds of VR setup. During this experiment, participants had to walk toward a target while avoiding another pedestrian. We chose this situation as it represents one of the majority interactions when walking in a crowd. Our results showed that gaze activity was qualitatively similar for all conditions with some quantitative differences. In particular, there were more head rotations in VR and a larger amplitude for gaze angle. In conclusion, VR seems to be adequate as a tool to study gaze activity during a pedestrian interaction. This work was presented at IEEE VR 2019 conference [Berton et al., 2019].

We extended this contribution in a second experiment where we were interested in the analysis of gaze activity during navigation in a crowded street. However, as this situation is more complex than a single pedestrian interaction, we first validated that our previous results were similar for this situation, by comparing real and virtual crowded situations. We then investigated the impact of crowd density on gaze activity in order to gain new insights on the neighbourhood of interaction. Our results suggest that as the crowd density increases, the exploration of the environment by the eyes becomes narrower without changing its frequency and that the gaze is concentrated on the pedestrians in front of the participants. Such results may indicate, for instance, to consider a constant number of pedestrians in the *"interaction neighbourhood"* regardless of the density of the crowd. This work was presented at IEEE VR 2020 conference [Berton et al., 2020].

However, when walking in a dense crowd, it is usual to collide with other pedestrians, which can be challenging to reproduce in VR. In a third contribution, we have investigated the effect of simulating such physical contacts on human behaviour in VR. To this end, we conducted an experiment where participants had to navigate through a dense virtual crowd with or without haptic rendering of collisions. Our results show that haptic rendering of collisions does not alter the trajectories taken by the participants, which is consistent with previous studies [Warren, 1998, Patla, 1997] indicating that vision is the main perceptual system used when walking. However, the introduction of haptic rendering of collision altered participants' local movements with an increased shoulder rotation as well as a decreased walking speed in order to move through the dense crowd. These results suggest continuing to use haptic rendering when performing VR experiments in dense crowds. This work has been submitted to the journal TVGC and is currently under minor revisions.

In addition, I have also been involved in several collaborative works on topics

related to crowd simulation and the study of human behaviour when interacting with virtual pedestrians. A study [Duverne et al., 2020] was performed to understand the influence of the social environment on human behaviour. In particular, we explored the effect of the transgression of personal space at a train station and in a sports fan zone, which was conducted in a real and virtual environment. In this work, I was involved in the analysis of VR data and the writing of the experimental design. Our results suggest that proxemics norms vary according to the subjective relationship of the individual to the social settings in real environments. However, while we were able to show that social norms still exist in VR, our results did not show a main effect of the social settings on participants' sensitivity to the transgression of proxemics norms. Another collaboration [van Toll et al., 2020] focused on crowd simulation where we proposed a method to reproduce several agent-based models using a cost function optimization. My contribution to this collaboration was mainly on the representation of the results. Finally, I actively participated in the development of a software (Chaos[1]) to visualize crowd movements based on numerical trajectories. This software also enables to record several kinds of data that can be used as a database for deep learning algorithms.

## 1.4 Overview

This thesis is structured as follows: Chapter 2 presents the state of the art in the different fields related to this thesis. In particular, it focuses on the presentation of different approaches to crowd simulation, as well as studies on trajectories and gaze activity during interactions between pedestrians in real and virtual environments. Chapter 3 describes the different platforms created and used to design our experiments. Chapter 4 introduces a first study on the impact of virtual reality on gaze activity during a collision avoidance task between two pedestrians. Chapter 5 presents a second study on the impact of virtual reality on gaze activity, in a more complex scenario, when navigating in a crowd. This study also explores the effect of crowd density. Chapter 6 investigates haptic rendering of collision when navigating in a virtual dense crowd. Finally, Chapter 7 presents a general conclusion of these contributions and future perspectives.

[1] https://project.inria.fr/crowdscience/project/ocsr/chaos/

Chapter

# 2 State of the Art

## Contents


**2.1 Crowd Simulation . . . . . 8**

2.1.1 Approaches . . . . . 9

2.1.2 Local Interaction Models . . . . . 11

**2.2 Pedestrian Interactions in Real Environment . . . . . 15**

2.2.1 Trajectory Analysis . . . . . 15

2.2.2 Vision and Locomotion . . . . . 21

**2.3 Virtual Reality . . . . . 29**

2.3.1 Technology . . . . . 29

2.3.2 Evaluation of Virtual Reality . . . . . 31

2.3.3 Experiment . . . . . 34

**2.4 Conclusion . . . . . 40**


In this thesis we are interested in the analysis of human behaviour when navigating in a virtual crowd, in order to subsequently improve crowd simulators. This thesis is thus at the intersection of three main domains which are: crowd simulation, human behaviour studies, and virtual reality, as shown in the Figure 2.1. This chapter first introduces crowd simulation, detailing the different approaches and models. Then, it presents the different studies on human behaviour when

interacting with pedestrians in a real environment. This presentation focuses especially on studies on trajectory analysis and studies on vision and walking, as these are the main research topics of this thesis. Finally, as virtual reality (VR) is a major tool in this thesis, the last section presents some of the biases induced by VR and studies on human behaviour.

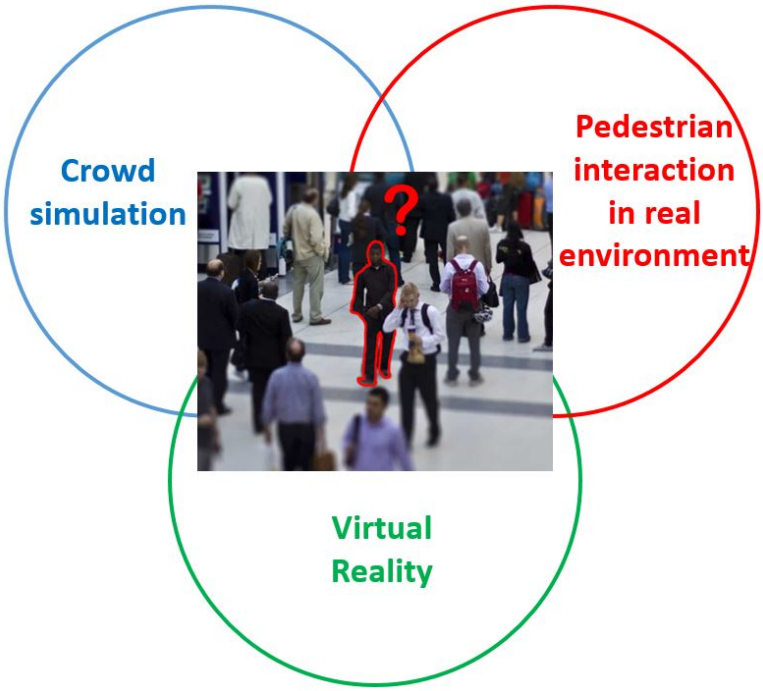



**Figure 2.1** – This thesis, on the analysis of human behaviour when navigating in a crowd, is at the intersection of three main domains: **Crowd Simulation**, **Human behaviour Analysis** and **Virtual Reality**

## 2.1 Crowd Simulation

This section introduces crowd simulation, and to do so it is first necessary to define the concept of a *"crowd"*. There is no exact definition, however, many works [Duives et al., 2013, Challenger et al., 2009, Hoogendoorn and Bovy, 2004, Wijermans, 2011] define a crowd as several pedestrians moving in the same place and at the same time. Furthermore, the number of pedestrians to define a crowd is quite variable, ranging from more than 2 [Challenger et al., 2009] to more than 100 [Duives et al., 2013]. Then a second important aspect about crowds is the *"collective behaviour"*. This notion is introduced by Sumter *et al.* [2012] and is used to describe the emergence of group level patterns from local interactions

of the individuals. Therefore, we can define crowd simulation as the creation of mathematical models to simulate collective behaviour in a virtual crowd. This topic has been of interest for several years, and over these years a wide variety of algorithms has been implemented. In Section 2.1.1, we present the main approaches that constitute the crowd simulation field, then we detail some models that focus on the simulation of interactions between pedestrians in Section 2.1.2.

### 2.1.1 Approaches

It is possible to divide the field of crowd simulation into three main approaches named: **Flow-based**, **Data-driven** and **Agent-based**, which are illustrated in Figure 2.2 and are each based on different principles.

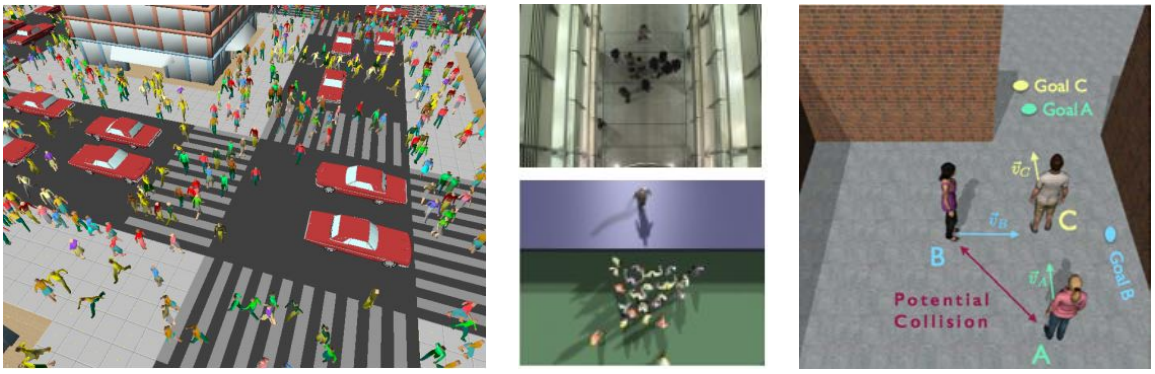



**Figure 2.2** – Illustration of the three main approaches for crowd simulation **Flow-based** [Treuille et al., 2006] (left), **Data-driven** [Lee et al., 2007] (middle) and **Agent-based** [Dutra et al., 2017] (right)

#### 2.1.1.1 Flow-based

In this approach, also called the macroscopic approach, the crowd is considered as a whole or as a *"Continuum"* [Hughes, 2002, 2003]. A *"Continuum"* is, according to Gan and Yong [2012], *"A body whose matter is continuously distributed and fills the entire region of space it occupies"*, with in our case the crowd as the body and pedestrians as the matter. These pedestrians are then set in motion while respecting some basic principles such as the preservation of matter and other constraints from various disciplines. For instance, some works [Shimizu et al., 2003, Pimenta et al., 2008, Kerr and Spears, 2005] consider the whole crowd as a continuous fluid, where the movement of the fluid particles (the pedestrians) is driven by the dynamics of fluids. Some other algorithms [Jin et al., 2008, Chenney, 2004, Treuille et al., 2006, Patil et al., 2010] used velocity fields in order to put in motion

the crowd without collisions with obstacles present in the environment. These velocity fields are either created manually, extracted from video or mathematically computed. These kinds of approaches are very interesting to simulate very dense crowds because of their complexity and the associated computation costs depend on the discretisation of the space chosen. However, they also lead to the creation of some artefacts such as the interpenetration between the visual representation of the crowd agents. A further disadvantage is that it is also not possible to individualise the behaviour of agents. To remedy this, Narain *et al.* [2009] proposed a hybrid approach in order to avoid constraining agents to follow a uniform model at all times. In regions with high crowd density, his algorithm use principles from fluids dynamics with agents represented as incompressible granular materials. For areas with lower density, agent velocity is obtained by interpolating velocity from a fluid dynamics model and the preferred agent velocity in order to reach their objective.

#### 2.1.1.2 Data-driven

For this approach, crowd simulation is performed using collected data. For instance, some works [Lee et al., 2007, Lerner et al., 2007] simulate the agent movements by looking for the closest real situation in order to replicate the same actions. In the case of Lerner *et al.* [2007], this search is based on the position and speed of the agents at a given time. Lee *et al.* [2007], on the other hand, code the state of an agent according to several variables (speed, formation of neighbours, ...) and use a statistical model (PCA) to predict the next action from the database of examples they have. Another method [Yersin et al., 2009] creates patches of crowd movements from real data and assembles them together. Charalambous *et al.* [2014] also proposed a perception-action graph to simulate agent's action. Each edge of this graph represents an agent personal state obtained from trajectories and each node is the action that transforms one state to another. Finally, new methods start to use machine and deep learning algorithms to model the crowd, predict and generate future movements of the agents in that crowd [Amirian et al., 2019, Zhong et al., 2016, Pellegrini et al., 2012, Zhao et al., 2019, Gupta et al., 2018, Alahi et al., 2016].

These methods therefore provide a relatively realistic simulation of crowd movements with a computation cost that differs according to the method chosen. However, a limitation is that they are dependent on the database provided as input.

#### 2.1.1.3 Agent-based

This approach is also known as microscopic approach. Each agent in the crowd is considered as an intelligent one with certain properties and a goal to reach. In most cases, a path planner gives the direction and the path to the goal that he has to reach [Van Toll et al., 2016, Bruneau and Pettré, 2017]. Furthermore, each agent uses a model to simulate the local interactions he has with the other pedestrians that are near him (interaction neighbourhood) and to avoid the obstacles present in the environment (trees, walls...). These local interactions with pedestrians can be numerous, the most recurrent being the avoidance of collision, but there are others such as following a person or walking in a group. In order to simulate them a lot of algorithms have been designed such as physics-based [Helbing and Molnár, 1995, Karamouzas et al., 2014, Reynolds, 1999], velocity-based [Paris et al., 2007, van den Berg et al., 2011, Moussaïd et al., 2011a, Pettré et al., 2009], vision-based [Dutra et al., 2017, López et al., 2019] or rule-based [Moussaïd et al., 2011b] models.

In this thesis we are particularly interested in this kind of approach, even if compared to other approaches, they are more expensive in terms of computation time. One of their major benefits is that they enable the simulation of heterogeneous crowds, which are more realistic in terms of local pedestrian behaviour. Indeed, pedestrians in a crowd do not have a uniform behaviour, for instance they do not walk at the same speed, some of them pay attention to their environment while others do not. Therefore, the agent-based approach provides a way to easily model such diversity by varying the parameters of the local interaction model and the interaction neighbourhood associated to each agent. The next section details the different existing algorithms in order to better understand the principles and thus be able to improve them.

### 2.1.2 Local Interaction Models

This section presents in detail several models of agent-based crowd simulators, by describing the computation of the local interactions as well as the interaction neighbourhood.

#### 2.1.2.1 Physics-based Models

Physics-based models use the laws of physics to set crowd agents in motion. In physics-based models, an attractive force is used between each agent and the goal they have to reach in order to put them in motion. Furthermore, interactions

between each agent are represented by repulsive forces for them to avoid collisions. Finally, the direction vector of an agent corresponds to the sum of all the forces that are applied to it, as illustrated in Figure 2.3. The first model to exploit such representation, named *"Social Forces"*, was introduced by Helbing *et al.* [1995]. In this model the intensity of the repulsive forces is correlated to the distance between the agents, so beyond a certain distance, these forces will be insignificant. Therefore, without any other process, the interaction neighbourhood corresponds to all agents at a certain distance. This model has been improved several times since, in particular by Karamouzas *et al.* [2014], who tied the intensity of the repulsive force to the time-to-collision between agents. This modelling directly changes the interaction neighbourhood (see Figure), with the selection of agents now being based not on distance but on the risk of collision, as illustrated Figure 2.3.

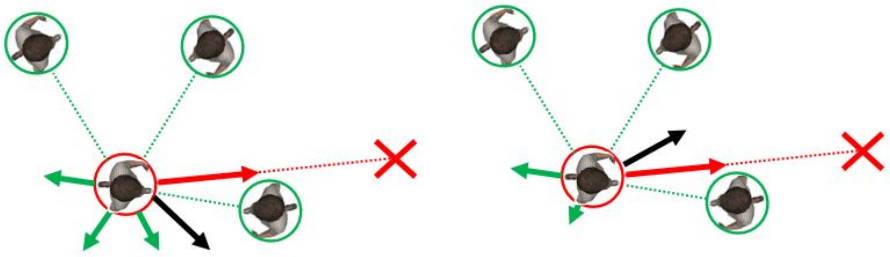


**Figure 2.3** – Illustration of physics-based model (Left [Helbing and Molnár, 1995] and Right [Karamouzas et al., 2014]) on a agent (red-circle), with the repulsive forces from the other agents (green arrows), the attractive force (red arrow) toward the goal and the final velocity vector (black arrow).

#### 2.1.2.2 Velocity-based Models

The general principle of velocity-based models comes from the work of Fiorini and Shiller [1998] who introduced the concept of *Velocity Obstacles*. The purpose of this algorithm is to compute for an agent $A$, the set of all velocities $VO_B^A$ for which a collision with agent $B$ (or an obstacle) will occur, as shown in Figure 2.4-left. Then, the velocity of agent A is modified in order to not be part of $VO_B^A$ while being as close as possible to the velocity vector towards the agent's goal. One of the most known algorithms in crowd simulation is called *Reciprocal Velocity Obstacles* (**RVO**) and is based on this principle. In **RVO**, Van Den Berg *et al.* [2008] modify the choice of the agent's velocity by using the average between the agent's current velocity and the desired velocity to avoid any collision. This difference avoids

a phenomenon of oscillation by the agents. **RVO** has been implemented and improved in the literature multiples times [van den Berg et al., 2011, Yeh et al., 2008, Golas et al., 2013, Godoy et al., 2014].

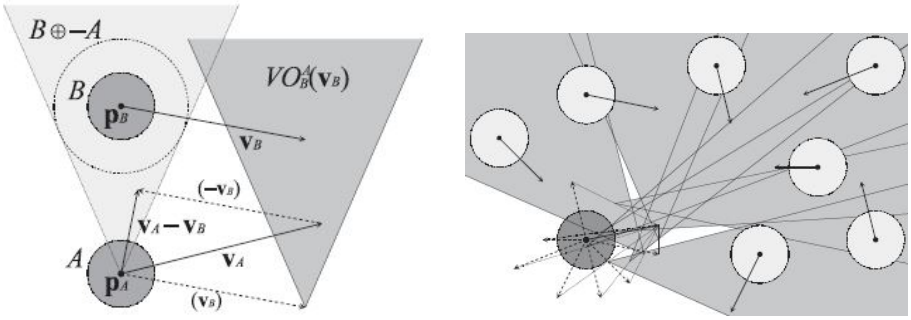



**Figure 2.4** – Illustration of the set of velocities $VO^A_B$ that cause a future collision for an agent A with an agent B (left), and extension to a case with multiple interactions (right)

Concerning multiple interactions (Figure 2.4-right), the set of velocities to avoid is represented by the union of $VO^A_{B^i}$ where $i$ is the number of agents in the interaction neighbourhood. In the case of RVO, this interaction neighbourhood is defined by a region around the agent, however we can find other methods for the selection of neighbours. In particular, Bruneau and Pettre [2017] proposed to use an area represented by a large semi-circle in front of the agent and a smaller semi-circle behind him, since a pedestrian perceives mainly the environment in front of him. Paris *et al.* [2007], for their part, use the field of vision of the agents as an interaction neighbourhood, furthermore, they do not consider agents hidden by an obstacle. Finally, Pettre *et al.* [2009] consider only the agents with which interaction occurs, and they restrict the interaction neighbourhood to a maximum of 7 neighbours.

#### 2.1.2.3 Vision-based Models

Vision-based algorithms use the field of view of each agent to extract information in the environment, process them in order to move them toward their goal without collision. In an obstacle-free environment, Warren *et al.* [2001] proposed a visual control law using the egocentric direction towards the objective to be reached and the rotation towards it weighted by the optical flow. This work was then extended by proposing a new visual control law [Warren and Fajen, 2008] using this time the bearing angle in order to navigate an agent toward his goal and avoiding the other obstacles present in the environment. Then some models proposed to compute the agent's motion by processing the image of the environment. For instance, Dutra *et*

*al.* [2017] computed two images representing the future time and distance at which the agent will be closest to other elements. Lopez *et al.* [2019] use the optical flow of the agent to create characteristics on each agent. These different pieces of information related to the risk of collision are then used in a cost function in addition to the agent's speed and goal. Then, this function calculates the speed of the agent to minimize the risk of collisions with all obstacles while maximizing the direction towards the agent's objective. This logic, illustrated in Figure 2.5, is very similar to the action-perception loop presented in Section 1.2. Furthermore, for these two last methods, the interaction neighbourhood is computed indirectly, since only pedestrians present in the agent's field of view are taken into account.

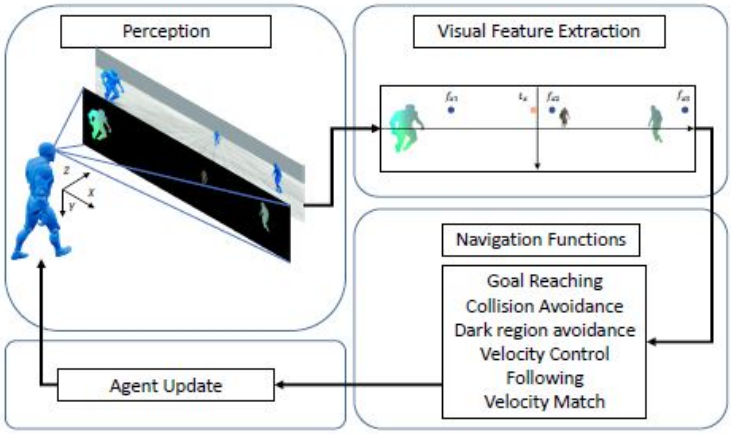



**Figure 2.5** – Illustration of the main loop used in Lopez *et al.* [2019] to update and move an agent.

In conclusion, the modelling of collision avoidance has been extensively studied for agent-based models, which is less the case regarding the interaction neighbourhood. Several definitions have been proposed either arbitrarily or based on basic concepts such as field of view. In this thesis, we are interested in experimental studies for understanding multiple interactions, and more especially the neighbourhood of interaction. The purpose of conducting experiments in specific situations is to understand the *human behaviour* in order to be able to later simulate it. Indeed, several works have already adopted this strategy in order to propose or improve agent-based model. For instance, in the case of simulation of following behaviour, Rio *et al.* [2014] proposed a visual control law from the results of two experiments where participants had to follow another pedestrian. Then they conducted more experiments [Dachner and Warren, 2014, Rio and Warren, 2014, Rio et al., 2018] in order to study multiple interactions. From their results, they improved their model on following behaviour and proposed a new definition

for the neighbourhood of interaction for such situations. The next section thus presents the various works carried out on the analysis of pedestrian interactions in a real environment.

## 2.2 Pedestrian Interactions in Real Environment

Navigating within a crowd involves many interactions between pedestrians such as following, walking together, as well as collision avoidance. In this thesis, we are particularly interested in collision avoidance behaviour, which is essential for safe navigation. Collision avoidance behaviour is strongly linked to the notion of proxemics. Proxemics is the study of people's perception and use of space [Hall, 1963]. The exploration of this field of study emerged in the 1960's as an interdisciplinary approach to understanding complex human behaviour in crowds. Proximity was shown to be influenced by cultural aspects [Hall, 1963], as well as by gender [Brady and Walker, 1978], age [Remland et al., 1995], attractiveness [Kmiecik et al., 1979] or speed of movement and density [Seyfried et al., 2005b].

In his work, Hall introduces several inter-personal distances in humans representing the spaces around them. Four spaces, namely intimate, personal, social, and public spaces have been described according to the nature of social interactions that occur within these spaces (Figure 2.6) [Hall, 1966]. In that context, the collision avoidance task, which requires the pedestrian to identify if and when a collision will occur [Cutting et al., 1995], not only aims at avoiding physical contact but also at preserving social distances. In particular, the term personal space (PS) is used to describe the minimum social distance from other pedestrians when walking [Gérin-Lajoie et al., 2005].

In this section, we review research works that focused on 1-to-1 and 1-to-N pedestrian interactions, especially during a collision avoidance task. We first considered studies that analysed pedestrian trajectories to understand the pedestrian movements during the interaction.

### 2.2.1 Trajectory Analysis

Research on human interactions and navigation can be divided into two major categories. The first one, referred as "pairwise experiment", concerns controlled laboratory experiments during which one participant is going to walk towards a goal while encountering another pedestrian. The second concerns studies involv-

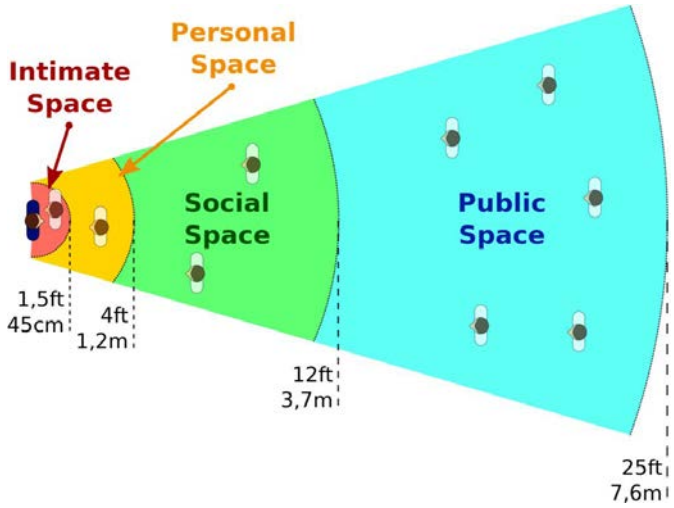



**Figure 2.6** – Proxemics distance

ing multiple human interactions, whether with a group of individuals or even a crowd. In order to record pedestrian trajectories when interacting with each other, several tools are at the disposal of scientists. When an experiment is performed in a laboratory with a small number of participants, researchers most often use optoelectronic systems to record human motion as illustrated in Figure 2.7-left. By using 3D markers fixed on the participants and tracking cameras (e.g., Qualisys[1], Vicon[2]), it is possible to record their trajectories. However, this kind of equipment cannot be used for outdoor use or with a large number of people. Therefore, some works rely solely on video recording such as the experiments described by Boltes and Seyfried [2013] and illustrated Figure 2.7-right. In this kind of work participant's trajectories can either be annotated by hand or detected using tracking algorithms.

#### 2.2.1.1 Pairwise Experiment

Pairwise experiments were performed in controlled laboratory settings. Typically, experimenters ask two walkers (either two participants, or one participant and one confederate), to walk through the experimental area to reach a target straight ahead. During this task, one walker encounters the other arriving from a certain angle, also known as the crossing angle, and must avoid any collision with him. However, no other indication is given to participants. Therefore, it is for them to

[1] https://www.qualisys.com
[2] https://www.vicon.com

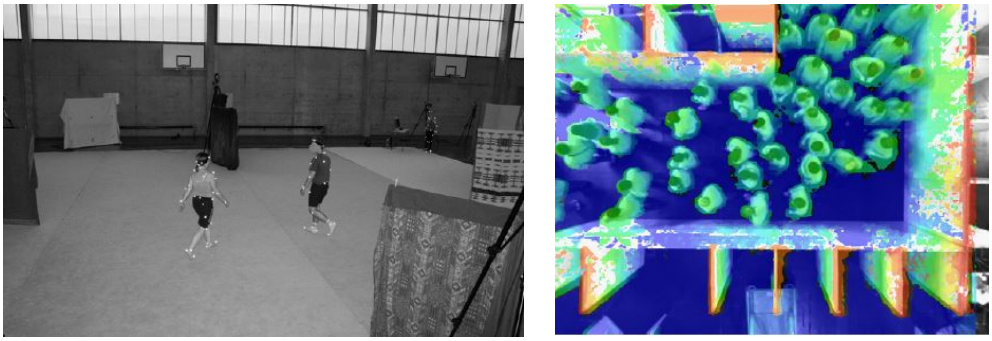


**Figure 2.7** – Illustration of two systems to record trajectories during experiments about interactions between pedestrians. Motion capture system (e.g, Vicon) are used in the left image from Olivier *et al.* 's experiment [2012], whereas videos and annotations are used in the right image from Boltes and Seyfried [2013] works .

judge whether a collision is actually going to occur, and to decide which movement they have to perform in order to avoid it.

In each experiment, the researchers focus on only one or two factors in order to understand their influence on the participant's locomotion. To this end, the creation of the experience is standardized so that the participant's behaviour should only be influenced by these factors. In their works, Olivier *et al.* [2012, 2013] designed an experiment where one participant encountered a confederate crossing at a 90° angle. The initial positions of the two walkers have been manipulated to ensure that they passed at a minimum distance $d$ from each other if they walked without any change in their locomotion. During the experiment, the authors varied $d$ between 0 and 2 meters in order to see the impact on the participant's locomotion. Furthermore, they set up walls between the two people, as illustrated in Figure 2.8, so they could not see each other at the beginning of the experiment but only after a certain time ($tsee$), in order to ensure that they had the time to reach comfort speed.

In order to analyse collision avoidance behaviour, Olivier *et al.* [2012] introduced a variable called *minimum predicted distance* ($mpd$). This variable, computed at a time (t), corresponds to the future crossing distance between two walkers if they keep walking at their current speed. The evolution of this minimal predicted distance, which is associated to the time to closest approach ($ttca$), is displayed in Figure 2.8. This metric relates to mutual motion adaptation since any change in one or both participants' trajectory will induce variations of $mpd$.

It has first been shown that walkers adapt their motion only if required. Indeed, motion adaptations were only observed when the initial $mpd$ was less than 1m. When considering situations where the initial future crossing distance is

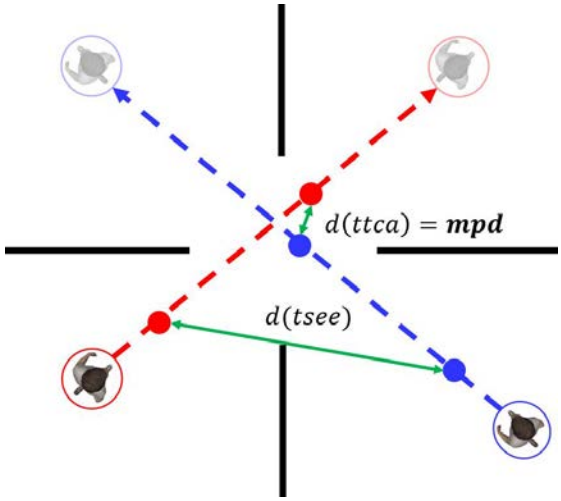



**Figure 2.8** – Setup used during Olivier *et al.* 's experiment [2012]. The participant (red) and the confederate (blue) start to see each other at time *tsee*. During the experiment the minimum distance between them (*mpd*) takes place at the time *ttca*.

less than 1m, the evolution of *mpd* during the interaction can be described into 3 phases, that were associated with 3 steps in motion adaptations, illustrated in Figure 2.9. First, $mpd(t)$ remains constant ("observation phase"): since no motion adaptation is performed. Second, $mpd(t)$ is increased so as to reach a comfortable crossing distance ("reaction phase"). Third, $mpd(t)$ remains constant at a comfortable distance until the actual crossing ("regulation phase") and shows that the collision avoidance task is solved before walkers cross each other.

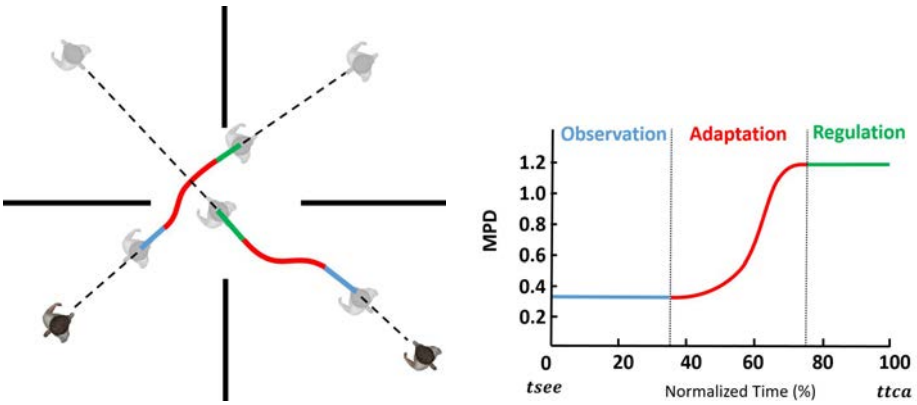



**Figure 2.9** – Collision avoidance between two pedestrians can be described as a 3 step process (Observation (blue), reaction (red) and regulation (green)) when considering the evolution of the minimum predicted distance (*mpd*), which is the future distance of closest approach.

While it was shown that collision avoidance is mutually solved by the walkers

involved in the interaction [Olivier et al., 2013] the contribution of each walker differs depending on their order at the crossing: the walker passing first contributes less than the second. This result was discussed with respect to the asymmetry of the personal space (PS). In addition, Huber *et al.* 's research[2014] focuses on the influence of the crossing angle on collision avoidance strategies. By studying several situations, they showed that collision avoidance requires a change in trajectory and speed. Moreover, the velocity adjustment was more important in the case of acute angles (i.e. 45° and 90°), which is consistent with previous work[Basili et al., 2013].

Some studies have also investigated the influence of individual characteristics such as height, gender, or age on locomotion during collision avoidance [Grundberg et al., Rapos et al., 2019, Knorr et al., 2016, Bourgaize et al., 2020]. Overall, participants showed similar avoidance strategies whatever their specific individual characteristics. However, some quantitative differences could be noticed with more risky behaviour in older adults [Grundberg et al.], or larger crossing distance with larger pedestrians [Bourgaize et al., 2020].

#### 2.2.1.2 Multiple Interactions

If we consider multiple interaction situations, we can first present studies that explored collision avoidance between a participant and two other walkers [Dicks et al., 2016, Meerhoff et al., 2018b]. Dicks *et al.* [2016] designed an experiment where a participant was interacting with one or two pedestrians on a head-on collision course, who could be reading a message on their mobile phone or not during a street crossing task. Their results showed that participants took more time to complete the crossing when they encountered two pedestrians. Furthermore, the participant's speed was lower when these pedestrians were looking at their phones. Meerhoff *et al.* [2018b], on the other hand, showed that tryadic interactions in a 90° collision avoidance task can result in both sequential or simultaneous interactions, but they mentioned that additional work is required to identify the conditions which invite for such interactions.

Also as part of a laboratory experiment but with a much larger number of participants, Seyfried *et al.* [2005a] were interested in the impact of crowd density on the walking speed. They conducted an experiment, illustrated in Figure 2.10-left, where $N$ participants ($N \in [1, 34]$) had to walk in a circular corridor with different widths. Their results show a negative linear relationship between the participant's velocity and the density of people (Figure 2.10-right). However Chattaraj *et al.* [2009] showed that this can be dependent on cultural differences.

In a similar context, Gorrini *et al.* conducted several controlled experiments on crowds walking in corridors. The first [Gorrini et al., 2013] focused on the impact of corridor curvature on walking speed with a crowd consisting of either individuals or groups. They found that groups walk slower than individuals and that the curvature of the corridor has an impact on the movement of the crowd. A sharp curve reduces the walking speed and thus the flow rate of the crowd. The second experiment [Gorrini et al., 2016] concerned the influence of crowd flow on pedestrians' walking speed. By changing the type of flow (unidirectional or counter-flow) and the flow's density, they showed that increasing the flow reduces walking speed. Moreover, in the case of a counter-flow condition, this effect was greater for groups than for single individuals, as this situation presents more turbulent trajectories.

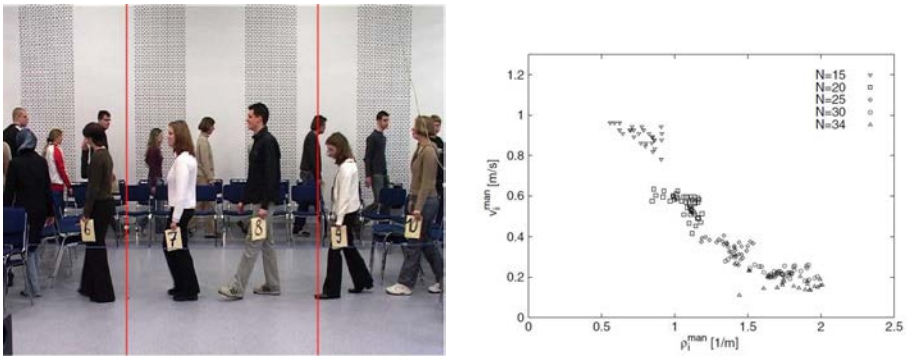



**Figure 2.10** – Left: laboratory setup used during Seyfried *et al.* 's experiment[2005a]. Right: Resulting fundamental diagram which represents the relation between $N$ participants velocity and density with $N \in [15, 20, 25, 30, 34]$

There are also several studies using outdoor situation videos to analyse crowd behaviour. In this context, Costa[2010] studied groups of people walking in the street, and he noticed that groups composed only of men were more dispersed and their speeds were higher than groups composed of women or mixed groups. He also noticed that groups composed of 3 people (known as triads) moved in a triangular arrangement ("<"). Gorrini *et al.* [2015], for their part, studied the constitution of a crowd in a tourist passageway using videos. They showed that this crowd was mainly composed of groups (84%), with the majority being groups of two (44%). Their results indicated that groups generally walked slower than individuals and that two-person groups tended to be less dispersed than other groups.

It can be noticed that all these studies focusing on crowds considered general attributes, such as average walking speed, and not on the behaviour of a specific subject in a crowd. Indeed, it can be difficult to understand trajectory formation

in such situations, as it is equivalent to inverting the injective process by which multiple sources of interactions combine to influence a single trajectory of lower dimension. To do so, another possibility is to understand what elements pedestrians have taken into account, which are ultimately part of the elements they have perceived over time, as explained in section 1.2. In that context, it has been shown that vision is a major perceptual system that guides locomotion [Warren, 1998, Patla, 1997, Berthoz, 2000]. We will then present in the next section the general concepts linked to the visual systems as well as the studies which investigate its role during human locomotion.

### 2.2.2 Vision and Locomotion

This section first introduces some terminology and terms in relation to the vision that are important to vision and gaze activity in Section 2.2.2.1. Then Section 2.2.2.2 focuses on the presentation of the gaze activity and its analysis. Finally, Section 2.2.2.3 presents the different experiences related to locomotion and gaze activity.

#### 2.2.2.1 Concepts about Vision

Vision is a perceptual system using the eyes (Figure 2.11) as a sensing organ to perceive the environment [Vickers, 2007]. In order to create an image of the environment, the rays of light reflected or emitted by all elements in the environment, enter the eye through the pupil, an orifice surrounded by the iris. Then, each ray of light goes through the cornea and the pupil in order to finally reach the retina. Finally, the retina converts these light signals into electrical signals and sends them to the brain via the optic nerve in order to process it. Furthermore, a portion of the retina name *fovea* is the location where the visual acuity is maximum, allowing thus to see visible small details on objects.

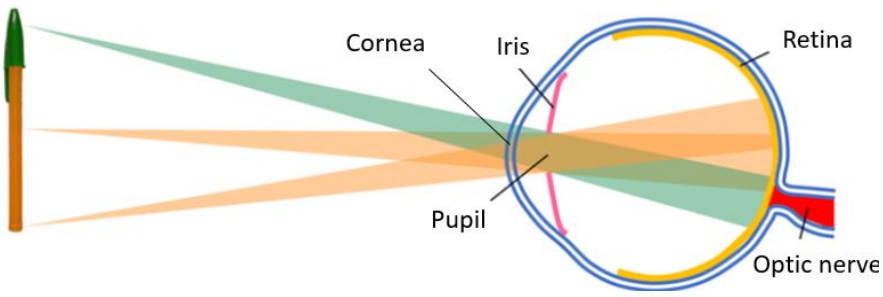



**Figure 2.11** – Anatomy of the eye

The portion of the environment perceived by the eyes is called visual field [Spector, 1990] or field of vision and is often expressed in degrees. For monocular vision (one eye), the field of view is approximately 100° horizontally and 60° vertically whereas for binocular vision (two eyes), the field of view is approximately 200° horizontally but do not change vertically. The gaze, being the center of the field of vision, is defined as the vector starting from the center of the *fovea* and passing through the center of the iris and can be expressed in different referential systems such as world, head or centre of the eyes. The visual field is further divided into several subregions [Pöppel and Harvey, 1973, Simpson, 2017] as shown in Figure 2.12. In its centre is the *fovea*, which corresponds to the zone where the visual acuity is maximum. Depending on the field of application of the researchers [Strasburger, 2020], this zone varies from $1° \sim 2°$ up to 5°. Then, from 5° to 10° is the *parafovea* zone. These two zones are often considered as the central vision and the rest of the visual field as the peripheral vision. We find then, the near-peripheral vision $(10° - 30°)$, the mid-peripheral vision $(30° - 60°)$ and the far-peripheral vision $(60° - 100°)$.

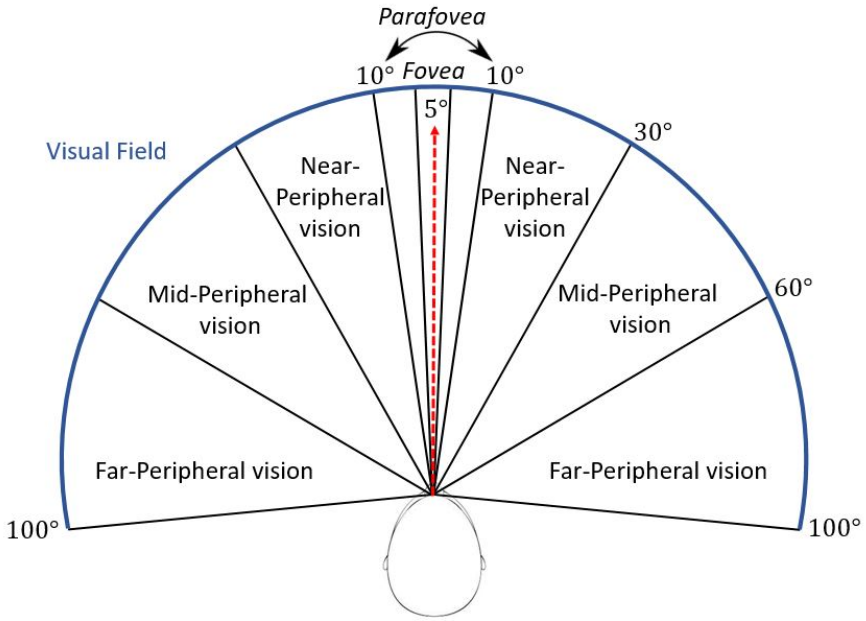



**Figure 2.12** – Illustration of the different regions of the visual field. With the *fovea* $(0° - 5°)$, the *parafovea* $(5° - 10°)$, the near-peripheral vision $(10° - 30°)$, the mid-peripheral vision $(30° - 60°)$ and the far-peripheral vision $(60° - 100°)$.

With these definitions of the various zones of the visual field, many researchers studied their utility in relation to tasks that participants were expected to perform in specific environments. For instance, Collier *et al.* [1931] showed that simple

pattern recognition is possible across the majority of the field of view (central and peripheral), however, the closer an object is to the *fovea*, the more details about it can be seen. For the case of locomotion, many researchers [Stoffregen et al., 1987, Jovancevic et al., 2006, Marigold and Patla, 2007, Marigold, 2008, Turano et al., 2005] showed how the presence of visual cues in the peripheral vision are important during walking in an environment alone or with other pedestrians. Moreover, regarding studies in a social context, it has also been shown that gaze conveys a certain amount of information during social interactions [Frischen et al., 2007, Nummenmaa and Calder, 2009, Itier and Batty, 2009]. In the frame of the theory of mind (ToM), which refers to the ability to explain the mental state of another person, some works [Itier and Batty, 2009, Calder et al., 2002, Baron-Cohen, 1997] indicate that gaze could be used as a relevant feature to understand people attention. Finally, gaze can also have an impact on the perception of others. In their work, Mason *et al.* [2005] demonstrated that a person is perceived as more likable by another person when a mutual gaze occurred instead of an averted gaze.

All these previous studies highlight the importance of the vision and the gaze activity when walking or interacting with other people. For this reason, in this thesis, we work on the hypothesis that the analysis of gaze activity in parallel with the analysis of the trajectory can provide additional information to understand the actions of pedestrians when navigating in a crowd. In particular, by looking at the information perceived before making a movement or an action.

#### 2.2.2.2 Gaze Activity

The study of gaze activity is the analysis of eye movement or 3D gaze points over time. When these data are analysed in the reference system of the head, it is referred as eye-gaze activity. Gaze activity, on the other hand, corresponds to the global coordinates of the gaze in the world space [Guitton and Volle, 1987], which therefore also accounts for head rotations. These data are recorded using eye trackers with different systems and methods [Kar and Corcoran, 2017] in order to estimate the direction of the gaze over time and record it. Moreover, it is necessary to calibrate these tools before performing a recording.

According to Findlay *et al.* [2003] three types of eye movements are related to the target selection by the perceiver:

- **Pursuit** is a smooth and continuous movement of the eyes to follow a target of the gaze (Figure 2.13-left).

- **Vergence** keeps the eyes pointing toward a target that is in a distant direction (Figure 2.13-middle). This causes both eyes to rotate in opposite directions.
- **Saccade** is a fast movement allowing the perceiver to move from one target to another (Figure 2.13-right). Unlike vergence and pursuit, the eye movement during saccade is extremely fast and consists of a multitude of jumps. In addition, it is also important to note that no useful visual information is acquired by the perceiver during the saccade [Parasuraman and Rizzo, 2008].

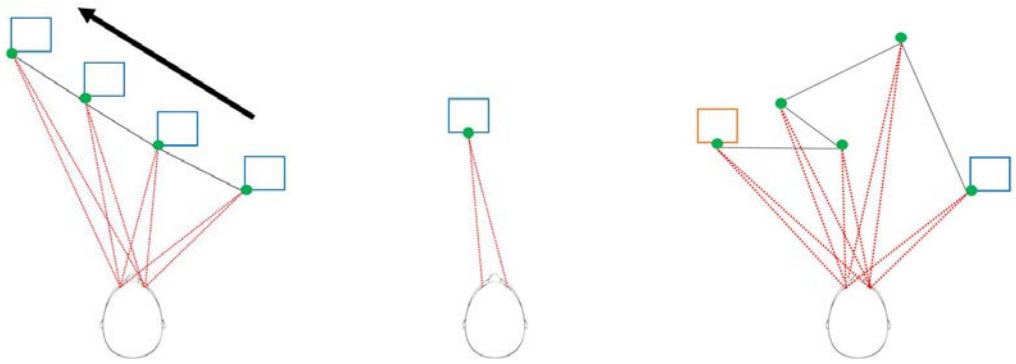


**Figure 2.13** – Illustration the three eye movements: **Pursuit** (left), **Vergence** (middle) and **Saccade** (right). For each illustration, the eye vectors are in red, the gaze points are in green and the rectangles represent the objects being looked at.

Furthermore, when walking, a different eye movement (**Vestibulo-ocular reflex**) appears, allowing the perceiver to compensate for the body and head movements in order to keep the image of the visual world steady on the retina. These movements occur automatically and independently of the perceiver's will. In their work, Leigh and Zee [2015] also introduce the term **visual fixation** corresponding to the movement of the eyes that keeps an object in the fovea.

In order to describe gaze activity, the term fixation is commonly used to define a period when eye movements are more or less stationary. Thus, by using it, it is possible to define gaze activity as a succession of fixations (lasting on average 200-400 ms) separated by saccades (lasting on average 30-50 ms) [Parasuraman and Rizzo, 2008]. Furthermore, there are micromovements of the eye within a fixation that are associated with drifts, tremors, or micro-saccades. In practice, a fixation is associated with a set of 3D points of the gaze being all within 2° 4° degrees of each other (size of the fovea). However, the exact definition of fixation can be unclear as explained by Hessels *et al.* [2018] in a review where they present

the different definitions used by 124 researchers. It is therefore important in every work on gaze activity study to clearly define what is considered to be a fixation and how it is detected.

Many methods have been developed to detect fixations and saccades. In their work, Salvucci and Goldberg [2000] present different spatial and temporal criteria that are the basis for the majority of these methods. Fixations, in contrast to saccades, are gaze points with low velocity, located around the same spatial region and therefore with a low dispersion for all the data in the same fixation. In addition, fixations also have a temporal constraint as they rarely last less than $100ms$ and are often between $200ms$ and $400ms$, whereas saccades last between $30ms$ and $50ms$. Using one or several criteria, Salvucii and Goldberg [2000] then proposed five basic algorithms for the classification of fixations. One of them in particular is based on the gaze points' velocity using thresholds to identify fixations. This kind of velocity-based method (**I-VT**) was later improved by various researchers. For instance, Komogortsev and Karpov [2013] use speed thresholds and point dispersion thresholds combined with a time window to detect fixations. Salvucci and Goldberg [2000] also proposed spatial-based methods using the duration of fixations coupled with a threshold on the dispersion of a set of gaze points (**I-DT**) or the location of these points in the space (**I-AOT**). Furthermore, in the past few years, researchers have started using deep learning approaches such as the one proposed by Starsev *et al.* [2019]

#### 2.2.2.3 Experiments

In the experimental field, there have been numerous studies that investigated gaze activity during locomotion tasks for diverse situations. For instance, some of them focused on analysing both participants' locomotion and gaze activity while they were walking on different kinds of ground [Marigold and Patla, 2007, Thomas et al., 2020], as illustrated in Figure 2.14-Top. In particular for indoor laboratory experiments, Marigold and Patla [2007] showed that participants' fixations were mainly on areas that were steeped on. In addition, a lot of these fixations were directed towards transition areas between two surfaces, which are finally the most complex surfaces. Thomas *et al.* [2020] conducted the same kind of experiments but for both indoor and outdoor situations. Their results indicate that both participants' speed and eye angle were lower when they walked on the more complex surfaces (Figure 2.14-Bottom).

In their work, Saeedpou and Parizi [2020] were interested in the influence

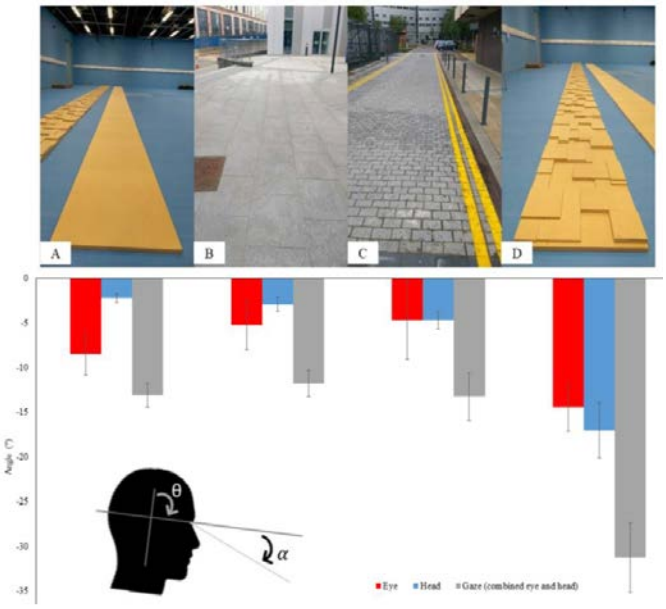



**Figure 2.14** – Top: Illustrations of the different indoor (A&D) and outdoor (B&C) types of ground used during the experiment of Thomas *et al.* [2020]. Bottom: Mean angle for eye (red), head (blue) and combination of eye and head (gray) for the four different ground.

of task planning on gaze activity and locomotion. In their experiments, they asked participants to perform three different scenarios where the task became progressively more complex. In the first one participants had to walk toward a bookcase with a cup on it, in the second participants had to do the same thing but also pick up the cup. Finally, in the third situation, participants performed all the precedent tasks but they also had to put the cup higher in the bookcase. Participants exhibited the same behaviour regarding the beginning of the task with an acceleration towards the bookcase and then a deceleration when they were close to it. However, the participant's peaks of velocity were higher for more complex tasks. In addition, the participant's gaze activity was also impacted by the task complexity, with higher fixation duration. This can indicate that to take information and process it, the duration of fixation has greater importance than the number of fixations. Moreover, these results also indicate that gaze activity is task-dependent, which is consistent with previous work [Yarbus, 2013]. In another situation, Cinelli *et al.* [2009] conducted an indoor laboratory experience where participants had to go through 2 motor-driven sliding doors. Two conditions were studied either the two sliding doors had the same velocity (symmetrical) or different ones (asymmetrical). There was no difference in participants' locomotion between the two conditions, with their speed decreasing before going through the doors. Furthermore, participants fixated longer on the second door for the

asymmetrical condition as the door movement was more complex.

However, these previous studies did not include interactions with other people. In this perspective, Laidlaw *et al.* [2011] conducted an experiment where participants were seated in a waiting room with another person (confederate) or a videotape of him. They found that participants looked more at the confederate when he was on videotape than when he was physically present, which suggests that gaze activity is also dependent on the social situation. Regarding experimental studies during dynamic tasks, Kitazawa and Fujiyama [2010] had participants walking on a platform on which a collision could occur with other confederates or static objects. The purpose of this experiment was to study the relationship between gaze and the Information Process Space (IPS) which is related to the notion of proxemics [Hall, 1963]. They noticed that participants do not gaze more at objects and other pedestrians located in the scene than at the ground. Furthermore, they deduced that the IPS shape is not a homogeneous circle around the walker, but presents a more important anterior area. For their part, Jovancevic-Misic and Hayhoe [2009] conducted an experiment where participants had to walk around an oval track while other people acted in specific ways with predefined risks of collision, as illustrated in Figure 2.15. They demonstrated that participants adapt their gaze strategies depending on the behaviour of surrounding persons as pedestrians with risky behaviours were more gazed at by participants.

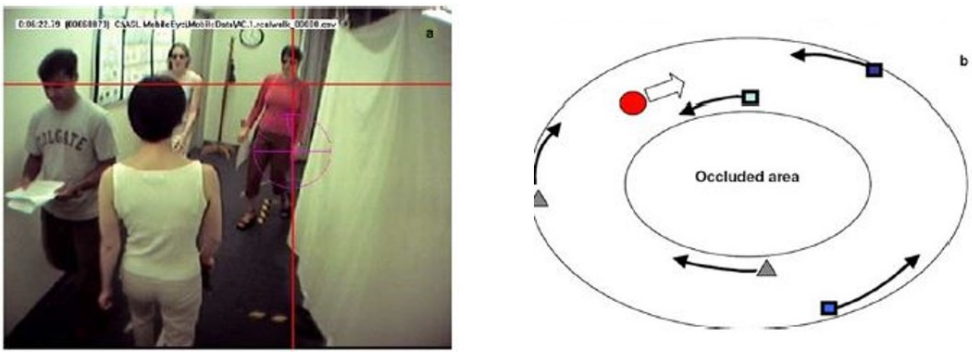



**Figure 2.15** – Experimental setup (oval corridor) used during the experiment of Jovancevic *et al.* [2009]. The participant is represented by the red dot and the confederates by the black squares and arrows. The top image represents a point of view of the participant during the experiment with his gaze represented by the red crosshair.

In their work, Croft and Panchuck [2017] studied avoidance strategies between two participants using a similar setup than for pairwise experiments (Section 2.2.1.1). Participants had to walk on either a constrained or an unconstrained path and avoid a collision with a confederate. Furthermore, this confederate was

walking either fast or slow with a crossing angle of 90°. They found that constraining the path had an impact on participant locomotion and avoidance strategies. Moreover, participants' gaze activity (e.g duration and number of fixations) was a good predictor to determine whether one pedestrian passed in front of the other. Regarding works with a crowd, Hessels *et al.* [2020] conducted an experiment where participants had to walk in a corridor filled with pedestrians and static objects, as shown in Figure 2.16. Two situations were studied. In the first, participants walked and avoided other pedestrians, while in the second they performed the same task but also pressed a button each time they saw a direct gaze toward them from oncoming people. In addition, the crowd followed a scripted scenario in order to repeat the same visual stimulus for each participant. For instance, 30% of the crowd were instructed to look at the participants, $10\% - 20\%$ of the crowd had to cross in front of the participants. They found no differences for locomotion (speed and collision) between the two situations. Furthermore, regarding gaze activity, there were also no differences on the ratio of fixations on obstacles, pedestrian's heads or pedestrian's body. This suggests that faces do not attract the gaze during locomotion whereas it was the opposite in static situations involving participants looking at images [Birmingham and Kingstone, 2009, Langton et al., 2008]. However, in the second situation, they noticed that when a mutual gaze occurred, fixation lasted longer.

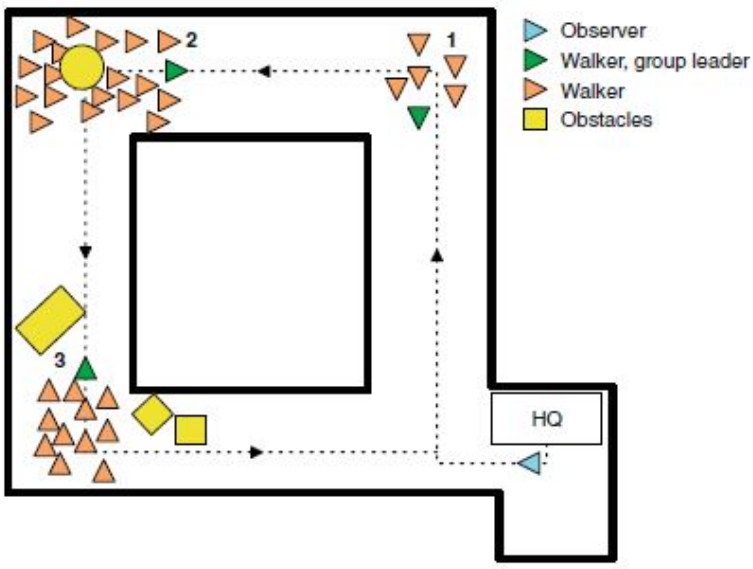



**Figure 2.16** – Schematic overview of the experimental setup used for Hessels *et al.* 's study [2020], with participant (light blue arrowhead), static obstacles (yellow) and three groups of pedestrians (orange arrowheads, with group leaders in green arrowheads). The black arrow represent the path participants had to take.

Finally, Gallup *et al.* [2012], conducted an experiment in the corridor of a public building, where they put at one side of it a device with a hidden camera and an attractive visual stimulus. From these videos, two reviewers estimated some information from the pedestrians, such as their gender, if they belong to a group and if they looked at the stimulus. They showed that gaze upon this visual stimulus increased if other pedestrians in the crowd displayed visual cues toward this stimulus, especially when these pedestrians were in front.

All these experimental studies analysing participant's behaviour during interactions with other pedestrians have shown that the experimental condition has a significant influence on human behaviour when walking. In particular, the initial conditions (e.g, speed [Olivier et al., 2012, 2013], crossing angle [Huber et al., 2014, Meerhoff et al., 2018b], task to perform [Saeedpour-Parizi et al., 2020, Yarbus, 2013]), the behaviour of other pedestrians [Jovancevic-Misic and Hayhoe, 2009], mutual gaze [Hessels et al., 2020], reading a message [Dicks et al., 2016]) or the presence of a visual stimulus [Cinelli et al., 2009, Gallup et al., 2012] have an impact on the participant's locomotion or his gaze activity. It is therefore crucial to be able to reproduce the exact same stimuli for all participants during the experiment. This is more or less possible for controlled laboratory experiments with a limited number of participants. In the case of a crowd, it is also possible, using scenarios, to have similar actions or ratios of pedestrians with a specific behaviour [Jovancevic-Misic and Hayhoe, 2009, Hessels et al., 2020]. However, despite all the possible attention, the visual stimulus will never be exactly the same for each participant during all the experiment. For this reason, we are looking to use Virtual Reality (VR) as a tool to conduct experiments on human behaviour involving interactions with pedestrians. The next section presents some of the different VR tools, but also the impact of VR on human perceptual systems, some of the limitations of VR, and the experiments that have been conducted to study human behaviour and locomotion in VR.

## 2.3 Virtual Reality

### 2.3.1 Technology

The expression virtual reality (VR) was introduced by Jaron Larnier in the 1980s, however there are some precursor elements such as the *"Sensorama"* (Figure 2.17-a) developed by Morton Heilig [1962]. In the words of Fuchs *et al.* [2006]: *"The purpose of virtual reality is to enable a person (or a group of people) to experience a*

*sensory-motor and cognitive activity in an artificial world, created digitally, which may be imaginary, symbolic or a simulation of certain aspects of the real world".* The first VR Headsets (HMD) began to appear in the early 1990s (Figure 2.17-b) and have now become common and affordable for the general public, in particular for their application in video games (Figure 2.17-c). There are currently more than a dozen HMDs on the market (e.g, Vive, Occulus, Valve Index...) each with its own advantages and disadvantages. In addition to HMDs, there is another virtual reality technology called CAVE ( for "Cave Automatic Virtual Environment"), which immerses a person through the use of 3D glasses and walls on which 3D videos are projected (Figure 2.17-d).



a) b) c) d)

**Figure 2.17** – From left to right several virtual reality devices through time with: Sensorama (a) in 1956, SegaVR (b) in 1991, Vive (c) in 2019 and Immersia Cave (d) in Rennes

In addition to the application of VR for video-games, several researchers became interested in studying human behaviour in the early 2000s [Loomis et al., 1999, Blascovich et al., 2002, Tarr and Warren, 2002]. In their discussion about experimental research in psychology, Loomis *et al.* [1999] explain that scientists in this field of application had to compromise between experimental control and ecological validity. According to them, VR offers the best tradeoff between these two components compared to traditional methods (Figure 2.18). Indeed, VR enables scientists to manipulate and have control over the whole experimental environment, which is impossible for in the physical world. For instance, when studying pedestrian interactions, it is possible to control all the pedestrians' actions over time, their behaviours, their attributes (e.g, gender, height, morphology, clothes...).

However, despite all the benefits that VR offers for experimental research, there are also some drawbacks such as the reduced field of view offered by HMDs or the possibility of finding artefacts in the virtual environment. In addition, it can be difficult for VR to accurately and reliably reproduce all the elements that the human perceptual system receives as input in a physical environment. Therefore, it is necessary to assess the differences in human behaviour in a real or virtual environment and thus be aware of the potential biases induced by VR.

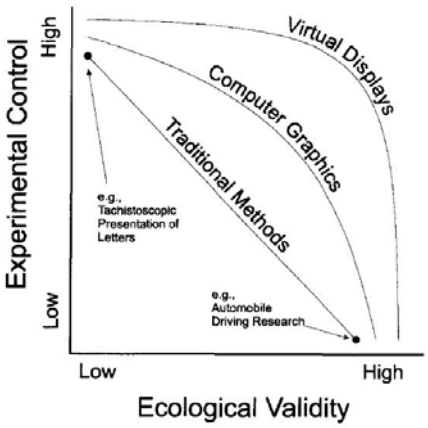



**Figure 2.18** – Plot of the compromise between experimental control and ecological validity for traditional methods and VR according to Loomis *et al.* [1999]

## 2.3.2 Evaluation of Virtual Reality

### 2.3.2.1 Locomotion and Personal Space

There are different methods for a user to move in a virtual environment (VE). These include the use of locomotion interfaces that combine an input device (joystic, keyboard...) and a control law which transcribe input data into movements in the VE. For instance a classic control law is the *"linear rate"*[Cirio et al., 2013] where the movements are mapped according to the participant's tangential and angular velocity. Another method is to enable the user to walk physically in the VE using cameras and 3D markers (included in the HMD or on the user's body). These systems record the user's position and movements in the real environment (RE) and stream them into the VE. Finally, treadmills are also used to move in a VE, especially over long linear distances. Many studies have therefore focused on evaluating the impact of VR and locomotion methods on participants' trajectories when immersed in a virtual environment. In this perspective, Cirio *et al.* [2013] proposed a framework to compare trajectories performed in a real and virtual environment (Figure 2.19).

For this purpose, they introduced several metrics based on the shape of the trajectory, the participants' performance, and other kinematic characteristics. They carried out an experiment in which participants had to walk through a door with different orientations in a real or virtual environment. They varied not only the input device of the locomotion interface (joystic, keyboard, gamepad) but also the control law (linear rate, inertial rate, joyman), the point of view of the participant's camera in the virtual environment (subjective camera, third person or

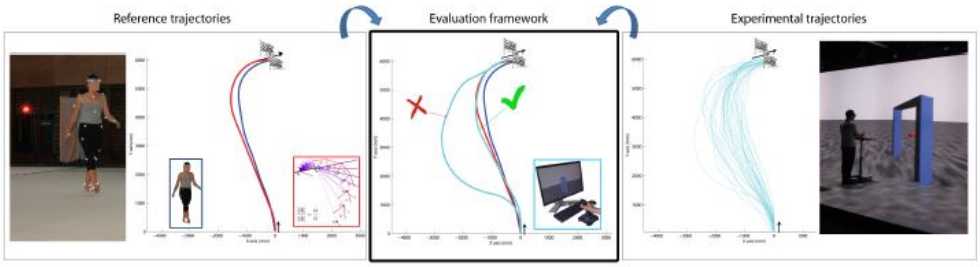



**Figure 2.19** – Evaluation framework (center) presented by Cirio *et al.* [2013], with the reference trajectories in the real world (left) and trajectories in the virtual environment (right)

fixed camera), the camera field of view (45°, 60°, 90°) and the type of VR display. Overall, they found no significant differences for participants' trajectories between real and virtual environments. Then other scientists focused on the evaluation of trajectories in a real or virtual environment during collision avoidance with either a static [Fink et al., 2007, Agethen et al., 2018] or dynamic [Olivier et al., 2017, Bühler and Lamontagne, 2018] obstacle. In particular, Olivier *et al.* [2017] performed a pairwise experiment in both real and virtual environments with a 90° crossing angle collision avoidance task. They used a CAVE as VR display and evaluate multiple locomotion interfaces, moreover, they also varied the initial $mpd$ between $-2m$ and $2m$. In Buhler and Lamontagne's work [2018], participants performed a collision task with three agents coming in front of them in both real and virtual environments. Overall the results for all these studies were similar. Trajectories in real and virtual environments share similitudes, especially regarding their shape. Participants also keep similar strategies to avoid obstacles, such as the side taken to perform the maneuver's [Bühler and Lamontagne, 2018]. All these studies also show that some quantitative differences exist regarding participant locomotion, especially participants' velocities being lower and minimal distance with obstacles being higher in VR.

Thereafter other studies [Gérin-Lajoie et al., 2008, Sanz et al., 2015] have looked at the impact of VR on personal space (PS section 2.2) during a collision avoidance task. Results show that the shape of the PS is similar in a real or virtual environment. However, the size of the PS is increased in VR, which is consistent with previous studies. Additionally, in their work Argelaguet-Sanz *et al.* [2015] showed that the size of the PS was also dependent on the type of obstacle. Indeed, in the experiment participants kept a larger distance with anthropomorphic obstacles than with inanimate objects.

#### 2.3.2.2 Distance Perception

Several studies have also looked at the impact of VR on distance perception [Loomis et al., 2003, Willemsen et al., 2004, Plumert et al., 2005]. Loomis *et al.* [2003], asked participants to look at spheres on the ground then to close their eyes, walk in the opposite direction, turn around and point the localisation of the spheres (triangulation task). They found significant differences between the correct position of the sphere and the one pointed by participants.



**Figure 2.20** – Real (right) and virtual (left) environment in which participants have to estimate the distance between themselves and the building during Plumert *et al.* 's experiment [2005].

Plumert *et al.* [2005], on the other hand, conducted an experiment where participants (adults and children) had to estimate the distance between themselves and a building in both real and virtual environments (Figure 2.20). The estimations were similar both in real and virtual environments, which is unexpected as it is in contrast to the precedent results [Loomis et al., 2003]. These differences might be explained by the difference in VR display (HMD [Loomis et al., 2003] or CAVE [Plumert et al., 2005]). Another difference is the object of interest. In Loomis *et al.* 's experiment [2003] participants estimated the distance with spheres very close to them (between $1m$ and $4m$) whereas in Plumert's work the building was at a range between $6m$ to $36m$. Finally, Willemsen *et al.* [2004] found that the differences in distance perception can be partially explained by the mass and moment of inertia of HMDs.

#### 2.3.2.3 Gaze Activity

Concerning the impact of VR on gaze activity, there is little work [Pfeil et al., 2018, Rubo and Gamer, 2018] to our knowledge. Previous work has shown that faces attract gaze in images or videos [Birmingham and Kingstone, 2009, Langton et al., 2008], which is not the case with tasks performed in the real world [Hessels

et al., 2020, Gallup et al., 2012, Laidlaw et al., 2011]. In order to investigate this issue in VR, Rubo and Gamer [2018] asked participants to sit in an VE where another person was carrying an object to sell in front of them. Data shows that gaze fixations were not mainly focused on the head of the virtual agent, which is consistent with tasks performed in real environments. In the work of Pfeil *et al.* [2018], participants were seated in front of a board with several lights on it (Figure 2.21). In both environments, the lights turned on and off in the same order. Their results show that the fixations were similar, however, the coordination of the head and eyes were different. They found more head rotations in VR which might be explained by the reduced field of view from the HMD.

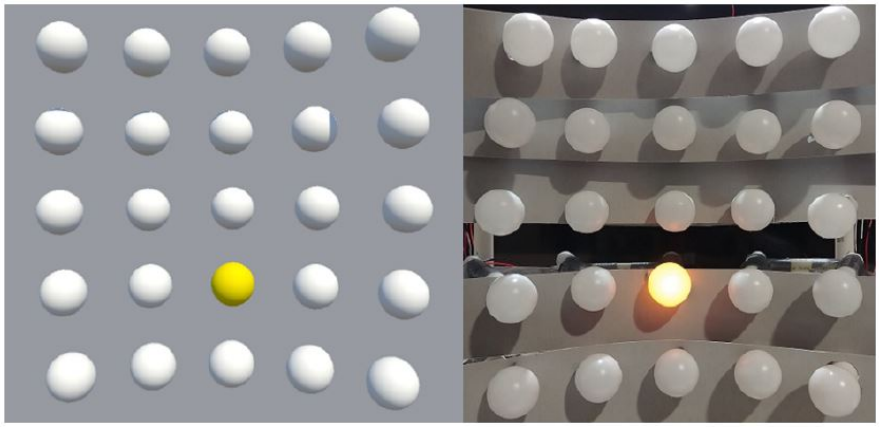


**Figure 2.21** – Board of light used during Pfeil *et al.* 's experiment [2018] in order to investigate the differences in gaze activity between real (left) and virtual (right) environment.

All these experiments, therefore, show that human behaviour is similar in a real or virtual environment, although some quantitative differences exist. The following section presents the different experiments carried out in VR in order to study human behaviour during walking, specifically when interacting with other virtual agents.

## 2.3.3 Experiment

### 2.3.3.1 Optical Flow

Since vision is the main perceptual system used when walking, several studies have focused on manipulating optical flow in VR to explore the impact on locomotion. In their work, Warren *et al.* [2001] conducted an experiment where participants walked towards an objective (a cylinder) in a VE. During the experiment, not only the vision but also the optical flow was manipulated by adding for instance

a texture on the floor or on the ceiling. Results show that whatever the condition participants walked toward the target, however, the addition of information in the optical flow helped to reduce the heading error toward the target. This indicates that during locomotion we both use the optical flow and the direction toward the goal. In addition, Chou *et al.* [2009] performed a similar experiment on the impact of optical flow on locomotion with different participant age groups. Their results show that the integration of optical flow information during a locomotion task is similar and therefore not dependent on age.

#### 2.3.3.2 Pairwise Experiment

As with the experiments in a real environment, many studies conducted pairwise (Section 2.2.1.1) VR experiments while manipulating the visual clues of the virtual obstacle that participants had to avoid. Bailenson *et al.* [2003] conducted an experiment where either participants had to walk in a room towards a virtual agent, or the virtual agent was coming toward them. They manipulated several variables such as the agent gender, whether the agent looked at participants or not, the crossing angle, and if the virtual agent was controlled by another human or not. They found that the clearance distance was higher when the virtual agent came in front of participants, as well as when mutual gaze occurred. In Rio *et al.* 's experiment [2014], participants had to follow an agent that randomly changed speed or a cylinder which visual angle was manipulated. Their results indicate that following behaviour aims to nullify the leader's optical expansion (change in visual angle) which finally indicate that the object followed stays at the same distance and direction. Several studies [Varma et al., 2017, Nummenmaa et al., 2009] also investigated the impact of gaze or head direction of a virtual agent during collision avoidance (Figure 2.22). Results show that participants tend to move in the opposite direction to the direction indicated by the agent's head or gaze. In addition, in their experiments Nummenmaa *et al.* [2009] recorded participants' gaze and found that their gaze pointed in the direction they were taking to avoid.

In the same vein, other research [Lynch et al., 2018, Mousas et al., 2019, Wang et al., 2019, Capozzi et al., 2018] focused on the impact of mutual gaze on participants' behaviour during a collision avoidance task in VR. For instance, in Lynch *et al.* 's work [2018], participants were immersed in a CAVE, and a virtual agent was coming from a 90° crossing angle. They found that there was no significant difference in participants' locomotion (mpd) and clearance distance. In the study of Moussa *et al.* [2019], participants had to walk to a goal while avoiding

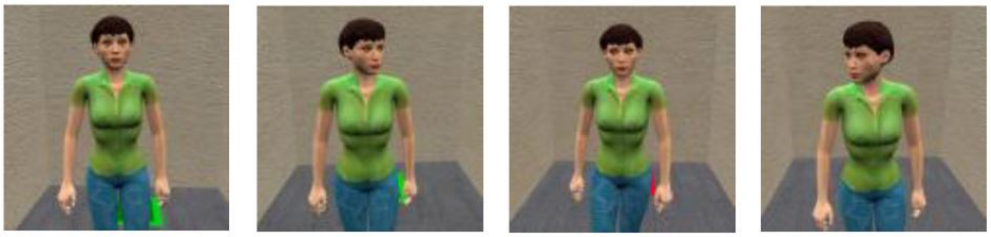


**Figure 2.22** – Example of several scenarios used during Nummenmaa *et al.* 's experiment [2009]: agent's head and eyes facing forward (left), agent's head and gaze towards participant's right (middle-left), agent's head forward and gaze toward participant's left (middle-right), agent's head and gaze toward participant's left (right)

a static virtual agent. Two variables were manipulated, either no representation for participants or the use of a self-avatar and virtual agent looked or not at the participant during the task. Their results showed that during tasks when the virtual agent looked at participants, the length and deviation of the path and the duration of the trial were greater. Furthermore, the presence of a self-avatar also impacted the participants' locomotion as participants took again a longer path. The difference of results between these two studies can possibly be explained by the differences in VR displays or the difference of task. Indeed in one experiment, the virtual agent was moving whereas in the other study it was static, furthermore one used an HMD and the other a CAVE, which change the size of the field of view. Other studies [Wang et al., 2019, Capozzi et al., 2018] also demonstrated that a shared gazed from at least two persons could lead to joint attention with another walker encountered. Finally, Lynch *et al.* [2017] performed a pairwise experiment with a virtual agent coming from a 90° crossing angle. During this experiment, they manipulated the visual representation of this agent which was either global cues (center of gravity represented by a sphere or a cylinder) or local cues (full human body, trunk, or legs), as illustrated in Figure 2.23. Participants displayed similar motion adaptations in order to avoid the agent for all conditions, however, clearance distance was lower for global cues than for local cues.

#### 2.3.3.3 Multiple-interaction

Virtual reality provides a high degree of control over the experimental conditions, especially over the virtual agents present in the VE. For this reason, many researchers have been interested in studies involving multiple interactions in VR. Silva *et al.* [2018] performed an experiment where participants had to walk toward a goal and avoid three incoming agents. These agents were represented either by

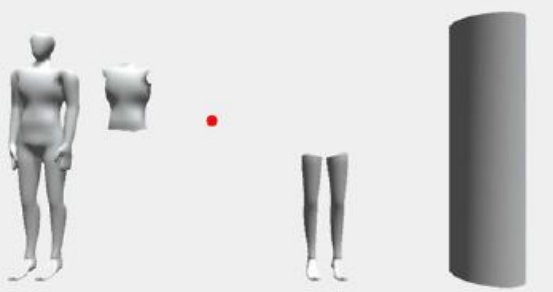


**Figure 2.23** – Illustration of the global (center of gravity represented by a sphere or a cylinder) and local (full human body, trunk or legs) visual cues used during Lynch *et al.* [2017] experiment.

a cylinder, full body, or full body with footstep sound. Clearance distance was higher when avoiding cylinder than full body agents, which is consistent with Lynch *et al.* [2017]. Furthermore, the presence of footstep sounds had no effect on the participants' locomotion, which indicates that collision avoidance relies more on visual cues than audible cues in this situation. Other researches [Bönsch et al., 2018, Huang and Wong, 2018, Volonte et al., 2020] have also been done on the influence of the emotions displayed by pedestrians (Figure 2.24) when navigating in a crowd. Virtual agents in the crowd displayed different positive, neutral or negative emotions such as happy, angry, sad which impact participant behaviour. When pedestrians exhibited positive emotions, the size of the participants' PS decreased [Bönsch et al., 2018], participants interacted more with pedestrians [Huang and Wong, 2018] and they took more time to perform tasks in the VE [Volonte et al., 2020]. Furthermore, participants reported that emotional virtual agents were more realistic [Huang and Wong, 2018]. Finally, Lobera *et al.*[2010] conducted a VR experiment where participants' skin conductance was recorded as a metric to assess their physiological arousal to the approach of one or more virtual agents, and showed that participants exhibited greater physiological arousal when the agents were close to them.

The impact of crowd density on human behaviour has also been investigated in VR. In their work, Dickinson *et al.* [2019] asked participants to move objects between tables in a crowded room with different densities (low, medium, high). Using Positive Affect Negative Affect Schedule (PANAS) [Watson et al., 1988] as a metric, they showed that a high crowd density has a negative effect on participants and impacts their velocity. Koilias *et al.* [2020], on the other hand, conducted an experiment where participants had to navigate through different types of crowds. In this experiment, they manipulated the density of the crowd but also its speed and direction. They found that crowds with a low density or



**Figure 2.24** – Illustration of positive and negative facial expression of virtual agents used during Volonte *et al.* 's experiment [2020].

a high speed increased participants' speed, and that crowd density and direction had an impact on participant's deviation and participant's path length.

With the technological advances on eye-trackers, several researchers have performed studies on gaze activity during VR navigation tasks. Jovancevic *et al.* [2006] asked participants to walk in VR among a few virtual humans and either to follow or avoid them. They found that the distribution of gaze fixations in the environment depends on the nature of interactions with virtual humans, i.e. they focus on following rather than on avoiding. This result is consistent with previous work showing that gaze is dependant on the task. More recently, Meerhoff *et al.* [2018a] conducted an experiment where participants navigated in a virtual crowd. They demonstrated that the participant's gaze is attracted to pedestrians with the highest risk of collision when walking in a virtual crowd.

Several scientists have also performed experiments on pedestrian interactions in VR for direct applications to crowd simulation. In Bruneau *et al.* 's experiment [2015], participants encountered a group of pedestrians more or less wide (interpersonal distance between pedestrians and the size of the group, as illustrated in Figure 2.25). They analysed participants' trajectories in relation to the Principle of Minimum Energy (PME) in order to quantify their decision threshold for going around or through, with the goal of applying such thresholds in crowd simulators. Furthermore, they also investigated the impact of group appearance in the decision process and observed an influence in trajectories when the appearance of the virtual agents were soldiers but not when they were zombies. A plausible explanation for this result is that soldiers are characters that can actually be encountered. Rio *et al.* [2014, 2018] investigated the influence of neighbours on human locomotion. They conducted experiments where participants walk with a

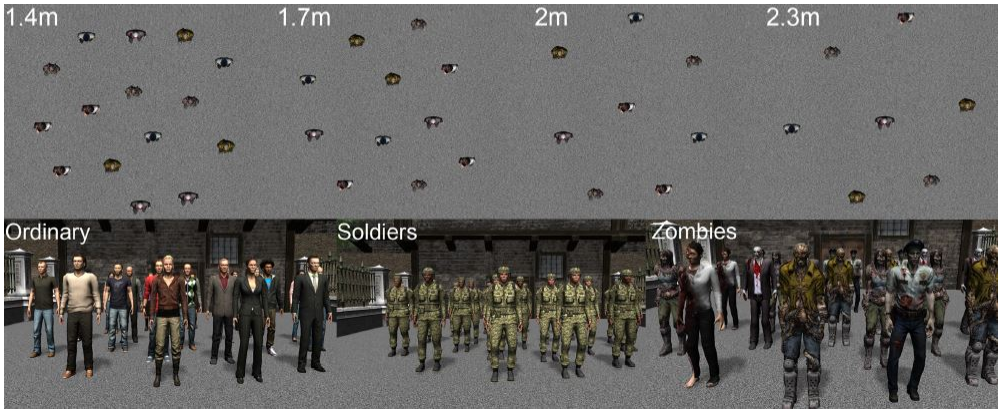



**Figure 2.25** – Illustration of initial condition for Bruneau *et al.* 's experiment [2015]: example of group size and interpersonal distance between pedestrians (top row) and group appearance (bottom row).

few or several virtual agents near them, then at some point, some agents in the crowd changed their direction (Figure 2.26). Results showed that the influence of neighbours can be modeled as a linear decay function of distance, which does not depend on the eccentricity of the other walkers within the participants' field of view. From this experiment, they represent the neighbourhood of interaction as a semi-circle with a radius of $5m$ and in which the contribution of the agents is proportional to the distance. Finally, Rios and Pelechano [2020], asked participants to walk in a virtual train station while an evacuation situation occurred. They manipulated different stress levels, such as an alarm ringing or a fire starting. In all conditions, participants tended to follow the crowd to reach the emergency exits regardless of the stress stimulus used. This behaviour was accentuated by the number of people in the crowd running towards the exit. Furthermore, results showed that the general stress level in the virtual scenario seemed to be dependent on the virtual human behaviours.

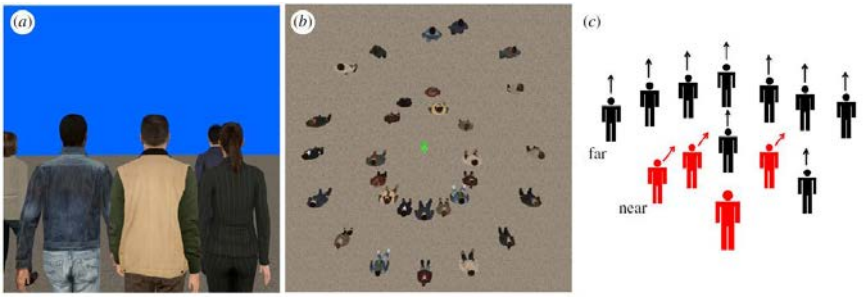



**Figure 2.26** – Illustration of Rio *et al.* [2018] experiment, with participant view in VR (left), aerial view of the virtual environment (middle) and diagram of a heading perturbation (right).

With the various technological advances in VR in recent years, this tool has been progressively used more and more to study human behaviour. VR is of great interest for this kind of research because of its control of the experimental environment and the manipulation of objects present in it. Several studies have provided a comprehensive assessment of some of the biases induced by VR during navigation tasks with other pedestrians. Subsequently, there has been recent research on the impact virtual agents' visual cues on participant's gaze activity during navigation task, and there has been also studies on the analysis of human behaviour in order to improve crowd simulators. However, there is little work on the coupled analysis of walking and eye activity during interactions with pedestrians.

## 2.4 Conclusion

In this chapter, we first introduced the state of the art for crowd simulation. In particular, we focused on the agent-based approaches which are composed of a local interaction model and an interaction neighbourhood. Scientists have been particularly interested in modelling pedestrian interactions and have thus implemented a variety of mathematical models. However, there is very little research on interaction neighbourhood models. In order to improve or develop new agent-based approaches, experiments on locomotion during pedestrian interactions have been conducted in laboratories or outdoors. These experiments were first focused on the analysis of participants' trajectories and then on the additional analysis of the participants' gaze activity. Indeed, during walking, vision is the primary perceptual system used. However, many elements present in the environment can have an impact on participants' locomotion or gaze activity. In order to overcome this, VR has become more and more used as an experimental platform to control the variables to be studied while preserving the same stimuli for all participants during an experiment.

We have therefore shown that the coupled analysis of locomotion and gaze can be very interesting in order to understand human behaviour. In particular, it can provide answers as to the neighbourhood of interaction when navigating in a crowd. Considering the experimental conditions on this topic, VR seems to be the most appropriate tool to perform such experiments. However, VR also introduces some biases into human behaviour, which have been studied intensely for locomotion but very little for gaze activity.

Our aim in this thesis is therefore first of all to implement an experimental platform to study human behaviour when interacting with other pedestrians in real and virtual environments. Using this platform, we will present and evalu-

ate the impact of VR on gaze activity, first during an interaction with a single pedestrian and then during more complex situations such as navigating in a virtual crowd. Following this assessment, we are interested in the impact of specific crowd properties (density, direction) on gaze activity and locomotion. However, as we want to immerse participants into denser crowds, we will need to simulate physical contacts as they are key features in social interactions. To this end, we will design an experiment to simulate them in VR and evaluate their impact on participants' locomotion.

Chapter

# 3 Experimental Tools

## Contents


**3.1 CrowdMP . . . . . 44**

3.1.1 Experimental Procedures . . . . . 45

3.1.2 CrowdMP Architecture . . . . . 46

3.1.3 Virtual Reality Displays . . . . . 49

**3.2 Tobii Tools . . . . . 50**

3.2.1 TobiiController . . . . . 50

3.2.2 Tobii Python Tool . . . . . 53


This chapter presents the different tools implemented and used in this thesis to design the experiments we conducted. The first tool, **CrowdMP** (Section 3.1), relies on Unity and implements a whole architecture of objects and scripts to simplify the creation of experimental experiences. In particular, this platform makes it possible to easily create virtual environments populated with pedestrians driven by a crowd simulator by adding prefabricated objects in Unity environment. It is also possible to immerse participants in it with different virtual reality setups by choosing the type of display for the main camera. This platform was implemented by Julien Bruneau, Fabien Grzeskowiak, and myself, with my participation being mainly on the addition of the plugins regarding the gaze activity in VR, the offline replay of experiments from the data recorded to perform complex computations, and the animations of virtual characters to have more realistic behaviours

from them. Then Section 3.2 describes eye-tracking tools designed to record and analyse the gaze activity of participants and what they looked at in a real environment. It contains a software implemented by Julien Bruneau in Unity managing the recording of an experiment with the Tobii glasses, and a set of python scripts implemented by myself for analysing the gaze data recorded.

## 3.1 CrowdMP

CrowdMP is a tool based on Unity game engine and whose main purpose is to ease the implementation of experiments that require any kind of crowds. It provides a basis to immerse participants in homemade virtual environments populated with custom virtual crowds, as displayed in Figure 3.1. In addition, it also provides many tools/plugins to customize as much as possible the experiment so that it can be used for a wide variety of purposes. To this end, it offers many different features that can be useful when running experiments such as user controls, trials management, data recording, instructions presentation, etc. It also offers features specific to running experiments with crowds such as a variety of character models, animations, and behaviours. Finally, it was developed to ease the addition of new features, to be adaptable to a wide range of future experiments.

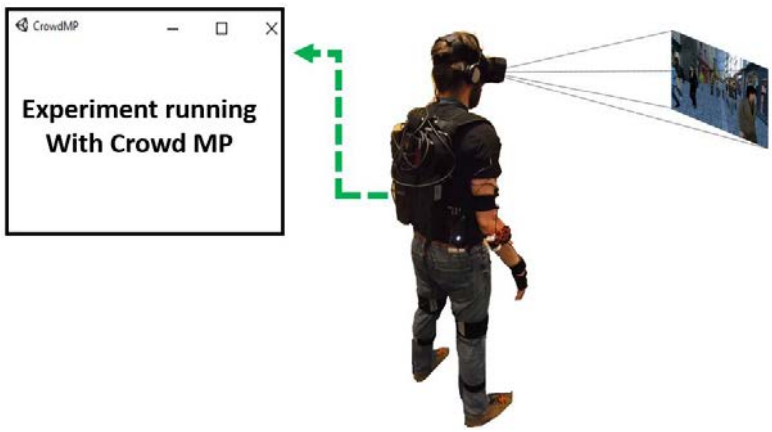



**Figure 3.1** – Illustration of participant who performs an experiment in VR with CrowdMP running on the computer located on the participant's back.

The remaining of this section presents the experiment procedures (Section 3.1.1) as well as the different components that are present in this tool (Section 3.1.2) and the list of VR devices supported by CrowdMP (Section **??**).

### 3.1.1 Experimental Procedures

An experiment designed using CrowdMP includes one or several trials that participants will have to complete. During this trial, CrowdMP loads for each trial the associated virtual scene and the elements necessary in order to run it. Furthermore, the order of the trials that participants have to complete is decided by the users. Therefore, a custom file architecture is used to represent an experiment as illustrated in Figure 3.2. For each participant is associated a *CSV* file that contains, in an order chosen by the user, a list of path to *XML* files. Each *XML* file represents a trial with all the information necessary for CrowMP to loaded and ran it, with some of these elements mandatory for all trials and other optional.

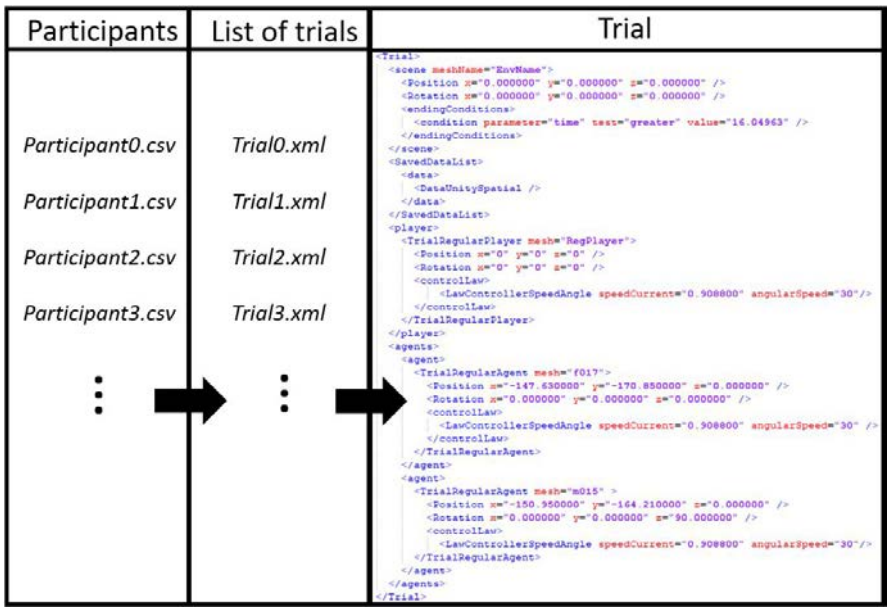



**Figure 3.2** – Organisation of participant and trial files used in CrowdMP, and example of a trial XML file.

- **Scene** (mandatory): Information about the virtual scene to load, with the name of the object in Unity, the position and rotation where it has to be load and the condition to finish the trial.
- **Recorder** (optional): Information about which kind of data will be recorded such as position and rotation of participants and all virtual agents, as well as gaze data or even animation.
- **Participant** (mandatory): Information about participant's starting position and rotation, the 3d model representing the participant in the virtual

environment (if needed), and also the control law that will drive his/her navigation.

- **Agents** (optional): List of all the agents present during the trial. For each of them, information about their starting position and rotation, the 3d model representing them, and the control law to drive their navigation

### 3.1.2 CrowdMP Architecture

CrowdMP is composed of two main blocks as illustrated in Figure 3.3 which provide a list of scripts, objects and tools in Unity to design with ease an experiment. The first block (in blue) contains the basic components that compose the core of CrowdMP and the second block (in green) contains every plugin added by current or previous users.

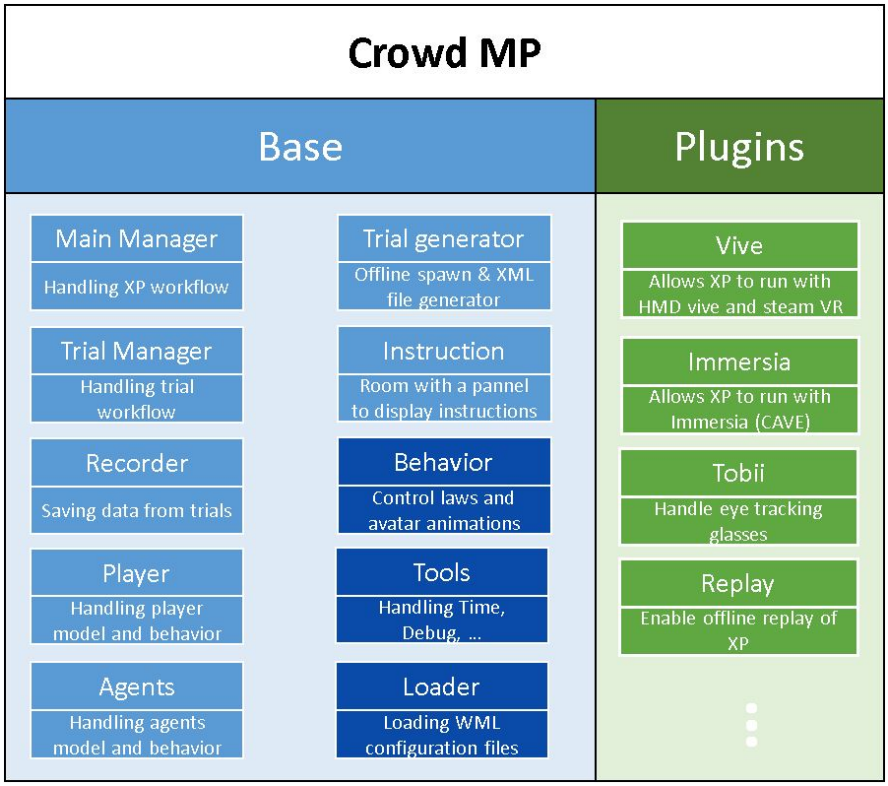



**Figure 3.3** – Code organization of the platform

In order to create an experiment, users have to interact directly with some of these components by selecting the one they need and putting them in the Unity scene as shown in Figure 3.4. Some of these elements are mandatory for an experiment, others are optional and depend on the experiment.




**Figure 3.4** – Example of the elements used to designed our second experiment [Berton et al., 2020].

- **MainManager** (mandatory): This script handles the overall management of the entire experiment. In particular, it manages the initialisation of each trial that constitutes the experiment (loading the XML files and the various assets), the transition between them and the end of the experiment. It is moreover possible to attach to this object a transition and end screen which will be displayed at the desired time.

- **TrialManager** (mandatory): This script controls the proceeding of a Trial, the update of each element during the trial and check when the ending condition happens. Each unity object that represent a virtual environment to be loaded must have a **TrialManager** associate to it.

- **Recorder** (optional): This component can be added to the scene object with the **TrialManager**, in order to enable the recording of data. The type of data to record has to be specified in the trial XML file as explained in Section 3.1.1.

- **Player** (mandatory): All the player models that have to be loaded during the experiment. Each model has a controller and a behaviour for the participant, but also an avatar to represent him if necessary.

- **Agents** (optional): All the models of the virtual agents that have to be loaded during the whole experiment with scripts to control them (navigation, behaviour and animation).

- **TrialGenerator** (optional): An object to generate a trial XML file from

the environment in Unity.

- **Instruction** (optional): A room which can be used to display instructions to the participants at the beginning of the experiment.

Furthermore, components in dark blue in Figure 3.3 are elements with which a user does not have to interact directly in Unity environment and that should not be modified. They include tools for managing time, debugging and a library to parse and load xml files, but also scripts to manage agent and participant behaviours, such as several control laws for navigating in the virtual environment and the animation of avatars. As for the block with the plugins, it contains all the scripts and objetcs implemented by users to satisfy their specific needs and customize their experiment. In this one one will find for instance components allowing to use different HMDs (e.g, Vive, Fove) or a CAVE (Immersia) but also a plugins to record gaze activity when a eye-tracker is present in one of this kind of displays. In addition, a plugin has been created to be able to replay an experiment, offline from the recorded data during this one. This plugin is particularly useful when it is necessary to perform complex operations. For instance, it possible to replay participants gaze activity with cameras computing segmentation of elements present in the environment and depth map. Finally a general overview of CrowMP main loop workflow is displayed in Figure 3.5, with the order to update each elements and how they synchronize together.

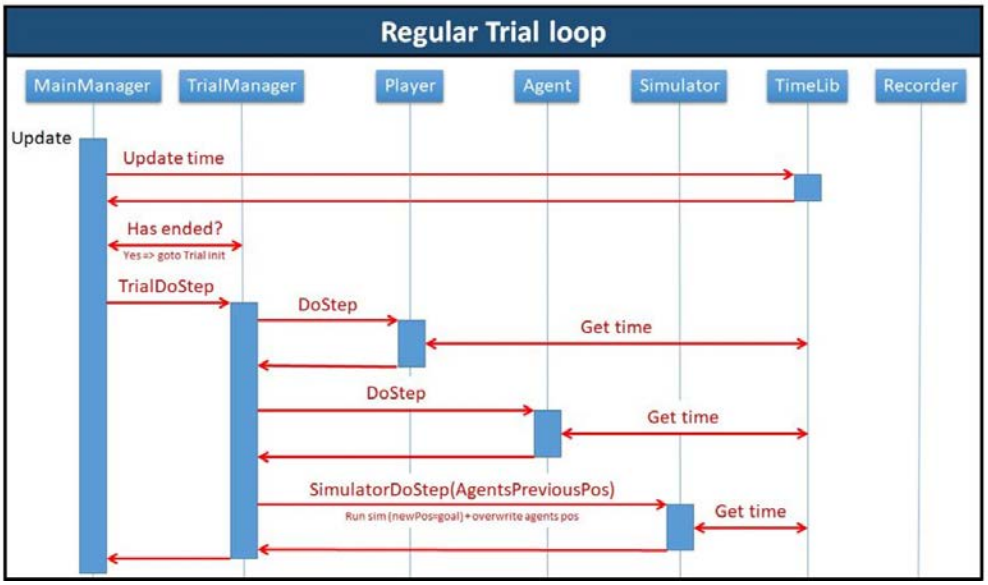



**Figure 3.5** – Workflow during a regular trial main loop

### 3.1.3 Virtual Reality Displays

The CrowdMP platform supports several VR devices, with the addition of the plugins which contain the Software Development Kit (SDK) associated with these VR devices.

- **HTC Vive:** It is one of the most well known and widely used VR devices (Figure 3.6-a). Its latest version offers a field of view of 110° with a resolution of 1440×1660 pixels, a framerate of 90Hz and a tracking system that enables to move in a space up to 10m × 10m. In addition, a version with an eye tracker (ViveEyePro) was released in 2019, which records the gaze at a frequency of $120Hz$.
- **Fove:** Release in 2015, it is one of the first VR HMD (Figure 3.6-b) with an integrated eye tracker which records the gaze at a frequency of $120Hz$. It offers a field of view of to 100° with a resolution of 2560×1440 pixels and a framerate of 70fps. In addition, while this HMD is shipped with a single tracking camera (i.e., with an extremely limited tracking space), our plugin in CrowdMP allows to track the FOVE positional information using several motion tracking systems (e.g., Vicon, Qualisys).
- **Pimax:** This HMD (Figure 3.6-c) is very recent and offers a 200° field of view with several resolutions according to the chosen models (5K: 2560×1440 pixels or 8K: 3840×2160 pixels) and a framerate of 144fps. Moreover, it offers a great freedom for the tracking system of the HMD, as it is both compatible with the Steam VR basestations (as for the **HTC Vive**) and with other types of motion tracking systems (e.g., Vicon, Qualisys).
- **Immersia:** This VR device is a CAVE (Figure 3.6-d) using a 4-screen Computer Assisted Virtual Environment. It is equipped with 13 projectors, has a 15MPixels resolution in total, and is 9m large, 3m high and 3m deep.



a) b) c) d)

**Figure 3.6** – From left to right list of virtual reality devices supported by CrowdMP: HTC Vive (a), Fove (b), Pimax (c) and Immersia (d).

## 3.2 Tobii Tools

This section presents first the **TobiiController**, an interface created with Unity in order to control the Tobii Glasses 2 (Section 3.2.1). This pair of glasses, shown in Figure 3.7, records the movement of the eyes (in 2D and 3D) of a person wearing them while also recording what that person sees. Furthermore, an interface is given with this eye-tracker allowing to create projects with participants and recordings but also to calibrate the eye-tracker before its use. Then Section 3.2.2 describes all the python tools to analyse the Tobii glasses data to study gaze activity

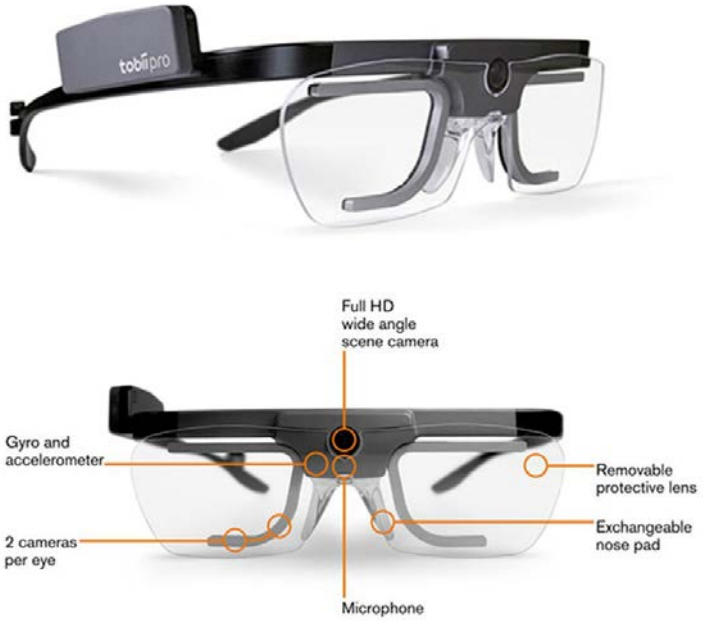



**Figure 3.7** – Tobii pro glasses 2, a pair of glasses that record the gaze activity of a person wearing them and what he/she is seeing.

### 3.2.1 TobiiController

The **TobiiController**, illustrated in Figure 3.8, is an interface that contains most of the features of the Tobii Pro Glass Controller that comes with the glasses, but also implements new ones. In particular, it includes the management of event that can be triggered by the user during a recording which can be very useful when users want to notify that a specific action occurred. In addition, this platform also enables users to indicate whether part of the recording consists on a validation

step of the eye-tracker, which is a crucial element of any gaze activity studies. Indeed, this step helps to validate if the eye-tracker records accurately the eye movements by asking users to follow a specific object with their eyes. Finally, the connection to the Tobii Glasses with this interface is by WiFi.

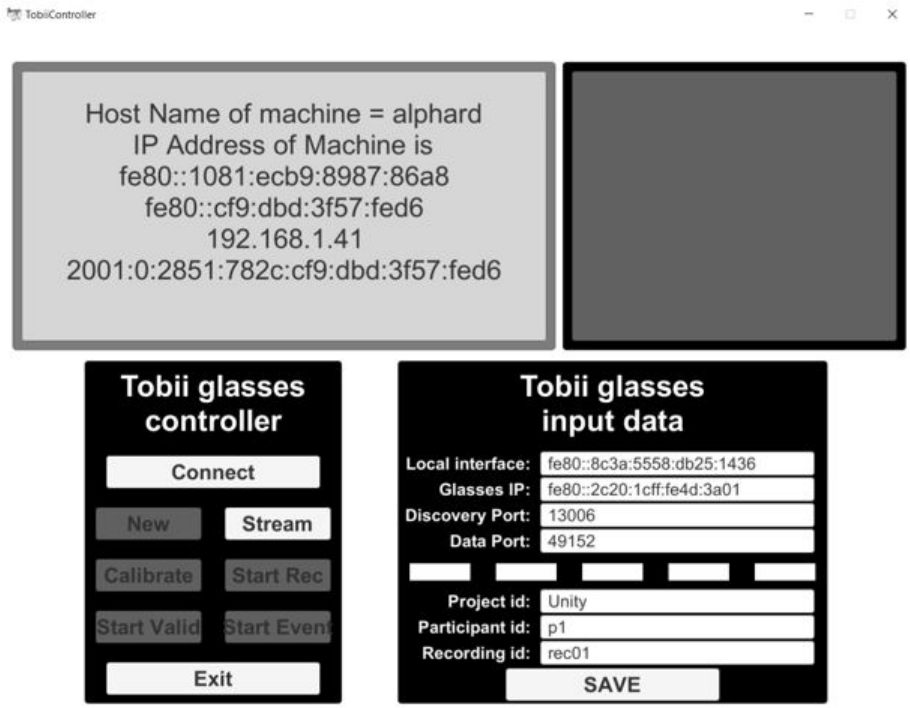



**Figure 3.8** – Tobii Controller Interface

This platform uses several parameters in input:

- Local Interface: IPv6 address of your computer (can be found using the command line "ipconfig" in the command prompt in windows).
- Glasses IP: IPv6 address of your glasses (can be found with the Tobii Pro Glass Controller).
- Discovery Port: port used to scan and find the Tobii glasses.
- Data Port: Port used to stream the data and send requests to the Tobii glasses.

Then several buttons are available in the Tobii Controller interface in order to manage a recording. The actions of these buttons are detailed here:

- Connect: Connexion to the Tobii glasses

- Stream: Activate the streaming of the data sent by the Tobii glasses
- New: Create a new recording for a participant in the right project
- Calibrate: Start the calibration of the Tobii glasses. The results of the calibration (either successful or unsuccessful) will be indicated in the top left screen.
- StartRec: Start the recording, this button will stay red until it is pressed again, which will terminate the recording.
- StartValid: Start the validation, this button will stay red until it is pressed again, which will terminate the validation.
- StartEvent: Start an event, this button will stay red until it is pressed again, which will terminate the event.

To use properly this interface, it is necessary to activate the buttons in the right order (see Figure 3.9). First of all, the connection to the glasses must be established (button Connect). Then the data must be sent from the glasses to the interface (button Stream). From this moment, the other buttons will be available. If the project, the participant or the name of the recording does not exist, it has to be created (button New). Then it is strongly recommended to perform a calibration before starting any recordings (button Calibrate). If all these steps have been checked, it is possible to start a recording (button StartRec).

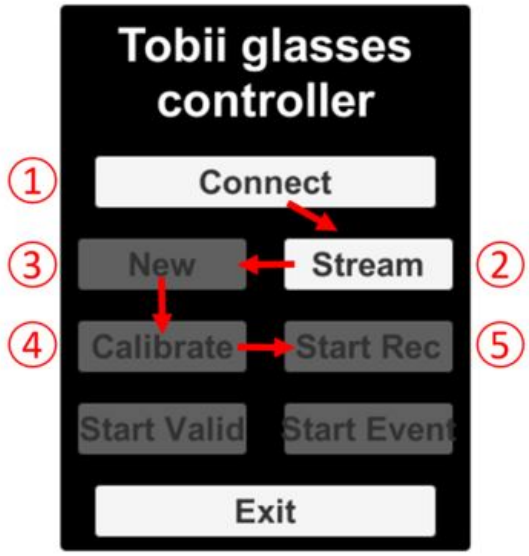



**Figure 3.9** – Buttons order to start a recording

For the “Start Valid” and “Start Event”, by pressing them, a tag will be insert in the recording data. This tag will indicate when the validation or an event starts and ends during the recording. Then it will be possible to split the recording and extract only these events using the python tools described in the next section.

### 3.2.2 Tobii Python Tool

In order to be able to extract and analyse the data recorded with the TobiiController, numerous python scripts have been implemented which can be divided into four categories.

- **Tobii Data Extraction**: This package of scripts can be used to parse the data recorded with the Tobii Controller, extract the content to recreate the projects with the participants and the trials performed by each of them. In addition, each trial is split using the events and calibration if any occurred, as shown in the Figure 3.10.

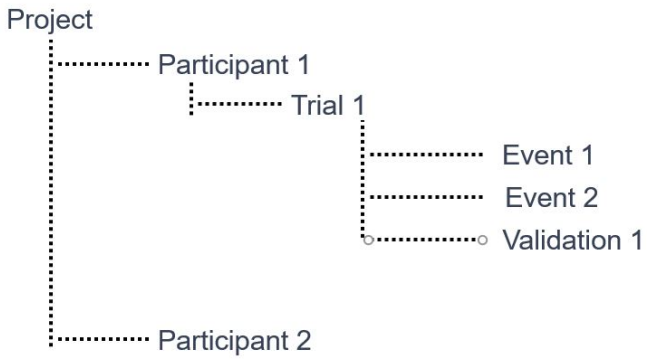



**Figure 3.10** – Architecture of the data extracted from the recordings of trials with the TobiiController.

- **Gaze Activity**: This collection of scripts enables the analysis of gaze data and the computation of fixations. Specifically, two methods (**I-VT** and **I-AOT**) from the literature [Salvucci and Goldberg, 2000] are implemented. These methods use the pixel coordinates of gaze points, however our goal would be to also implement methods using 3D gaze coordinates will also be implemented.

- **Object Recognition**: An implementation from internet of **YOLO** neural network [Redmon and Farhadi, 2018] is present in this set of tools. This algorithm is used to detect in an image various objects belonging to the 80

classes of the COCO database [Lin et al., 2014]. In future works, we are going to add implementation of algorithms that perform pose estimation such as DeepPose [Toshev and Szegedy, 2014].

- **Videos**: A package of scripts to create videos with the computed fixations and objects detected in the participants' videos, as shown in the Figure 3.11.




**Figure 3.11** – Illustration of a recording from a participant wearing the Tobii Glasses. The current (red) and precedent (blue) fixations are displayed with also all the objects detected (black bounding boxes) by the algorithm Yolo [Redmon and Farhadi, 2018] with the name of their classes.

The next chapters present the experimental studies that have been created using these two platforms and which aim at analysing locomotion but more importantly participants' gaze activity during interactions with pedestrians in a real and a virtual environment.

Chapter 4

# Influence of Virtual Reality Setup on Gaze Activity during Collision Avoidance

## Contents


**4.1 Methods . . . . . 57**
4.1.1 Participants . . . . . 58
4.1.2 Task . . . . . 58
4.1.3 Environment . . . . . 58
4.1.4 Experimental Design . . . . . 59
**4.2 Analysis . . . . . 61**
4.2.1 Collected Data . . . . . 61
4.2.2 Time Variables . . . . . 62
4.2.3 Collision Avoidance Trials . . . . . 62
4.2.4 Kinematics of the Collision Avoidance Task . . . . . 63
4.2.5 Gaze Behaviour . . . . . 64
4.2.6 Statistics . . . . . 65
**4.3 Results . . . . . 66**
4.3.1 Locomotion . . . . . 66
4.3.2 Gaze Behaviour . . . . . 67
**4.4 Discussion . . . . . 70**
4.4.1 Collision Avoidance Behaviour . . . . . 70

4.4.2 Fixations and Gaze Allocation . . . . . 71
4.4.3 Gaze and Head Angles . . . . . 72
4.4.4 Limitations . . . . . 73
**4.5 Conclusion . . . . . 74**

In this chapter we present the first scientific contribution of this thesis, which is about the impact of virtual reality (VR) on gaze activity during collision avoidance. From the analysis of the literature in Chapter 2, little is known about differences in gaze behaviour during a collision avoidance locomotor task in real and virtual environments. This is however a daily situation of interest to understand how humans interact together. Furthermore, as explained in Section 2.3, there are a multitude of various VR systems with fundamentally diverse properties. In particular, there are great differences regarding mobility or fields of view which can thus impact the behaviour of people immersed in VR. In addition, there are also some differences in the technologies built into VR systems to record the gaze, e.g., in terms of accuracy or robustness to lighting conditions. The objective of this chapter is therefore to evaluate and compare how different VR setups influence gaze behaviour during collision avoidance between walkers. To this end, we designed an experiment involving a collision avoidance task between a participant and another walker, during which we recorded participants' gaze using eye-tracking devices. For comparison, this experiment was performed in both real and virtual environments using different virtual setups (including a cave, a screen and a Head-Mounted Display), as illustrated Figure 4.1.




**Figure 4.1** – Conditions studied in this chapter. (a) Real situation, (b-e) Participants avoiding a virtual character while (b-c) wearing a HMD and (b) walking or (c) navigating using a game controller, (d) standing in a CAVE and navigating using a game controller, (e) interacting with a computer Screen.

In the context of collision avoidance task between walkers, the contributions of this work are as follows:

- We propose a methodology based on several objective criteria to evaluate both gaze and kinematic behaviours in virtual and real environments.

- We provide recommendations for the design of VR platforms to perform human locomotion studies in the context of collision avoidance between walkers.

The remaining of the chapter is organized as follows. Section 4.1 presents the general description of our experimental study. Section 4.2 describes the different criteria used to compare the trajectories and gaze behaviour. Finally, results are presented and discussed respectively in Section 4.3 and 4.4.

## 4.1 Methods

We designed an experiment to evaluate and compare gaze behaviour during a collision avoidance task. We considered a real baseline condition as well as four virtual conditions. Virtual conditions varied by the type of visual display (Cave, HMD or screen). As it has been demonstrated that participants' gaze can be impacted by action requirements [Dicks et al., 2010], we also considered two types of navigation techniques (physically walking or using a game controller). Our hypotheses are:

- **H1**: the nature of the visual information retrieved from the environment to achieve the collision avoidance task (i.e., gaze allocation) will be similar between the real and the virtual conditions, similarly to what was previously observed for locomotion behaviour between VR and real conditions. Moreover, we hypothesize that gaze allocation (e,g: time spend to look at elements in the environment) will not be affected by the type of VR display.

- **H2**: the type of display will however affect gaze movements. Since each type of display does not convey visual information in the same way (e.g., HMD and screen have limited field of view), we expect that gaze movements will adapt accordingly. Especially, we expect that displays with limited field of view will induce larger head (HMD) or eye movements (Screen because we used a chin rest) to explore the environment and compensate for the limited field of view.

- **H3**: the type of navigation controller will not affect gaze behaviour. Again based on previous studies that demonstrate that locomotion is performed similarly in real and virtual conditions in spite of the use of various locomotion techniques, we expect that the nature of the visual information to control virtual locomotion will remain similar, and therefore will induce similar gaze behaviour compared to the real baseline condition.

### 4.1.1 Participants

Seventeen unpaid participants, recruited via internal mailing lists among students and staff, volunteered for the experiment (6F, 11M; age: avg.=23.6$\pm$3.3, min=19, max=29). They were all naive to the purpose of the experiment, had normal or corrected-to-normal vision, and gave written and informed consent. The study conformed to the declaration of Helsinki, and was approved by the local ethical committee. Because of tracking issues that appeared specifically for five participants (motion capture tracking issues, or eye gaze calibration issues), only the data from twelve participants was finally used in this paper.

### 4.1.2 Task

Participants were asked to navigate in a real or virtual environment towards a target while avoiding any collision with another walker crossing their way (a virtual character or a real confederate). As shown in Figure 4.2, the other walker had an orthogonal trajectory to the one of the participant, he could come either from the right- or left-hand side, he walked straight at constant speed (with no adjustment of his trajectory). Participants reached a target in front of them, visible from the beginning. Walls hid the walker to participants at the beginning of the trial to let them reach their comfort speed before they react to the virtual walker. We varied the risk of collision with the other walker by defining 13 offset positions for the walker, that resulted into advance or delay on the participants' motion (6 giving advance, 6 delaying him and 1 symmetrical situation with full risk of collision). For each participant, the experiment was conducted in two sessions which altogether lasted between 1h30 and 2h.

### 4.1.3 Environment

Participants performed the task in a real and a virtual environment. The real environment was delimited by a $9m$x$9m$ square, with four walls placed along its the diagonal at $2.2m$ from the center (see Figure 4.2). They were instructed to reach the opposite corner of the square. Furthermore, the confederate walker did not perform eye contact and did not react to participants.

In comparison, virtual conditions took part in a virtual copy of the physical environment (dimensions, start and target positions, walls) which we designed in Unity 2017.4. The target was represented as a green cylinder. The virtual walker animated based on motion capture of the real confederate. Note that animations did not include gaze movement or facial expression to keep similar behaviour of

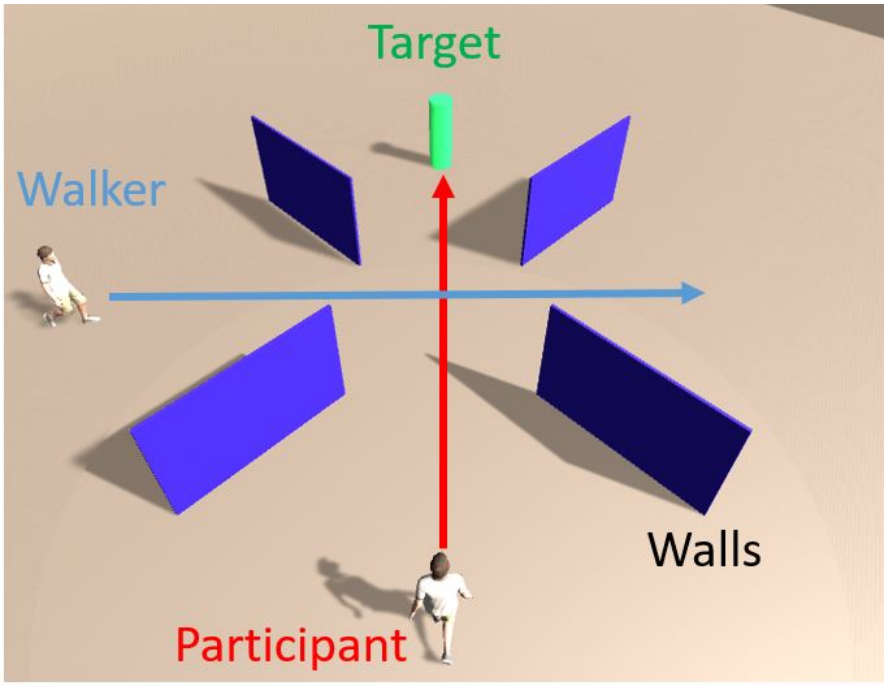



**Figure 4.2** – Participants were asked to navigate towards a target, while avoiding any collision with a real or virtual confederate crossing their path perpendicularly. Walls prevented participants from seeing the real or virtual confederate before reaching their comfort speed.

the walker with respect to the real condition.

### 4.1.4 Experimental Design

Participants were asked to perform the task presented in Section 4.1.2 in the five following conditions (summarized in Table 4.1):

**Real**: this is our baseline condition to compare VR situations to. Participants were asked to walk at comfort speed and to avoid any collision with the confederate walker. A 28-camera Vicon system (120Hz) was used to record their trajectory, where participant and confederate positions were approximated as the centre of their head using bike helmets equipped with reflective markers. Participants also wore Tobii eye-tracking glasses which recorded both their gaze behaviour ($50Hz$) and what they saw of the environment (scene camera: 25Hz, 90° field of view).

**HMD-W**: participants were immersed in the virtual environment using a FOVE HMD (70Hz, 100° field of view), which comes with an integrated eye-tracker[1]

[1]It is important to note that we used three different models of eye-trackers in this experiment

(100Hz). They performed the collision avoidance task with the virtual character by walking in the co-localised physical setup (Figure 4.1.b). As for the Real condition, the 28-camera vicon system was used to track participants' trajectories using markers positioned on the HMD, as well as to update the participants' viewpoint in the virtual environment.

**HMD-C**: participants were immersed in the virtual environment using a FOVE HMD (see above), which was also used to record their gaze behaviour. However, in this condition we were interested in a different navigation technique, where participants used a game controller (Logitech F710 S) to avoid collisions with the virtual character, while seated at a desk. With no action on the controller, participants moved straight with a default speed of $1.33m.s^{-1}$ (comfort walking speed from [Bohannon, 1997]). They were able to linearly adapt their speed from $0.8m.s^{-1}$ to $2m.s^{-1}$ with the longitudinal axis and their angular speed from $-25deg.s^{-1}$ to $25deg.s^{-1}$ with the lateral axis of the controller. This control scheme was previously validated by Olivier *et al.* [2017] for such collision avoidance tasks.

**Cave**: participants were immersed in the virtual environment using a 4-screen Computer Assisted Virtual Environment (CAVE). the system is equipped with 13 projectors, has a 15MPixels resolution in total, and is 9m large, 3m high and 3m deep. Participants wore volfoni 3D glasses for active stereo vision, as well as the eye-tracking Tobii glasses, assembled with a custom-built rig. Participants were standing in the middle of the CAVE and asked to avoid collisions with the virtual character using a joystick (Logitech Attack 3), which used the same control scheme described above.

**Screen**: participants sat at a desk in front of a 24-inch screen, and were asked to avoid collisions with the virtual character using a joystick (Logitech Attack 3), which used the same control scheme described above. Gaze behaviour was recorded using The Eyetribe (60Hz) positioned under the screen, and a chin rest was used to increase tracking accuracy.

Participants performed the experimental task for each of these conditions. This

(Tobii glasses, Eyetribe desktop eye-tracker, and FOVE built-in eye-tracker) to fit the constraints of the VR devices. E.g., it was not possible to use eye-tracking glasses in the HMD, and we therefore used the built-in device.

**Table 4.1** – Summary of the conditions presented in this experiment

| Condition | Eye-tracker | Navigation |
|---|---|---|
| Real | Tobii | walking |
| HMD-W | Fove (built-in) | walking |
| HMD-C | Fove (built-in) | game controller |
| CAVE | Tobii | game controller |
| Screen | Eyetribe | game controller |

was combined with 13 levels of initial risk of collision with the confederate or the virtual walker. Finally, we included two repetitions of each situation, where the walker to avoid came either from the left or from the right. In total, each participant performed 130 trials (5 Conditions $\times$ 13 Risk levels $\times$ 2 Repetitions). They were also invited to perform up to 20 additional training trials at the beginning of each condition (most participants performed 4 to 5 in each case).

## 4.2 Analysis

### 4.2.1 Collected Data

During the experiment, participant and walker trajectories were recorded, as well as participants' head rotation and gaze behaviour (origin and direction). The image of what the participant saw (Figure 4.3 a-b) was recorded either with the Tobii glasses for the real environment, or with Unity for the virtual environment. This image was divided into three different items that could be targeted by the gaze (**Walker**, **Target** or **Environment**). This segmentation (Figure 4.3 c-d) was done either using shaders in Unity for the virtual environment, or manually with the help of a CNN network [Caelles et al., 2017] on the gaze video for the real environment. Furthermore for each trial, we re-sampled the recorded data (trajectories, eye movements) at a frequency of $60Hz$.

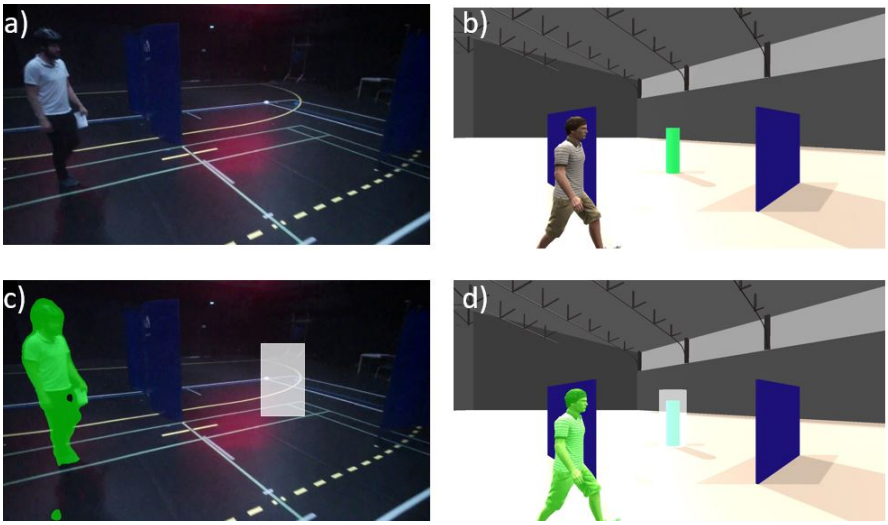



**Figure 4.3** – Real and virtual environments (a-b) and their respective segmentation (c-d). **Walker** is in green, **Target** in white and the rest is considered as **Environment**

### 4.2.2 Time Variables

To describe the interaction period[2], we defined two variables: the moment when the confederate walker is not obstructed by the wall anymore, which corresponds to the first instant when the participant can first see him ($Tsee$), and the moment when the two walkers are the closest to each other ($Tcross$). We then normalized the time along this interaction period from 0% ($Tsee$) to 100% ($Tcross$).

### 4.2.3 Collision Avoidance Trials

AS our objective is to study gaze behaviour during a collision avoidance task. We therefore needed to distinguish trials where a collision avoidance was performed (i.e., the initial risk of collision was high enough to trigger collision avoidance manoeuvres) from those that did not require trajectory adjustments. To this end, we based our analysis on the minimum predicted distance ($mpd$) [Olivier et al., 2012] explained in Section 2.2.1.1

More specifically, these two types of trials (with or without avoidance manoeuvres) will differ in the evolution of $mpd(t)$. Trials with trajectory adjustments to avoid collision typically have a significant increase of $mpd$ in time, i.e.,

[2]The term ”interaction“ has been used in previous work to provide a temporal description of the task. Even though the virtual character is not reactive, and thus, the interaction is not really present in our experiment, we decided to keep calling this period ”interaction“ for the sake of clarity

$mpd(Tsee) < mpd(Tcross)$, while the latter type of trials typically has a more constant $mpd$, with the exception of the effect of trajectory noise (e.g., swaying in locomotion), i.e., $mpd(Tsee) \approx mpd(Tcross)$. As a result, a threshold $mpd_{CA}$ on the value of $mpd(Tsee)$ over which Collision Avoidance is not necessary anymore can be identified in order to divide our data into trials with collision avoidance and trials without collision avoidance. This threshold was automatically computed by fitting the following model describing the evolution of $mpd(Tcross)$ in relation to $mdp(Tsee)$ on our data (see Figure 4.4 for illustrative representation):

$$\underset{a, mpd_{CA}}{\text{minimize}} \sum_{i}^{N} (f(Tsee^i) - Tcross^i)^2$$

$$\text{subject to } f(x) = \begin{cases} a \times x + b & \text{if } x < mpd_{CA} \\ & \text{subj. to } b = (1-a) \times mpd_{CA} \\ x & \text{otherwise} \end{cases}$$

It is important to note that we computed the best parameters $a$ and $mpd_{CA}$ which minimized the sum of squared residuals (SSR) independently for each condition, in order to compute a threshold adapted to the data from each condition.

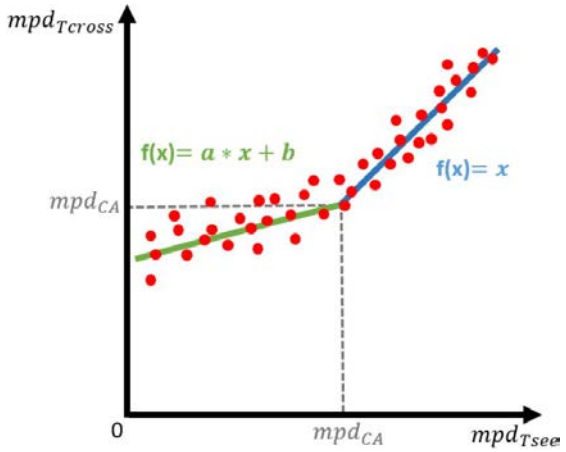



**Figure 4.4** – Schematic model used to calculate $mpd_{CA}$

### 4.2.4 Kinematics of the Collision Avoidance Task

In the real condition, we asked the confederate not to react to participants. We assessed the absence of reaction by computing his acceleration during the interaction period.

Then, for all conditions, we analyse the kinematic characteristics of the collision avoidance task by computing for each trial:

- Number of collisions: we defined a collision as occurring when the distance between the two walkers (computed from center to center) is less than $50cm$.
- Number of inversions: by linearly extrapolating the participant trajectory from current positions and velocities, we can estimate the future crossing order at time $Tcross$. We count the number of trials for each condition where an inversion of this order occurs along the trial.

For trials where there was a collision avoidance, we computed:

- Clearance Distance: the actual distance (in m) between the two walkers at $Tcross$ (i.e., $mdp(T_{cross})$).
- Speed: average participant's speed over the interaction period.
- Mpd evolution: to characterize trajectory adaptations, we computed the *mpd* evolution during the normalized interaction as well as its temporal derivative.

### 4.2.5 Gaze Behaviour

To understand gaze behaviour leading to avoidance adaptations, we analysed participants' gaze behaviour only for trials with motion adaptations. We considered two main aspects of participants' gaze behaviour namely fixations, and head and eyes movements. In the case of fixations, we also define the *gaze allocation* as the ratio of time spent looking at each object during the whole interaction.

**Fixation.** According to Parasuraman and Rizzo [2008], the gaze behaviour can be described as a succession of fixations separated by fast eye movement called saccades. An important task in eye-tracking studies is to well define these two movements [Hessels et al., 2018]. Depending on the task and situation different definitions can be found in the literature, and we therefore used the definition of fixations given by Kitazawa and Fujiyama [2010], whose experimental task shared common properties with ours: a fixation was defined as a continuous gaze on the same object for more than $80ms$. Furthermore, for each temporal window of $80ms$, all the gaze points were required to be within a range of 3.0 degrees from the initial point. Considering this definition, our dependent variables related to fixations during the interaction were:

- Number of fixations per second
- Average duration of fixations in seconds.

- Gaze Allocation: we reported, in %, where participants looked at considering 3 allocations: the confederate walker, the target and the surrounding environment.

**Gaze and head angles.** To consider gaze and head movements during the interaction period, we computed the following angles in the $XY$ plane, where both the confederate walker and the target were represented as a circle with respectively a radius of $25cm$ and $50cm$:

- $\text{Angle}_{HH}$: angle between the head vector and the heading. It was not computed for the Screen condition as participants' head was immobilized using a chin rest, and therefore always aligned with the heading direction.
- $\text{Angle}_{GH}$: angle between the gaze vector and the head vector.
- $\text{Angle}_{GW}$: angle between the gaze vector and the other walker.
- $\text{Angle}_{GT}$: angle between the gaze vector and the target.

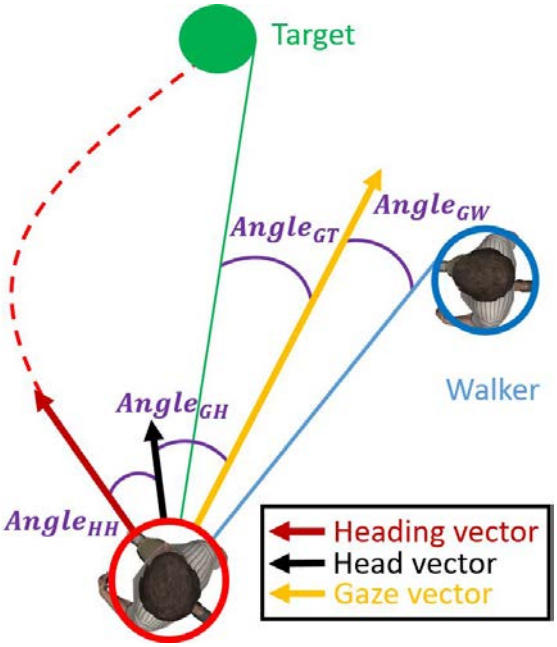



**Figure 4.5** – Illustration of the 4 angles we computed to relate head and eye movements.

### 4.2.6 Statistics

For all dependent variables, we set the level of significance to $\alpha = 0.05$. A Shapiro Wilk test was performed to evaluate whether data followed a normal distribution. If the distribution was not normal, a Friedman test was performed to evaluate the effect of the condition on these variables. Post-hoc comparisons were performed

using a Wilcoxon signed rank test with a Bonferroni correction. If the distribution was normal, a one-way analysis of variance (ANOVA) with repeated measures was performed. Greenhouse-Geisser adjustments to the degrees of freedom were applied, when appropriate, to avoid any violation of the sphericity assumption. Bonferroni post-hoc tests were used to further analyse significant effects. For all variables which described an evolution during the normalized time of interaction, we evaluated the effect of the condition using Statistical Parametric Mapping (SPM) methods [Friston et al., 2007]. This analysis allows comparing time-series data of different trials taking into account their variability at each time-step.

## 4.3 Results

### 4.3.1 Locomotion

In the real condition, we confirmed that the confederate walker had approximately a constant speed ( $1.4 \pm 0.1 m.s^{-1}$ in average).

**Thresholds for adaptation** We computed the threshold $mpd_{CA}$ separately for each condition, and trials where mpd($Tsee$) was lower than $mpd_{CA}$ of the condition. Trials were then considered to contain trajectory adjustments to avoid a potential collision. The threshold for each condition are reported in Table 4.2. Overall, we found a lower threshold for the real condition than for the virtual conditions, which is further discussed in Section 4.4.

**Table 4.2** – $mpd_{CA}$, average (±SD) speed of participants, average (±SD) number of collisions, and average (±SD) number of inversions of crossing order with respect to the experimental conditions.

| Conditions | $mpd_{CA}$ (m) | Speed ($m.s^{-1}$) | Collisions | Inversions |
|---|---|---|---|---|
| Real | 0.97 | 1.34±0.16 | 0.16±0.38 | 1.00±0.85 |
| HMD-W | 1.40 | 1.08±0.16 | 0.17±0.39 | 1.58±1.08 |
| HMD-C | 1.29 | 1.27±0.07 | 0.25±0.45 | 0.50±0.67 |
| CAVE | 1.40 | 1.25±0.05 | 0.33±0.65 | 1.16±0.83 |
| Screen | 1.46 | 1.23±0.06 | 0.58±0.90 | 1.00±0.85 |

**Average speed** of participants (Cf.,Table 4.2) was affected by the condition ($F(1.72, 18.88) = 10.12, p < 0.0001, \eta^2 = 0.48$), with lower values observed for the

HMD-W condition.

**Number of collisions and inversion** are illustrated in Table 4.2. These values were low (NB, values computed for each participant on the 26 trials of each condition). Furthermore, there is no significant effect of the condition on the number of collisions (p=0.79) or inversions (p=0.26).

**Clearance Distance**, presented in Figure 4.6, was affected by the condition ($F(2.35, 25.92) = 16.72, p < 0.0001, \eta^2 = 0.60$). Bonferroni post-hoc tests showed that it was lower in the real condition than in all the virtual conditions.

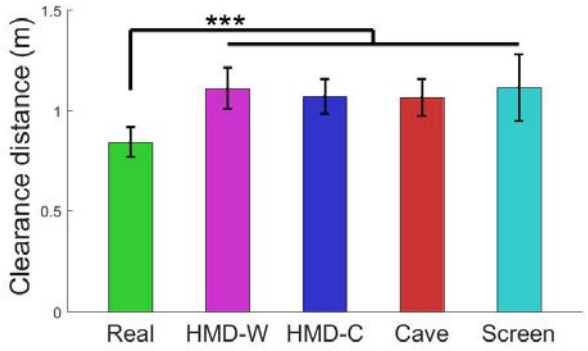



**Figure 4.6** – Average clearance distance (±SD) depending on the experimental condition. Significant post-hoc comparisons are highlighted with stars (***p<0.001)

**Mpd** evolution and its time derivative during the interaction period are illustrated in Figure 4.7. Qualitatively, we can notice that all the curves have similar shapes. From a quantitative point of view, SPM analysis showed an effect of the condition on mpd values during all the interaction ($p < 0.001$). Post hoc tests revealed that mpd in real conditions was lower than HMD-C, Screen and Cave conditions during all the interaction period, and than HMD-W from 25% to the end of the interaction ($p < 0.001$). No effect of the condition was however shown when considering mpd time derivative.

## 4.3.2 Gaze Behaviour

### 4.3.2.1 Fixations

**Number of fixations** per second during the interaction period is illustrated in Figure 4.8. A significant effect of the condition was observed ($F(4, 44) = 4.88, p =$

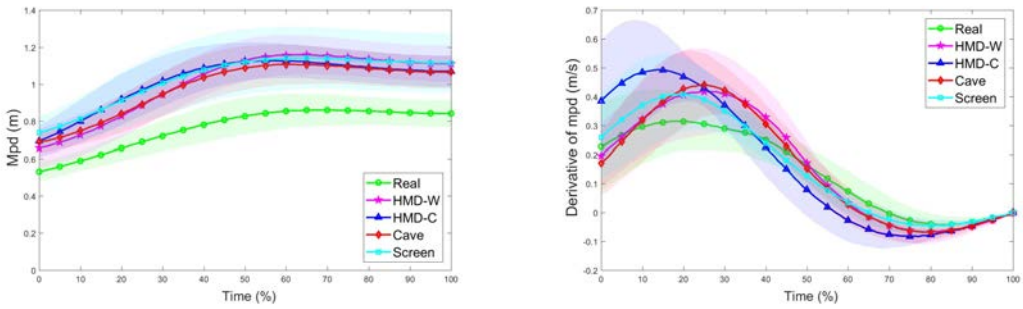



**Figure 4.7** – Kinematics adaptations to avoid a collision during the interaction are depicted through minimum predicted distance (mpd) (top) and its temporal derivative (bottom).

$0.0024, \eta^2 = 0.31$). Post-hoc analysis showed that there were less fixations in the Cave than in the conditions involving physical walking, namely HMD-W and Real.

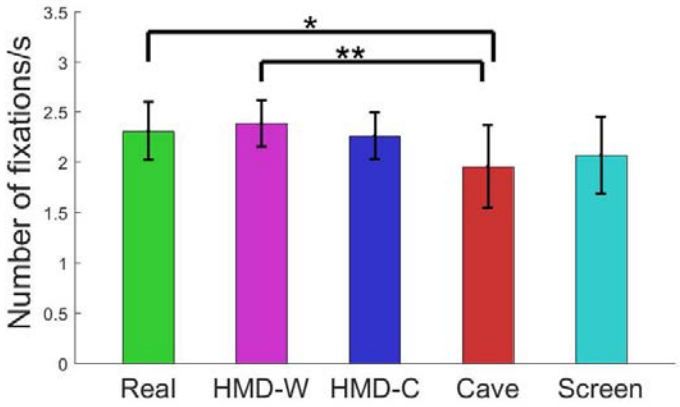



**Figure 4.8** – Number of fixations per second during the interaction for each condition

**Average duration of fixation** is illustrated in Figure 4.8 and was influenced by the condition ($\chi^2$ (4)=28.46, p<0.0001). It was longer for the Screen than for the Cave and the HMD-W, smaller for the Cave than the HMD-C and smaller for HMD-W than HMD-C.

**Gaze allocation** is summarized in Figure 4.10. A significant effect of the condition was observed on the ratio of fixation directed towards the confederate walker ($F(4, 44) = 4.25, p = 0.0053, \eta^2 = 0.28$), the target ($F(4, 44) = 3.67, p = 0.011, \eta^2 = 0.25$), as well as the environment ($F(4, 44) = 3.45, p = 0.015, \eta^2 = 0.24$). Post-hoc analysis showed that gaze allocation towards the confederate walker was lower in the Cave than in the HMD-W condition. Gaze allocation towards the target was higher in the HMD-C than in the Screen condition. Finally,

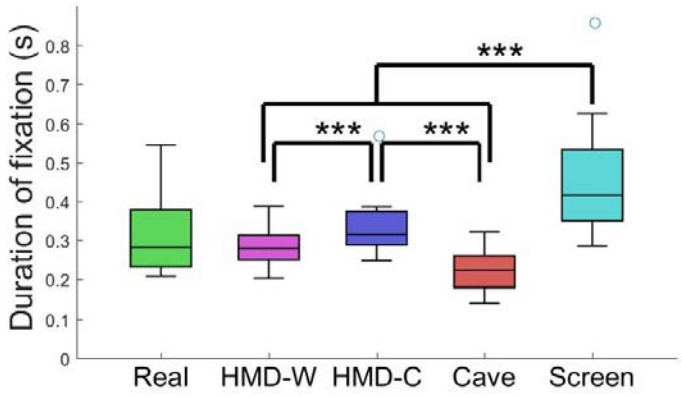



**Figure 4.9** – Average duration of fixation during the interaction for each condition.

gaze allocation towards the environment was lower in the HMD-C than in Cave condition.

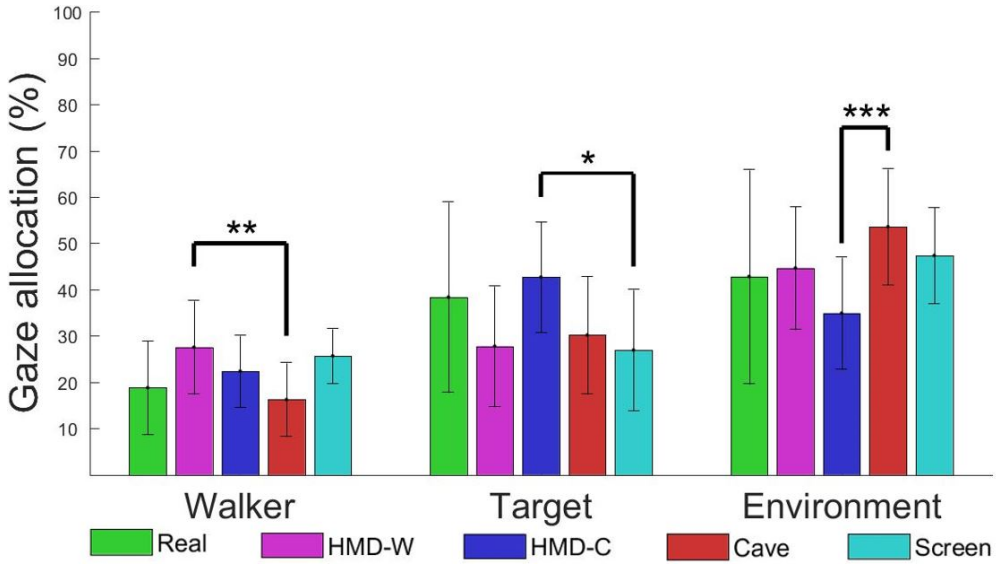



**Figure 4.10** – Average (± SD) gaze allocation towards the confederate walker (left), the target (center) and the environment (right) during the interaction period.

#### 4.3.2.2 Gaze and Head Angles

Figure 4.11 illustrates the evolution during the interaction period of all the angles related to gaze and head. SPM analysis showed an effect of the condition for all angles. Significant differences from the post-hoc SPM analysis are depicted on the top of each sub-figure by an horizontal bar on the corresponding period of the interaction.

**Angle**$_{HH}$ in the Real condition was significantly smaller than for HMD-W during the first half of the interaction, and than for the Cave from 21% to 31% of the interaction period. Also, angle$_{HH}$ was bigger in HMD-W than in HMD-C at the end of the interaction.

**Angle**$_{GH}$ in the Screen condition was significantly bigger (up to twice) than in all the other conditions at the beginning of the interaction. A significant difference was also noticed between the Real and the HMD-C conditions, angle$_{GH}$ being bigger in the HMD-C conditions from 0% to 9% of the interaction.

**Angle**$_{GW}$ in the Real condition at the beginning of the interaction was bigger than in HMD-W and Screen conditions and smaller than in the Cave conditions. It was also smaller than in the HMD-C and the Cave conditions at the end of the interaction. Also the angle$_{GW}$ in the Cave conditions was bigger than all the other three VR conditions at the beginning of the interaction.

**Angle**$_{GT}$ in the Real condition at the beginning of the interaction was smaller than all the VR conditions. It was also smaller at the beginning of the interaction in Cave compared to Screen and HMD-W.

## 4.4 Discussion

The main objective of this study was to evaluate and compare the influence of various VR setups on the gaze behaviour during a collision avoidance task. To this end, we designed an experiment were participants performed the task in real and virtual conditions and we tracked both their trajectory and gaze.

### 4.4.1 Collision Avoidance Behaviour

Kinematics analysis showed that there were some quantitative differences in the metrics of interaction between real and virtual environments. Especially, the threshold $mpd_{CA}$ that triggers motion adaptation was lower in the real condition than in virtual conditions, which is consistent with the larger clearance distance observed in the VR conditions. However, when considering the derivative of *mpd*, i.e., *mpd* variations and not absolute values, the absence of a significant difference between the studied conditions suggests that motion adjustments are performed similarly in both real and virtual environments. This corroborates results from previous studies [Olivier et al., 2017, Fink et al., 2007, Bühler and Lamontagne,

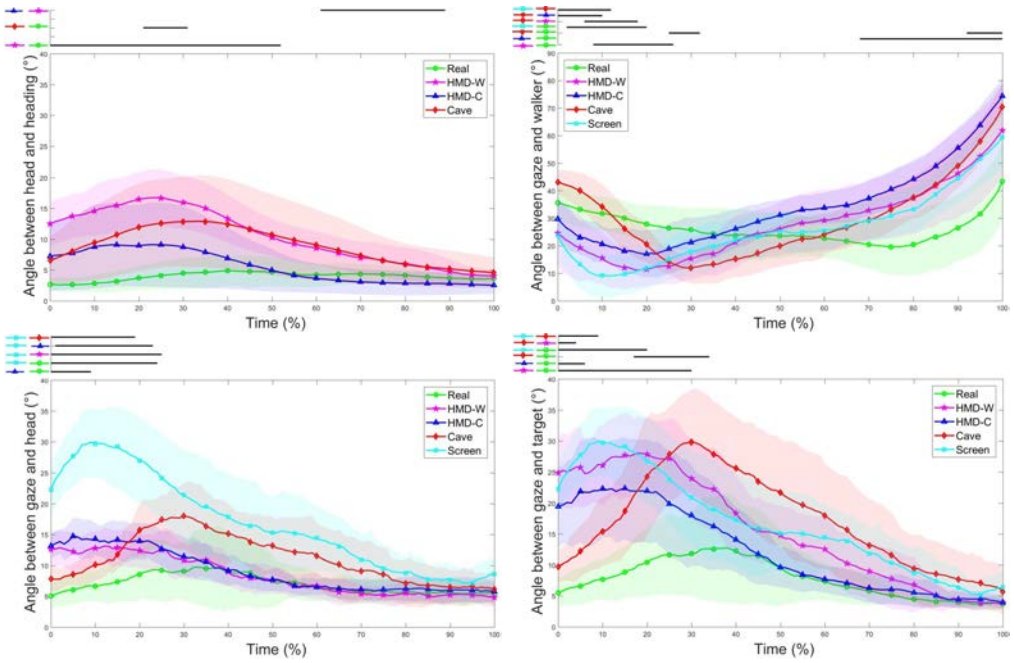



**Figure 4.11** – Evolution of the four angles $\text{Angle}_{HH}$ (top left), $\text{Angle}_{GH}$ (bottom left) , $\text{Angle}_{GW}$ (top right) and $\text{Angle}_{GT}$ (bottom right) during the normalized interactions for all the conditions. SPM analysis showed an effect of the condition on each angle. Significant pairwise post hoc results are depicted above each figure where horizontal black lines indicate during which period of time the angle differs between 2 conditions.

2018]. Furthermore, considering the VR conditions only, we did not find any significant difference between *mpd* values along the whole collision avoidance task. This suggests that collision avoidance behaviour is similar regardless of the employed VR setup. We found however a difference in the average walking speed that was lower in the HMD-W condition: participants wore a HMD but had to move among physical obstacles, which may have induced a safer locomotion speed. This result is also consistent with previous works [Bühler and Lamontagne, 2018, Agethen et al., 2018].

### 4.4.2 Fixations and Gaze Allocation

Apart from the number of fixations in the Cave condition, our results showed similarities in the number and duration of fixations between real and virtual conditions. The smaller number of fixations in the Cave condition may be linked to technical issues where we found that eye gaze recognition was slightly impaired by the lighting conditions (which is specifically discussed in Section 4.4.4).

Regarding gaze allocation (i.e., ratio of what participants looked at in the scene), results only showed a few statistical differences between the VR conditions

(Figure 4.10), where participants looked longer on average at the walker in one condition than another (HMD-W > Cave), or looked longer on average at the target (HMD-C > Screen). However we did not find any significant difference between the Real condition and any of the VR conditions.

Overall, our results validate our Hypothesis **H1**. Comparing Real and VR conditions, participants looked at a similar visual content. They spend similar proportion of time on the various elements of the scene, with similar visual patterns. However we observed a few differences between the studied VR setups. This indicates that in spite of significant difference with Real conditions, some setups tend to change gaze behaviours in Virtual Reality and could become more important when studying other type of tasks for example. We discuss more extensively those limitations below.

### 4.4.3 Gaze and Head Angles

While the evaluation of what participants looked at during the interaction is important, another interesting aspect is to evaluate whether the way they looked at this content was similar in the studied conditions.

First, we measured the evolution of the angle between the participants' gaze vector and both the walker and target, respectively by computing $\text{Angle}_{GW}$ and $\text{Angle}_{GT}$. We observed that most of the differences between the angle evolution appeared at the start of the interaction, where participants could first see the other walker (namely, the observation phase [Olivier et al., 2012]). We can observe in Figure 4.11 (top and bottom right) an overall decrease of $Angle_{GW}$ with a simultaneous increase of $Angle_{GT}$, showing that participants tended to look progressively more towards the walker and less towards the target in the first part of the interaction. As the walker is perceived with a possible risk of collision, this pattern is in correlation with previous work [Meerhoff et al., 2018a, Jovancevic-Misic and Hayhoe, 2009]. However, we observed that the amplitude of these two variations was smaller in the real condition compared to the VR conditions, suggesting that peripheral vision might have played an important role in the real condition. These two variations also seem to happen later for the Cave and Real conditions than for the other VR conditions, which again seem to demonstrate the effect of the larger peripheral vision available in these two conditions. Therefore, it is likely that the limited peripheral vision in the HMD conditions forced participants to actively look for the other walker earlier in the interaction than in the Cave and Real conditions.

Then, we evaluated how participants' gaze direction resulted from the Eye-

Head-Body relative orientations. In particular, we explored how the angle of the head relative to the heading evolved over time ($\text{Angle}_{HH}$), as well as the angle of the gaze relative to the head ($\text{Angle}_{GH}$). We found statistical differences at the beginning of the interaction between the Real condition and all the VR conditions, for either $\text{Angle}_{HH}$ or $\text{Angle}_{GH}$, which is in correlation with previous work [Pfeil et al., 2018] for a stationary situation. In the case of the HMD-W and Cave conditions, statistical differences were found only for $\text{Angle}_{HH}$, differences which were longer in time for HMD-W. Conversely, for the Screen and HMD-C conditions statistical differences were found only for $\text{Angle}_{GH}$. This seems to show that participants moved more their head in relation to their body in the HMD-W and Cave conditions, but displayed similar Eye-Head patterns. Such observations could be explained by the possibility of turning the torso at the same time as the head in order to initiate a rotation in the case of the HMD-W and the Cave conditions, which was not the case for the other two conditions. These results validate **H2**. Because of the strong differences for HMD-W, we hypothesize that such differences in behaviour might have been caused by the limited field of view of the HMD, which might have forced participants to induce larger head motions to look for the walker at the beginning of the interaction. This hypothesis should be evaluated through further studies. Similarly, such differences in the Screen condition were expected as participants' head was restricted because of the use a chin rest (required to improve the accuracy of the eye-tracker), which would have caused them to perform the scene exploration only using their eyes.

The results on $Angle_{GH}$ between the HMD-C and Real conditions were however more surprising, especially as we expected the limited field of view to lead to similar patterns than for HMD-W. This difference on Eye-Head-Body relative orientations between HMD-W and HMD-C when compared to the Real condition could therefore also be due to the difference in the locomotion interface, namely walking or seating using a game controller. This suggests that not only the type of VR device, but also the locomotion interface, might have an impact on the Eye-Head-Body coordination, which partially invalidates **H3** (at least in terms of differences in eye movements). Therefore, further studies should be conducted to thoroughly evaluate these effects.

### 4.4.4 Limitations

Even with high-end eye-tracking devices, accurately tracking eye gaze can be influenced by the experimental conditions, such as lighting or participant's eye color. For the experiment reported in this paper, data from four participants

were discarded because of eye-tracking issues regarding the quality of the calibration (3 participants with the Tobbi glasses, 1 participant with the FOVE HMD). In particular, we found the Tobbi glasses to be very sensitive to the room illumination while tracking the gaze, both in the Real and Cave situations. While this impacted calibration for three discarded participants, we also found that it led to more noise in the data in the Cave condition across participants, including the inability to identify eye gaze in some frames and influencing the identification of fixations in the data. We believe that such a problem might explain at least partially the lower average of fixations per second for the Cave condition (Figure 4.8), where illumination is typically reduced for 3D projection purposes. Regarding gaze tracking issues for the one participant in the HMD-W condition, we hypothesize that it could come from the head movements, a combination of head movements and a HMD not tightly fixated, or bad calibration because of eye color. These technical limitations therefore influenced our sample size, which should be increased in future experiments to explore more subtle effects and to strengthen our conclusions.

While the presented experiment is to our knowledge the first attempt to study in VR the kinematics of a collision avoidance task in conjunction to gaze behaviour, it also involves a specific controlled interaction in restricted physical and virtual spaces. For instance, interactions were possible in a $9m$X$9m$ area, which can also influence participant behaviour. Therefore, further studies would benefit from evaluating kinematics and gaze behaviours in larger physical setups, as well as more complex situations.

## 4.5 Conclusion

In this chapter, we explored the use of VR to study locomotion and gaze behaviour during collision avoidance between walkers, where we studied a collision avoidance task between two walkers in both real and virtual environments. We then compared the real condition with four VR conditions (including a CAVE, a screen and a Head-Mounted Display). Our results show that fixations and gaze allocations were similar in both real and virtual environments. However the exploration of the environment, to seek visual information, was different in both real and virtual environments. Regarding motion adaptations, our outcome confirms previous work, where collision avoidance behaviour is qualitatively similar in VR and real conditions. This indicates, not only the type of VR device but also the locomotion interface has an impact on eye movement, which should therefore also be taken into consideration. In conclusion, our results suggest that VR has potential

to study qualitatively the gaze behaviour for this kind of situation (e.g, collision avoidance between two pedestrians). Therefore, the aim of our next chapter is to extend these analyses to a more complex situation, which is navigation in a crowd. We then examine also different factors, such as the density of the crowd, which can influence the activity of the gaze.

Chapter 5

# Navigation in Crowds: Influence of Virtual Reality and Crowd Density

## Contents


**5.1 Introduction . . . . . 78**
**5.2 Overview . . . . . 79**
5.2.1 Apparatus & Task . . . . . 79
5.2.2 Participants . . . . . 80
5.2.3 Analysis . . . . . 81
**5.3 Real vs. Virtual Validation . . . . . 84**
5.3.1 Procedure . . . . . 85
5.3.2 Analysis & Results . . . . . 86
5.3.3 Discussion . . . . . 89
**5.4 Effect of Crowd Density on Eye-gaze Behaviour . . . 90**
5.4.1 Procedure . . . . . 90
5.4.2 Analysis & Results . . . . . 91
5.4.3 Discussion . . . . . 93
**5.5 General Discussion . . . . . 94**
5.5.1 Crowd Simulation . . . . . 94
5.5.2 Limitations . . . . . 94
**5.6 Conclusion . . . . . 96**

## 5.1 Introduction

In the previous chapter we showed that gaze activity during a collision avoidance between two pedestrians in a real or virtual environment was similar. However, it is unlikely that we only interact with a single person when walking. This is why in this chapter we are interested in the analysis of gaze activity when navigating in a crowd, which is a daily life situation. As before, it is first of all necessary to evaluate the impact of virtual reality on gaze activity for this complex situation. In a first experiment, we asked therefore participants to walk an existing street (Figure 5.1-left), and recorded both their eye-gaze activity and their visual environment. We then reproduced the same situation in VR using a digital replica of the street (Figure 5.1-middle-left and middle -right). We compared the eye-gaze activity of participants under these two conditions.

Then, a second objective of this chapter is to further understand how eye-gaze activity is influenced by the number of people we are interacting with in a crowd, as a mean of better understanding interaction neighbourhood as well as collision avoidance manoeuvres. More precisely, as gaze is indicative of how people take into account other individuals to navigate in crowds [Meerhoff et al., 2018a], we wonder how eye-gaze activity features will be influenced by the density of the crowd. Therefore in a second experiment, we evaluated the influence of crowd density on eye-gaze activity. By asking participants to navigate a virtual street populated with different densities of virtual characters (Figure 5.1-right).



**Figure 5.1** – Our objective is to analyse eye-gaze activity within a crowd to better understand walkers' interaction neighbourhood and simulate crowd behaviour. We designed two experiments where participants physically walked both in a real and virtual street populated with other walkers, while we measured their eye-gaze activity (red circle). We evaluated the effect of virtual reality on eye-gaze activity by comparing real and virtual conditions (top row) and investigated the effect of crowd density (bottom row).

Our contribution in this chapter is therefore twofold. We are contributing to the validation of VR as a tool for studies coupling eye-gaze activity and navigation, showing important similarities of virtual compared to real behaviours. We also propose new ways to improve crowd simulation algorithms by improving knowledge about how the interaction neighbourhood of walkers might be visually evaluated by viewers.

The remaining of the chapter is organized as follows. Section 5.2 presents an overview of the experimental design and analyses common to both experiments. Section 5.3 describes the first experiment about the comparison of gaze activity in a real and virtual crowded environment and show the results found. Section 5.4 presents the second experiment about the influence of crowd density on gaze activity and the results observed. Section 5.5 presents a general discussion about these two experiments. Finally, section 5.6 provides a general conclusion and future perspectives.

## 5.2 Overview

Our objective is to explore and further understand the interaction neighbourhood of people walking in busy environments, with the particular interest of relying on the analysis of the walker's eye-gaze activity. We chose to perform this study in VR, to facilitate the control of experimental conditions, the replication over several participants, as well as the measure of the eye-gaze activity. To this end, we conducted two experiments, the first one enabled us to study the bias induced by VR on eye-gaze activity when navigating in a crowd (Section 5.3). The second experiment focused on the impact of crowd density on eye-gaze activity (Section 5.4). We decided to carry out these experiments based on the task of walking in a busy street. The advantage of using such a task is to correspond to a daily-life situation, with no ambiguity on how to perform it: participants simply have to walk and to follow the direction of the street as they commonly do. Having a clear and simple task is important to us, as we know that the nature of the task has a direct impact on eye-gaze activity [Yarbus, 2013].

### 5.2.1 Apparatus & Task

Participants walked the real, or digital reproduction, of Vasselot street, in the city of Rennes, France (see Figure 5.1). The digital reproduction was designed by Archivideo, with professional centimetric geometrical precision and textures generated from real photos. Slight differences between the real and virtual environment

were however still present, due to minor differences in the exact localization or aspect of some objects, such as chairs at the terraces of cafés, billboards, etc. In both RE and VE, we were interested in recording participants' eye-gaze activity while they interacted with other pedestrians in the street:

- Real Environment (RE): participants wore the Tobii pro glasses 2 eye-tracking, which recorded both their eye-gaze activity (50Hz recording, 4 eye cameras) and a video of their visual field (scene camera: 25Hz, 90° field of view, H.264 1920x1080 pixels)
- Virtual Environment (VE): participants were immersed in the VE using a FOVE HMD (70Hz, 100° FoV), which comes with an integrated eye-tracker (100Hz). The virtual scene was rendered using Unity. Participants freely moved in a physical space (gymnasium) of $20m \times 6m$, while their position was tracked with a 23-camera motion capture system (Qualisys).

In both RE or VE, participants were asked to navigate in the street while avoiding collision with pedestrians and to stop when they were in front of a specific shop. They performed multiple round trips between two specifics shops (separated by 20m). Figure 5.1-middle-left gives an example of the virtual conditions, where participants were asked to navigate in the virtual street by walking in the gymnasium.

For the virtual condition, the virtual humans were driven by RVO [Van Den Berg et al., 2008], an open-source crowd simulator often used in video-games [Snape et al., 2012]. Its computational performances enable to have multiple agents avoiding collisions with other obstacles without impacting the framerate, which is crucial for VR experiments. In our experiments, RVO parameters were the following: each agent was represented by a $0.5m$-radius cylinder, took into account a maximum of 7 neighbours in a $5m$ space around them, was assigned a random speed $\in [0.95, 1.25] m/s$, and was set up to perform collision-avoidance manoeuvres $3s$ before a potential collision. We chose a distribution centered around $1.1m/s$ instead of $1.3m/s$ for the agent's comfort speed as participant are walking slower in VR [Bühler and Lamontagne, 2018, Agethen et al., 2018].

### 5.2.2 Participants

Twenty-one unpaid participants, recruited via internal mailing lists among students and staff, volunteered for the experiment. They were all naive to the purpose of the experiment, had normal or corrected-to-normal vision, and gave written and informed consent. The study conformed to the declaration of Helsinki,

and was approved by the local ethical committee. Data from one participant were removed from the first experiment a posteriori because the tracking ratio was lower than 80% in the RE. Similarly, because of incorrect calibration of the eye-tracking device in the virtual conditions, data from 4 other participants were removed a posteriori from the first experiment, and of 3 participants from the second experiment. Therefore, only the data from sixteen participants (4F, 12M; age: avg.=24.9 ± 3.2, min=20, max=30) was used for the first experiment and the data from eighteen participants (4F, 14M; age: avg.=25.5 ± 4.0, min=20, max=36) was used for the second experiment.

### 5.2.3 Analysis

#### 5.2.3.1 Eye Data Collection

For the RE, a camera placed at the center of the eye-tracking glasses, just above the nose (see Figure 5.2-a), recorded what participants saw over time. Gaze coordinates correspond to pixel positions in this video. For the VE we recreated the same protocol by placing a camera with the same characteristics between the participant's two eyes in the VE (see Figure 5.2-b). As for the RE, it was impossible to record the participants trajectory, in this chapter we used the 2D location of the gaze in the recorded video of the environment seen by the participant to study the eye-gaze activity. However, it is important to distinguish the difference between gaze activity and eye movements. Eye movement refers to the local coordinates of the gaze relative to the head. Gaze activity, on the other hand, corresponds to the global coordinates of the gaze in the world space [Guitton and Volle, 1987], which therefore also accounts for head rotations. In our case, we recorded eye movements, and assume that the head movements contribution is not significant. Qualitatively speaking we observe that participants do not significantly move the head when performing the task at hand, i.e. reaching destination 20m further down a 5m-wide street. The movement of the eyes recorded will therefore be close to the gaze activity, and this is the reason why we talk about eye-gaze activity in this work. Furthermore in case of sudden movements, it has been shown that the eyes initiate the movement, then the head, and finally the body, both in RE or VE [Patla, 1997, Reed-Jones et al., 2009], thus resulting in a saccade. In this chapter we are only interested by the location of fixations, and therefore will not analyse eye movements during saccades.

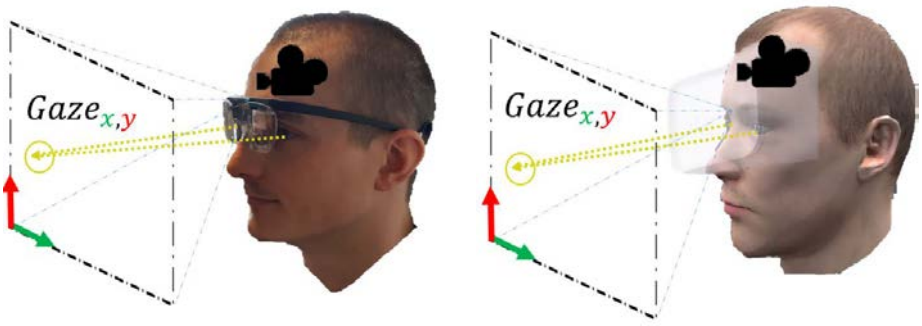



**Figure 5.2** – Setup used to collect gaze location for both real (a) and virtual (b) conditions. For each condition, 2D gaze location is displayed in the image recorded by the real or virtual camera (in black).

#### 5.2.3.2 Fixations Computation

Our eye-gaze activity analysis relies, as in many other studies on gaze, on the measure of visual fixations. As for other studies, our first important task is therefore to accurately register the fixations and the saccades [Hessels et al., 2018]. Depending on the task and situation several methods have been proposed in the literature to compute fixations [Kar and Corcoran, 2017], each with advantages and limitations depending on the situation. In our situation, gaze fixations are computed based on the 2D gaze location of participants in the recorded images (real or virtual) over time, which is different than for the precedent chapter with 3D gaze location were used for gaze data.

The method used to compute fixations is inspired by the dispersion-based algorithms (I-DT) described by Salvucci *et al.* [2000]. We first compute the maximum distance between 2D gaze locations and their centroid over a sliding window of 100 ms. After identifying all the points where this distance is less than a threshold of 1.5°, we process each identified point by accumulating in the same fixation the neighbouring points if they respect these two conditions:

- $mean(GazeD) < 1.5°$,
- $std(GazeD) - StdInit < 0.4°$,

Where $GazeD$ is the list of distances between the fixation's centroid and all gaze points currently belonging to that fixation, and $StdInit$ the standard deviation of $GazeD$ at the start of the fixation. In our analyses, we chose a circle with a radius of 1.5°, as the diameter of the fovea is about 3°, while the threshold for

variance was chosen empirically and set to 0.4°. Additional information about the pseudo-code for computing fixations is provided as supplemental material.

#### 5.2.3.3 Independent Variables

As our goal is to understand eye-gaze activity while walking in crowds of different densities, as well as the biases that can be introduced by VR, we considered two main aspects of participants' gaze fixations. The first aspect relates to the characteristics of fixations, namely the average duration and the amplitude of saccades. The second aspect relates to where participants looked in the scene, through the coverage of the fixations:

**Fixation descriptors**

- *Average duration of fixations (ms).* This feature informs about time spent on each object to extract visual information.
- *Amplitude of saccades (degree).* This feature informs about the distance separating successive targets. It is computed as the distance between two successive fixations.

**Eye-gaze spatial distribution** These features describe eye-gaze activity over the whole navigation task. For each condition and each participant, we computed a fixation map according to the definition given by Wooding *et al.* [2002]. To create a fixation map, we start with a blank image and then, for each fixation, we added a gaussian at the center of the fixation. We use a standard deviation $\sigma$ of 1.5° for the gaussian, which will approximate the fovea. The fixation map will be represented as a heat-map, and for the sake of visibility, we will display the logarithm values of this map. From this map, we calculated two metrics:

- $Peak_{Fixation}$ is the maximum value of the fixation map normalized over the number of fixations. It describes where participant's focused more their gaze during the task.
- *Coverage* is the number of pixels in the fixation map superior to a threshold $D_{crit}$ over the size of the image, as defined by Wooding *et al.* [2002]. We choose $D_{crit}$ so has not to consider isolated fixations, i.e., fixations at a larger distance than $2\sigma$ (3°) from any another fixation. As a result, $D_{crit}$ is computed as follows:

$$D_{Crit} = \frac{1}{2\pi\sigma^2} + \frac{1}{2\pi\sigma^2}e^{-2} \tag{5.1}$$

where the first term of Equation 1 is the peak value of the gaussian representing each fixation in the fixation map, and the second term the value at $2\sigma$.

#### 5.2.3.4 Statistical Analysis

We set the level of significance to $\alpha = 0.05$. A Shapiro Wilk test was performed to evaluate whether the distribution of our data followed a normal distribution. When comparing Real vs. Virtual conditions (Experiment 1), we conducted paired t-tests. In Experiment 2, we investigated the effect of density in Virtual conditions by conducting either a Friedman test with Wilcoxon-signed rank post tests when the distribution was not normal, and a one-way repeated measures analysis of variance (ANOVA) with post-hoc paired t-tests otherwise. Greenhouse-Geisser adjustments to the degrees of freedom were applied, when appropriate, to avoid any violation of the sphericity assumption. For the post-hoc tests, we adjusted the p value to account for multiple comparisons using the Benjamini-Hochberg procedure [1995] with a false discovery rate of 0.1.

## 5.3 Real vs. Virtual Validation

The goal of this first experiment is to evaluate whether the eye-gaze activity of a human walking a street is biased in VR conditions compared to real ones. While our previous chapter showed qualitative similarities when considering an interaction between two walkers, it is not yet established how these results generalize to more complex scenarios involving larger numbers of pedestrians. In particular, our experiment aims at assessing the following hypotheses:

**H1.1** The scene is displayed through a HMD in the VE. This reduces the field of view of participants in comparison with RE. As a consequence, we expect gaze spatial distribution to be different in the VE. Consistently, we expect the amplitude of eye saccades, as well as the area covered by the gaze, to be smaller in the VE.

**H1.2** The features of the RE are accurately reproduced in VE (same street, buildings, geometry and same density of people) and the participants' task remains identical. We therefore expect the duration of gaze fixations to be similar in both conditions, as participants should take similar visual information to perform the task in RE and VE.

**H1.3** The task is to walk toward the opposite side of the street, which is a central point in the participants field of vision. We thus expect to observe gaze fixations to be centered in the field of view.

### 5.3.1 Procedure

Participants were asked to physically walk through the real, and then, the virtual street (see Section 5.2). They performed the RE first because the parameters of the virtual condition were adjusted to be as similar as possible to the RE, in terms of visual density of pedestrians encountered. As all participants could not perform the RE under the exact same experimental conditions, we chose to minimize differences in terms of brightness and crowd by conducting the real condition during lunchtime over several days. The virtual counterpart was then conducted approximately one week later.

To enable comparisons between RE and VE, we first estimated for each trial the number of people they saw in the RE. We then created a specific stimuli that reproduced this same number. To this end, we detected and tracked people visible in the video recorded through the Tobii glasses (see Figure 5.3a-b) using a combination of two neural networks: Yolo [Redmon and Farhadi, 2018] and DeepSort [Wojke et al., 2017]. Based on tracking information, we categorized people into 3 categories: standing, walking in the same direction, or walking in the opposite direction to the participant. We generated scenarios with similar features by spawning virtual characters in the street. This process is illustrated in Figure 5.3, while the similarities between the generated RE and VE are analysed and discussed below.

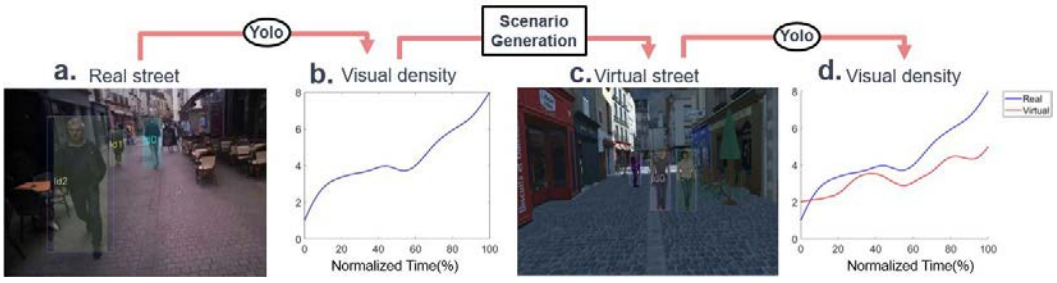



**Figure 5.3** – Generation of the virtual scenarios. a) Video recording when a participant walked in the real street. b) Density of people seen by participants over the normalized duration of the trial, estimated by tracking people visible in the video recording using deep learning algorithm. c) Virtual scenario, reproducing qualitatively similar situations in the virtual conditions in terms of virtual characters encountered, as seen by the participant. b) Density of individuals actually seen by the participant over the normalized duration in both the real (blue) and the virtual trial (red).

Finally, participants wore Tobii eye-tracking glasses in the RE, whilst the Fove HMD was used in the VE. They performed 4 trials in each condition, as well as 2 initial training trials to get accommodated to VR. The experiment lasted approximately 10min for the RE, and approximately 15min for the VE.

## 5.3.2 Analysis & Results

### 5.3.2.1 Comparison between Real and Virtual Stimuli

As mentioned above, the VE were generated so as to reproduce similar distributions of people compared to each corresponding real trial. These generated scenarios were also verified by the experimenter prior to the VE. To this end, we ran the same tracking techniques on virtual stimuli to estimate the number of characters seen by participants. Figure 5.4 presents the average number of individuals seen by each participant across trials, for both RE and VE. While inter-participant differences exist, and are expected as the RE could not be controlled in terms of pedestrian activity in the street, results show that the number of individuals seen is quite similar in both RE and VE, suggesting that the real and virtual stimuli presented were mostly similar in this aspect.

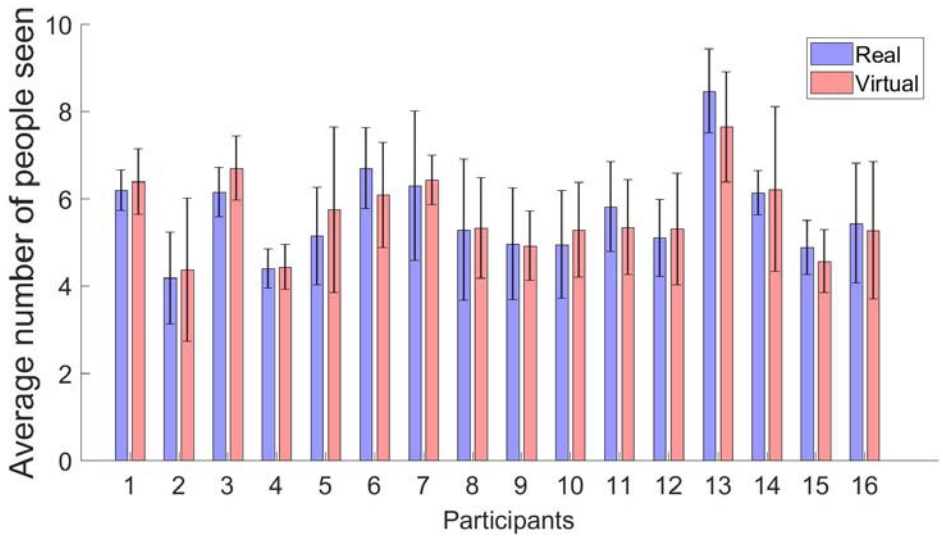
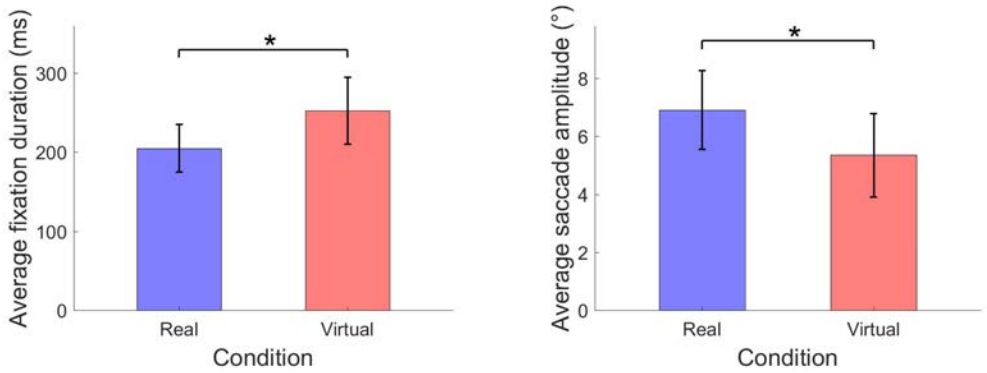



**Figure 5.4** – Average number of people seen by participants across trials for both RE and VE. The black bars represent the standard deviation for each participant and conditions.

### 5.3.2.2 Fixations and Saccades

The average duration of fixations is illustrated in Figure 5.5-a) and is influenced by the condition ($t(15)=3.9$, $p<0.005$, $r=0.71$), where the duration of fixations

was significantly longer in VE ($252.5 \pm 42.4ms$) than in RE ($204.8 \pm 30.4ms$). The amplitude of saccades is illustrated in Figure 5.5-b) and is also influenced by the condition (t(15)=4.1, p<0.001, r=0.72), where the amplitude of saccades is significantly larger in RE ($6.9 \pm 1.3°$) than in VE ($5.3 \pm 1.4°$).

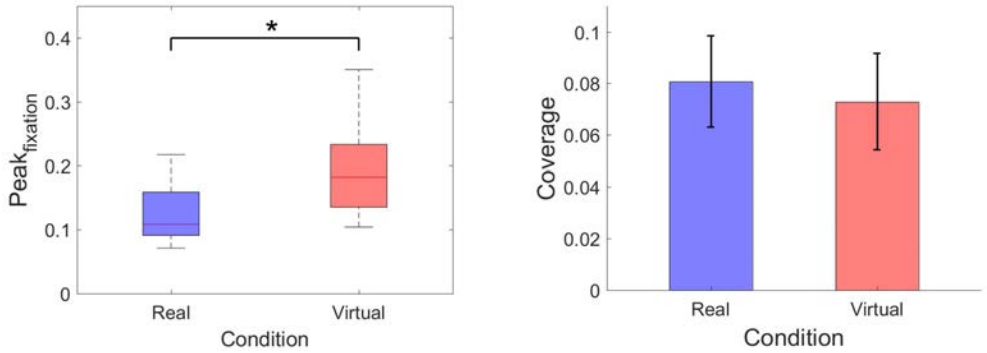



a) b)

**Figure 5.5** – a) Average duration of fixations and b) average amplitude of saccades for RE and VE.

#### 5.3.2.3 Gaze Spatial Distribution

The average $Peak_{Fixation}$ is illustrated in Figure 5.6-a), and is influenced by the condition ($T = 0$, $Z = 3.51$, $p < 0.001$, $r = 0.88$), where results show that the peak value is significantly higher for the VE (0.18) than for the RE (0.11). Coverage is illustrated in Figure 5.6-b), but we did not find significant differences between RE and VE ($p = 0.14$).




a) b)

**Figure 5.6** – a) $Peak_{Fixation}$ and b) Coverage for RE and VE.

A visual representation of these results is provided in Figure 5.7, which displays the fixation maps for both the RE and VE. In particular, it is noticeable that gaze

locations seem to follow a centered distribution for both conditions. While the center of this distribution is approximately at the center of the $1920 \times 1080$ image on the x-axis for both conditions (RE: $Peak_x = 965 \pm 56pixel$, VE: $Peak_x = 962 \pm 68pixel$), it appears to be higher on the y-axis for the RE than for the real one (RE: $Peak_y = 717 \pm 102pixel$, VE: $Peak_y = 526 \pm 65pixel$).

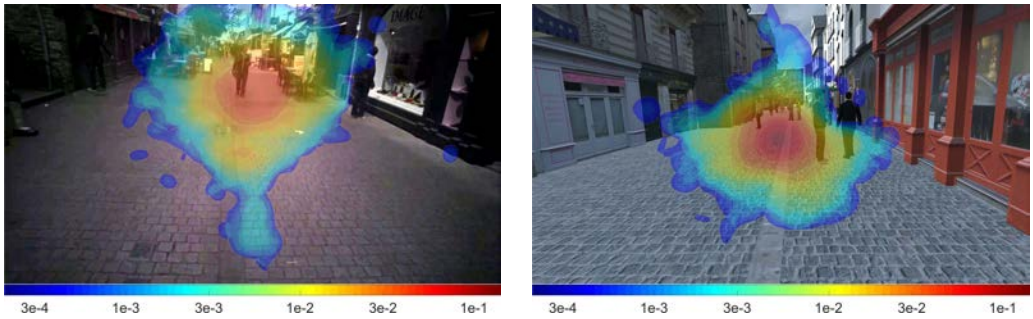



**Figure 5.7** – Heat-map (log-transformed) of the gaze fixation distribution for both RE and VE

.

To identify whether this difference could be due to a shift of the horizontal reference axis between the RE and VE (i.e., an angle between the Tobii camera axis and HMD virtual camera axis), we asked two volunteers to identify (by clicking with the mouse button) the horizon line at the end of the street in a selection of 100 images randomly extracted from the video recordings of real trials, as well as in 100 images from virtual trials. The distribution of the altitude of the horizon line in images follows a normal distribution with a average coordinate of $804.5 \pm 112.8$ pixel for the real images and an average coordinate of $597.6 \pm 86.6$ pixel for the virtual ones, showing therefore a difference of 206.9 pixels between these two centres (equivalent to an angular difference of 9.7°). This difference is similar to the y-axis peak, and might therefore explain the previous significant difference.

#### 5.3.2.4 Gaussian Model for Gaze Prediction

The spatial distribution of the fixations follows a centered distribution that decreases exponentially around the center, as shown in the log scale heat-maps in Figure 5.7. We fit a Gaussian model on this distribution, which estimated parameters ($\mu x$, $\mu y$, $\sigma_x$ and $\sigma_y$) are presented in Table 5.1 with the corresponding coefficient of correlation ($R^2$, with $p<0.05$). It is also interesting to mention that the vertical difference in the location of the center of gaze fixations as estimated

by our Gaussian model fitting is of 7.75°, which is relatively close to the angular difference observed in the altitude of horizon estimated from the real and virtual images (9.7°).

| Conditions | $\mu_x$ | $\mu_y$ | $\sigma_x^\circ$ | $\sigma_y^\circ$ | $R^2$ |
|---|---|---|---|---|---|
| RE | 0.15° | 8.10° | 7.69° | 6.93° | 0.77 |
| VE | −0.57° | 0.35° | 6.11° | 5.08° | 0.77 |

**Table 5.1** – $\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$ of the gaussian distribution for the gaze location in the image and $R^2$ between this distribution and the initial distribution with respect to the experimental conditions.

### 5.3.3 Discussion

In this first experiment we showed that there are several differences in the walker's eye-gaze activity between the RE and VE. First, saccades are of a greater amplitude in the RE than in the VE. In addition, the value of $Peak_{Fixation}$ is different, indicating that the gaze is more intensely focused on the center of the visual field in VR. Moreover, the fitting of the Gaussian model on the fixations location distribution resulted in a correlation coefficient of 0.77 for both conditions, that we estimate to be sufficiently high to validate this choice. A larger value is computed for $\sigma$ in the RE. All these differences are consistent with hypothesis **H1.1**. Nevertheless, no statistical difference is observed for coverage, which does not allow us to fully confirm the hypothesis. Note that our hypothesis was founded on the restriction of the field of view due to the use of HMDs. However, the performed task, i.e., walking straight in a street, may not require the exploration of peripheral areas. If true, users may not have to compensate such a reduction.

A statistical difference is however present for the average duration of fixations, thus invalidating our second hypothesis **H1.2**. We interpret that the cause of this difference is possibly due to the level of detail in the digital replica of the street. In spite of the high quality of the digital scene, there are always missing details, e.g., birds, or sounds, that could have attracted the visual attention of our participants in RE, and provoked faster fixations.

We also find a significant difference in the altitude of the gaze center between RE and VE. This difference appears to be well explained by a difference in the orientation of the eye-trackers reference horizontal axis between the two conditions. Finally, participants display a similar eye-gaze behaviour as they have a centered distribution for gaze location toward their goal, validating thus Hypothesis **H1.3**.

Therefore, our experiment highlights the importance of the digital content when performing gaze studies in VR, as eye-gaze activity seems to be affected by the richness and realism of this content. Despite our experiment being based on high-fidelity urban digital mockups, we cannot claim that we were able to reproduce all the details of reality. Nevertheless, these results provide valuable insights to understand which aspects of eye-gaze activity are qualitatively similar between RE and VE, to further explore using VR aspects about navigating in a crowd which are not possible in real situations. We thus believe that VR is a valid tool to study such activity, while experimenters should remain aware that some elements in real environments may distract walkers' gaze more.

## 5.4 Effect of Crowd Density on Eye-gaze Behaviour

The goal of this experiment is to evaluate the impact of crowd density on eye-gaze activity, by leveraging the use of VR to accurately control the number of displayed characters. When density increases, walkers have more frequent interactions and face more people. In this experiment we aim at assessing the following hypotheses:

> **H2.1** As the amount of visual information increases with density, we expect that the duration of fixations will decrease accordingly. By increasing the scanning frequency, participants will be able to consider more elements in the VE.
>
> **H2.2** In addition, we also expect participants' gaze to be more focused towards the center of the visual field, since pedestrians with the highest risk of collision are typically the ones in front of them. In particular, we expect that the amplitude of the saccades, as well as the area covered by the gaze, will decrease as the density of the crowd increases. This also implies that the gaze will be more intense in the center of the field of vision.

### 5.4.1 Procedure

In this experiment, participants were immersed in the same VE as described in Section 5.2, and were instructed to navigate in the virtual street through a crowd walking in a unidirectional flow in the opposite direction. Our objective for this experiment is to study the impact of crowd density on eye-gaze activity. In particular, for a specific crowd density $d$, we generated a scenario with $d$ virtual

characters every 15 meters at the start of the trial. These characters were driven by the RVO crowd simulator, similarly to the previous experiment (see details in Section 5.2.1). Participants were asked to navigate through the crowd for approximately 20m, so different speeds of virtual humans did not affect the visual densities for such a short period. Participants were presented in random order with 6 different conditions of densities: $d \in [2, 5, 10, 14, 18, 24]$ (Figure 5.10). For each density, participants had to perform 4 repetitions. Participants performed in total 24 trials, and the experiment took on average 20 minutes per participant.

## 5.4.2 Analysis & Results

### 5.4.2.1 Fixation Descriptors

The analysis on average duration of fixations shows an effect of density ($\chi^2(5) = 14.15873$, $p < 0.01463$). However, this is not confirmed by post-hoc pairwise comparisons. Nevertheless, for illustrative purposes, this result is displayed in Figure 5.8-a. Density has however a strong effect on the amplitude of saccades ($F(2.69, 45.8) = 18.4$, $p < 0.000001$, $eta_p^2 = 0.51$), where post-hoc analysis shows that the amplitude of saccades is significantly larger when navigating in a crowd with a density of 2 and 5 than for any other condition (Figure 5.8-b).

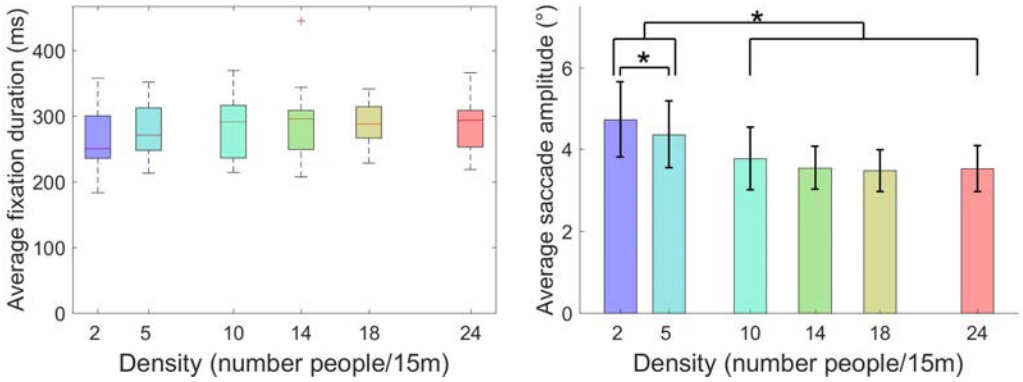



**Figure 5.8** – a) Average duration of fixations and b) average participant's amplitude of saccades depending on crowd density.

### 5.4.2.2 Gaze Spatial Distribution

The average $Peak_{Fixation}$ is illustrated in Figure 5.9-a). An ANOVA shows an effect of the density ($F(5, 85) = 2.74$, $p < 0.05 (= 0.024)$, $eta_p^2 = 0.14$), which is not confirmed by post-hoc pairwise comparisons. We however find an effect of

density on Coverage ($\chi^2(5) = 15,78$, $p = 0,008$), where post-hoc analysis shows that the coverage is larger when navigating in a crowd with a density of 2 than for any other condition (Figure 5.9-b).

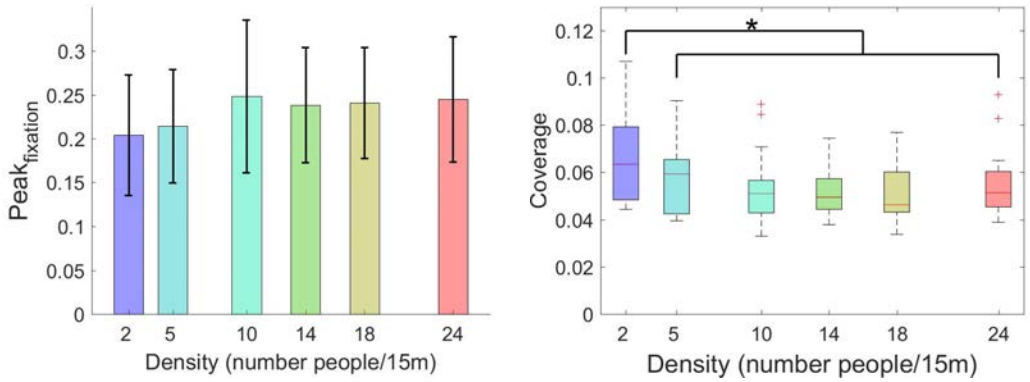



**Figure 5.9** – a) $Peak_{Fixation}$ and b) Coverage depending on crowd density.

Fixation maps are also displayed for each condition in Figure 5.10. For each density, the gaze location follows a centered distribution, furthermore it seems that the coverage by the gaze is decreasing as the density of the crowd is increasing, especially on the vertical-axis.

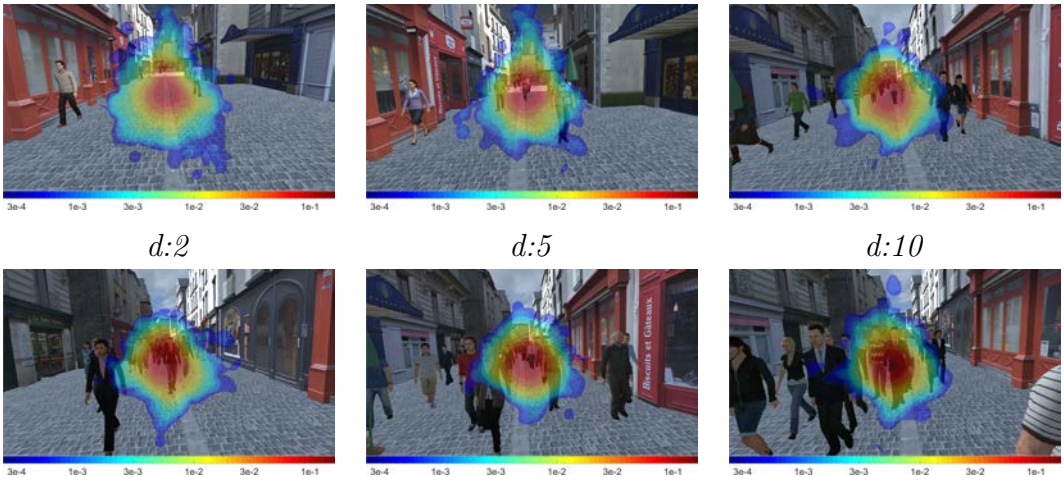



**Figure 5.10** – Virtual street with all the different crowd densities. For each density, the fixation area (log-transformed) is displayed on top of the image

### 5.4.2.3 Gaussian Model for Gaze Prediction

As in the precedent experiment, the spatial distribution of the fixations follows a centered distribution that decreases around the center, as shown in the log

scale heat-maps in Figure 5.10. We fitted a Gaussian model on the gaze location distribution. The estimated Gaussian parameters ($\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$) are reported in Table 5.2, with the corresponding correlation coefficient ($R^2$, with p<0.05). Qualitatively, the $\sigma$ values decreased with the increase of density, especially on the vertical-axis. dispersion.

**Table 5.2** – $\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$ of the gaussian distribution for the gaze location in the image and $R^2$ between this distribution and the initial distribution with respect to the density conditions.

| Conditions | $\mu_x$ | $\mu_y$ | $\sigma_x^\circ$ | $\sigma_y^\circ$ | $R^2$ |
|---|---|---|---|---|---|
| $d:2$ | $-0.72°$ | $0.02°$ | $5.06°$ | $5.82°$ | 0.75 |
| $d:5$ | $-1.08°$ | $-0.11°$ | $4.38°$ | $5.39°$ | 0.79 |
| $d:10$ | $-1.22°$ | $-0.55°$ | $4.21°$ | $4.64°$ | 0.83 |
| $d:14$ | $-0.69°$ | $-1.01°$ | $4.26°$ | $4.53°$ | 0.84 |
| $d:18$ | $-1.05°$ | $-0.91°$ | $4.07°$ | $4.01°$ | 0.87 |
| $d:24$ | $-0.87°$ | $-1.36°$ | $4.26°$ | $4.45°$ | 0.83 |

### 5.4.3 Discussion

In this second experiment we show some differences in the walker's eye-gaze activity when crowd density increases. In particular, we show statistical differences in the amplitude of the saccades and the area covered by the gaze. These two variables increase when the crowd density decreases. Furthermore, this result is also supported by the Gaussian model fittings. The $\sigma$ parameter, which is correlated with the coverage, decreases with the increasing crowd density, especially on vertical-axis. However, there is no statistical difference for $Peak_{Fixation}$, which indicates that walker's gaze is not more intense in the center of the field of vision as expected. From all these results, we can only partially validate our second hypothesis **H2.2**. Nevertheless, it seems that a Gaussian model could be used to estimate the distribution of gaze location as the correlation coefficient is high for each condition of density.

Concerning the duration of the fixations, post-hoc analysis did not allow to validate that density has an impact on this data, thus invalidating our hypothesis **H2.1**. Participants took the same amount of time to search for visual information regardless of the crowd density present in front of them. Finally, it can be noted that regardless of crowd density, participants display mostly a similar eye-gaze behaviour, as they show a centered distribution of gaze positions toward their objective. These results are in line with previous work [Yarbus, 2013] showing that eye-gaze activity is dependent of the task. In conclusion, these results indicate

that participants' gaze is more focused toward the direction of the goal they have to reach as crowd density increases. Their scanning range decreases, showing that they visually take into account mainly pedestrians and visual cues in front of them.

## 5.5 General Discussion

### 5.5.1 Crowd Simulation

The motivation of this work was the modelling and simulation of crowd behaviour. In particular, the experiment presented in Section 5.4 explored the question of interaction neighbourhood, and how it varies with increasing densities. We first show that density has no significant effect on the duration of fixations. From the perspective of crowd modelling, this let us think that the number of interaction neighbours tends to remain constant with density. Indeed, an increase in the number of neighbours would probably have required participants to visually scan them faster. For a similar situation such as an opposite crowd in a street, we thus recommend to work with a constant number of neighbours in simulations (i.e., the number of simulated interactions for each agent). A more detailed analysis of the observed characters would however be required to evaluate whether some of them are observed multiple times.

Concerning the selection features to use to select neighbour agents, we recommend ordering them by risk of collision (which can be estimated in different ways, as discussed in [Meerhoff et al., 2018a]). Indeed, even though coverage did not significantly change with density, our results reveal that gaze tends to refocus around the visual center when density increases: this is the area where characters presenting the highest risk of collision in this bidirectional traffic condition.

### 5.5.2 Limitations

Section 5.3.3 attributes changes in eye-gaze behaviour between RE and VE to the lack of some details in the virtual scene. This interpretation is corroborated by previous studies [Slater et al., 2009]. In particular, the digital scene did not incorporate sound simulation. Locomotion and collision avoidance, is mainly controlled from visual information. As an example, Silva *et al.* [2018] showed that, with 3 pedestrians or more, the emission of sound has no effect on the manoeuvres performed by participants. However, gaze activity was not explored in this work, and a sound can certainly easily attract our attention, thus impacting some

characteristics of our gaze activity [Coutrot et al., 2012].

In the second experiment on crowd density, virtual humans had to avoid any collision with the participant. We did not want to have participants traversed by virtual characters, that would have negatively impacted immersion. In RE, the level of attention paid by surrounding people to their navigation can vary a lot. As an example, the use of cellular phone while walking might impact gait kinematics when avoiding other pedestrians [Licence et al., 2015]. These differences in the behaviour of neighbours may also induce a change in the participants' behaviour. We believe that, generally, the reduction of the field of view by the Fove HMD may have an impact on the eye-gaze behaviour. This was already outlined by previous studies [Berton et al., 2019, Pfeil et al., 2018]. However, in our case, we think that this impact was limited, especially because the task was to walk straight ahead in a street. The goal as well as the oncoming obstacles were always visible in the central vision area. Nevertheless, we are interested in extending the number of situations and to address the case of crossing traffic. The width of the field of view would certainly take a greater importance then. It is also possible that performing the task in a street (RE and VE) may have overly normalized eye-gaze behaviour, leading to little effect of density on the studied variables. In a next step, we would therefore be interested in studying such eye-gaze behaviours in more open places, in order to evaluate whether normalizing the reaction of our participants in a closed-space (street) was indeed overly constraining their eye-gaze activity. In addition, it is important to note that the method we used to compute fixations cannot be applied to all situations, as it assumes small head movements. We used this method for a fair comparison with RE where the capture of head movement in a street is challenging. With the use of VR, we would then be able to compute these motions in order to adapt our analysis to more complex scenario. Furthermore, several recent studies explored the coupled analysis of locomotion and gaze and considered, for instance, walking speed [Dietrich and Wuehr, 2019]. In this work we have focused solely on the eye-gaze activity, but we intend to focus on this type of study in our future work. Finally, it would have been interesting to have a larger number of participants in this study, so as to improve the accuracy of the fitted Gaussian models as well as to study more sensitive metrics, such as the inter-individual variability in gaze data in VR [Dorr et al., 2010], which we plan to explore in the future.

## 5.6 Conclusion

In this chapter we have carried out two experiments on the study of eye-gaze activity while walking in a street. Our first experiment was to study the activity of the gaze while navigating in a real and a virtual environment in order to evaluate the impact of VR on the eye-gaze activity. Our results show a qualitatively similar eye-gaze activity with some quantitative differences. In our second experiment we studied the impact of crowd density on a walker's gaze in VR. Our results demonstrate an influence of density on gaze deployment, where it decreases as the density increases while navigating in a street full with an opposite crowd. For such situation, this indicates that in high densities, walkers have a tendency to focus more their gaze in front of them. Consequently, they will visually take more into account people in front them than people in their surroundings. We are able to provide guidelines for the design of models of local interaction for crowd simulators.

More importantly, this work opens new perspectives for future research. The modeling of crowds raises plenty of questions. In first place, we have studied the effect of density in the case of an oncoming traffic in a street only. It is first required to explore more traffic conditions to get a deeper understanding of interaction neighbourhood. For example, in the case of crossing flows or in a more open space, we expect the gaze to explore more the peripheral areas of the field of vision. It would be also very interesting to try to saturate one's visual field with many interactions of importance (e.g., many characters all on a collision course) to explore the limits of visual integration and observe if walkers apply a specific strategy in such cases. Nevertheless, if peripheral vision gets more important in new scenarios, it would be certainly required to re-evaluate eye-gaze activity when using HMDs with a wide field of vision, that are becoming more and more popular. Finally, as we have highlighted the possible role of other sensory channels on visual attention (e.g., sound or touch), we would like to integrate higher fidelity scenes and VR rendering techniques in our experimental VR platform.

Chapter

# 6

# Haptic Rendering of Collisions During Navigation through a Virtual Crowd

## Contents


**6.1 Virtual Realty and Collision Rendering . . . . . 99**

**6.2 Experimental Overview . . . . . 101**

6.2.1 Materials & Methods . . . . . 101

6.2.2 Environment & Task . . . . . 104

6.2.3 Protocol . . . . . 105

6.2.4 Participants . . . . . 106

6.2.5 Hypotheses . . . . . 106

**6.3 Analysis . . . . . 107**

6.3.1 Collected Data . . . . . 107

6.3.2 Trajectories . . . . . 107

6.3.3 Body Motions . . . . . 108

6.3.4 Collisions . . . . . 110

6.3.5 Presence and Embodiment . . . . . 110

6.3.6 Statistical Analyses . . . . . 113

**6.4 Results . . . . . 113**

6.4.1 Trajectory Analysis . . . 113
6.4.2 Body Motion . . . 114
6.4.3 Collisions . . . 115
6.4.4 Presence and Embodiment . . . 117
**6.5 Discussion . . . 119**
6.5.1 Trajectories . . . 119
6.5.2 Avoidance Behaviour . . . 119
6.5.3 Haptic Rendering After-effects . . . 120
6.5.4 Embodiment & Presence . . . 121
6.5.5 Limitations . . . 121
**6.6 Conclusion . . . 122**

In this chapter, we are now exploring another dimension of this thesis, which is how to improve immersion while navigating in a virtual crowd in VR. More specifically, we focus on the simulation of physical contact in virtual reality (VR) when navigating in a virtual crowd, as illustrated in Figure 6.1. Indeed, when walking in a dense crowd, it is not uncommon to bump into other pedestrians. In order to record as accurately as possible human behaviour in such situations, it is then necessary to simulate these physical contacts, which is not an easy task in VR. For this reason, studies of collective behaviour in VR are often limited to cases considering distant interactions only [Ríos et al., 2018, Rio and Warren, 2014], so as not to require any haptic feedback.

In this work we employ a set of wearable haptic interfaces able to provide compelling vibrotactile sensations of contact to the user's arms. Our objective is to investigate whether and to what extent the rendering of contacts influences the user's behaviour during navigation in dense crowd, as well as limits the occurrence of certain well-known artifacts, such as when the user's virtual avatar interpenetrates other virtual characters. We conducted an experiment where participants were equipped with four wearable haptic interfaces (two on each arm), and asked to navigate in a densely-crowded virtual train station.

The rest of the chapter is structured as follow. Section 6.1 summarizes the state of the art on the topic, comparing what has been presented in the literature to what we are proposing in this paper. Then, Section 6.2 describes the experimental setup, methods, and task. Section 6.3 presents the metrics we considered, based on the study of local body movements, trajectories, energy, contacts, embodiment, and presence. Section 6.4 discusses the results and analyses the differences between the considered conditions using inferential statistical analysis methods.



**Figure 6.1** – Our objective is to understand whether and to what extent providing haptic rendering of collisions during navigation through a virtual crowd (right) makes users behave more realistically. Whenever a collision occurs (center), armbands worn on the arms locally vibrate to render this contact (left). We carried out an experiment with 23 participants, testing both subjective and objective metrics regarding the user's path planning, body motion, kinetic energy, presence, and embodiment.

Section 6.5 discusses our findings as well as their implications for crowd experiments in VR. Finally, Section 6.6 draws the final remarks and discusses future work on the topic.

## 6.1 Virtual Realty and Collision Rendering

Immersive VR must actively engage one's senses, so as to make the human user feel truly part of the virtual world. This feeling of presence is achieved through multiple types of features and information that flow from the virtual environment to the human user. Rich sensory feedback is of course a fundamental pillar of this feeling of presence in VR, which includes the rendering of virtual contacts. Pacchierotti *et al.* [2017] presented a review paper on wearable haptic devices used to render contact sensations at the fingertip and hand. Notable examples of this technology used for VR applications are [Leonardis et al., 2016, Chinello et al., 2017, Schorr and Okamura, 2017, Aggravi et al., 2018]. For example, Leonardis *et al.* [2016] presented a wearable skin stretch device for the fingertip. It moves a rigid tactor in contact with the skin, providing skin stretch and making/breaking contact sensations. This device was used to render the haptic sensation of grasping and lifting objects in VR. Chinello *et al.* [2017] presented a wearable fingertip device composed of two parallel platforms: the upper body is fixed on the back of the finger, housing three small servo motors, and the mobile end-effector is in contact with the finger pulp. This device was used to simulate contacts with arbitrarily-oriented surfaces in VR and AR [Meli et al., 2018]. Collisions are of course not

limited to the hand, especially when dealing with crowded environments. For example, Mestre *et al.* [2016] performed an experiment where participants had to walk through a door with a variable width in VR with or without haptic rendering or 3d body representation. Haptic rendering was a vibrotactile feedback at the shoulders to notify users in VR when they were in close vicinity of the door structure (from 10 cm until contact). Participants showed the same basic behaviour (shoulder rotation when passing a narrow door) with and without haptic rendering or body representation. Furthermore, the optimal behaviour to avoid collision with the door's borders required the presence of body representation and vibrotactile feedback. Louison *et al.* [2018] tested wearable vibrotactile feedback in an industrial VR training application. Participants wore a haptic sleeve on their right arm, which provided vibrotactile feedback sensations via ten vibrating motors. Users had to activate or deactivate a series of targets in VR, following a given order. Contacts between the environment and the user's arm were rendered by vibrotactile feedback, where trials with vibrotactile feedback showed a higher spatial awareness and less contacts with the environment. Bimbo *et al.* [2017] employed eight vibrotactile motors in two armbands to provide the user with feedback information about the collisions between a robotic arm and the cluttered environment it operated in. Regarding interaction with virtual characters, Krogmeier *et al.*[2019] designed an experiment where participants had to bump into a virtual character, with or without haptic rendering of contacts. This haptic rendering was performed using the "Tactsuit"[1], equipped with with 70 haptic point to render contacts. In this preliminary study, they showed that this kind of haptic feedback improves presence and embodiment. In another context, Krum *et al.*[2018] were interested in the impact of different locomotion techniques and priming haptic rendering on proxemics and subjective social measures during an interaction with a virtual character. The priming haptic rendering corresponded to a simulated touch by the virtual human. Their results showed that priming haptic rendering did not influence participant's proxemics but influenced the subjective social measures. For instance, it improved the sympathy and the relation toward the virtual character. Furthermore, Faure *et al.*[2019] asked participants to perform a collision avoidance task with a virtual character, while walking on a treadmill. They used a cable mechanism to render physical contacts, which increased the minimal distance between the participant and the virtual character, as well as made participants tend to initiate the avoidance strategy sooner.

[1] https://www.bhaptics.com/

All these studies focus on the interaction with one virtual character. However, to the best of our knowledge, no study designed an experiment where participants had to navigate in a virtual dense crowd while wearing haptic rendering devices. The use of tactile feedback therefore raises questions about their effect on participant behaviours, which sets the objective of our study as described in the next section.

## 6.2 Experimental Overview

The purpose of this study is to investigate the effect of haptic rendering of collisions on participants' behaviour during navigation through a static crowd in VR. To explore this question, we immersed participants in a virtual train station and asked them to perform a navigation task which involved moving through a crowd of virtual characters. In some conditions, collisions with the virtual characters were rendered to participants using 4 wearable vibrotactile haptic devices (motorized armbands). Our general hypothesis is that haptic rendering changes the participants' behaviour by giving them feedback about the virtual collisions. Moreover, we also expect that even after removing haptic rendering, an after-effect still persists on the participants' behaviour.

### 6.2.1 Materials & Methods

#### 6.2.1.1 Apparatus

For the purpose of immersing participants in the virtual environment and investigating the potential effects of haptic rendering while navigating in groups of characters, we used the following devices, which are summarized in Figure 6.2:

- **Motion Capture**: to record participants' body motions, as well as to render their animated avatar in the scene, we used an IMU-based (Inertial Measurement Unit) motion capture system (Xsens[2]). The Xsens system provides real-time, easy-of-use, reliable and accurate human motion tracking. IMU sensors were equipped on the participants using motion capture suit and straps, while body tracking was handled by the Xsens MVN Animate software and streamed to Unity in real time.

[2] https://www.xsens.com/




**Figure 6.2** – Devices worn by participants during the experiment.

- **HMD**: to immerse participants in the virtual environment, we chose to use a Pimax[3] virtual reality headset, in particular because of the wide field of view provided in these situations of close proximity with other characters (specifications: 90 Hz, 200°$fov$, 2560 × 1440 resolution). The HMD was used with 4 SteamVR 2.0 base stations, providing a tracking area of approximately 10×10 m. This setup enabled participants to physically walk in the real space, while their walking movements were displayed on their avatar in the size-matched virtual environment.

- **Haptic Rendering**: To render haptic collisions between participants and the virtual characters, we equipped participants with four armbands (one on each arm and forearm) [Scheggi et al., 2016]. Each armband is composed of four vibrotactile motors with vibration frequency range between 80 and 280 Hz and controlled independently. Motors are positioned evenly onto an elastic fabric strap (see Fig. 6.3). An electronics board controls the hardware. It comprises a 3.3V Arduino Mini Pro, a 3.7V Li-on battery,

[3] https://www.pimax.com/

and a Bluetooth 2.1 antenna for wireless communication with the external control station.

- **Computer**: to let participants move freely in the environment, they were equipped with a MSI VR One backpack computer, which was running the experiment. All the devices were connected directly to this computer (specifications: NVidia GTX 1070, Intel Core i7-7820HK processor, 32GB RAM).

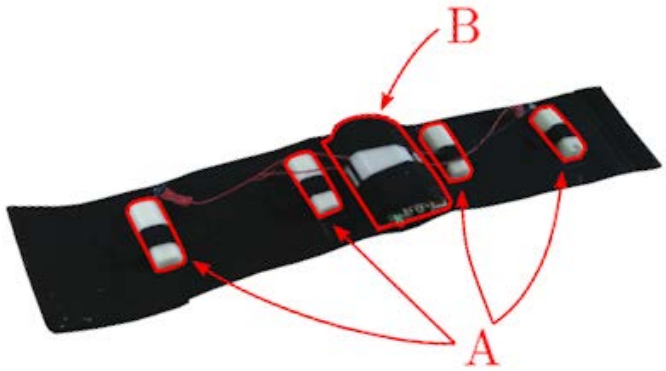



**Figure 6.3** – Wearable vibrotactile armband, composed of four vibrating motors (A). The electronics is enclosed in a 3D-printed case (B) [Scheggi et al., 2016].

#### 6.2.1.2 Haptic Rendering

Haptic rendering requires collisions to be detected in the VR environment. Since haptic devices were worn on participant's arms, we detected collisions between their avatar (animated using the Xsens motion capture system) and the virtual crowd. To this end, we segmented each avatar's arm into three parts (arm, forearm, and hand), and attached to each segment a Unity capsule collider that reported on collisions with other objects in the scene.

When a collision was detected, that is if one of the six segments of the avatar entered in collision with the geometry of any virtual crowd character, one of the four haptic devices was activated. More specifically, colliders on the left (resp. right) virtual forearm and hand activated the armband located on participants' left (resp. right) forearm, while colliders on the left (resp. right) virtual upper arm activated the armband located on participants' left (resp. right) upper arm.

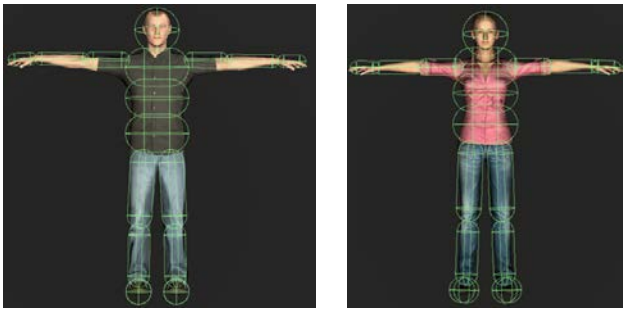
In terms of vibrations, each vibro-motor of an armband was driven using a single parameter called *vibrotactile rate*, which controlled both the amplitude and the frequency of vibration. During the experiment, all the motors of an activated armband were therefore controlled using the same *vibrotactile rate*, which varied according to a 10 Hz-period sine wave profile. The variation of the vibrotactile

rate resulted in a frequency of vibration in the range of [57–126] Hz. Although these motors can vibrate up to 255 Hz, we decided to limit their range after participants in a pilot study reported the full vibrating range to be too strong.

Communication with the armbands was performed at 4 Hz, meaning that collisions with a duration lower than 250 ms were not rendered to participants, and that there was a maximum delay of 250 ms in activating (resp. stopping) the armbands after a collision was detected (resp. ended).

### 6.2.2 Environment & Task

Participants were immersed in a digital reproduction of the metro station "Mayakovskaya" in Moscow, amongst a virtual static crowd (see Figure 6.5). A total of 8 different configurations of the scene were prepared in advance and used in the experiment. A configuration is defined by the exact position of each crowd character in the virtual station. In each configuration, the crowd formed a squared shape, and character positions followed a Poisson distribution resulting in a density of $1.47 \pm 0.06$ character/m$^2$. Such a distribution combined with such a level of density ensures that a gap of 0.60 m on average exists between each character. The crowd is composed of standing virtual characters animated with various idle animations (only small movement but standing in place). In each configuration, characters were animated according to two types of behaviour, either waiting (oriented to face the board displaying train schedules, moving slightly the upper body) or phone-calling (with a random orientation). We used several animation clips for each of the two behaviours, in order to prevent the exact same animation clip to be used for two different virtual characters.



a) b)

**Figure 6.4** – Male (a) and female (b) avatars used to represent the participants in the virtual environment. For both avatars the capsule around each segment represents the solid used to compute collisions.

At the beginning of each trial, participants were initially standing at one corner

of the square crowd, embodied in a gender-matched avatar (see Figure 6.4). They were instructed to traverse the crowd so as to reach the board displaying train schedules, and to read aloud the track number of the next train displayed on the board before coming back to their initial position. They were physically walking in the real room, while their position and movements were used to animate their avatar. This task required participants to reach the opposite corner of the space in order to read information on the board, while forcing them to move through the virtual crowd. Also, the screen displayed the train information only when participants were at less than 2 m from it (i.e., when they reached the green area displayed in Figure **??**.b). Furthermore, we provided the following instruction to participants prior to the experiment: "*Walk through the virtual train station as if you were walking in a real train station*".



a) b)

**Figure 6.5** – Snapshots of the environment under two different points of view. Participants started from the blue cross on the floor, and were instructed to reach the screen board. Figure (b) displays an example trajectory in a red doted line. The screen displayed the train information only when participants reached the green area.

### 6.2.3 Protocol

Upon arrival, participants were asked to fill in a consent form, during which they were presented the task to perform. They were then equipped with the equipment listed in Section 6.2.1.1. Calibration of the Xsens motion capture system was then performed to ensure motion capture quality, as well as to resize the avatar to participants' dimensions. Once ready to run, participants performed a training trial in which they could explore the virtual environment and get familiar with the task.

The experiment then consisted of 3 blocks of 8 trials, where the blocks were presented for all participants in the following order: *NoHaptic1*, *Haptic*, and *No-*

*Haptic2*. The *Haptic* block corresponded to performing the task with haptic rendering of contacts, while the *NoHaptic* blocks did not involve any haptic rendering of contacts. The experiment therefore consisted in performing first a block without haptic rendering, in order to measure a baseline of participants' reactions. The purpose of the second block was then to investigate whether introducing haptic rendering influenced their behaviour while navigating in a crowd, while the purpose of the last block (without haptic) was to measure potential after-effects. In each trial, participants performed the task described in Section 6.2.2 once. Each block was comprised of 8 trials, corresponding to the 8 crowd configurations presented in Section 6.2.2, performed in a random order. At the end of each block, participants were asked to answer the *Embodiment* and *Presence* questionnaires (Tables 6.2, 6.3, 6.4 & 6.5) while remaining in the virtual environment. Finally, at the end of the experiment, participants filled in a demographic questionnaire.

### 6.2.4 Participants

Twenty-three unpaid participants, recruited via internal mailing lists among students and staff, volunteered for the experiment (8F, 15M; age: avg=26 $\pm$ 6, min=18, max=43). They were all naive to the purpose of the experiment, had normal or corrected-to-normal vision, and gave written and informed consent. The study conformed to the declaration of Helsinki. The study conformed to the declaration of Helsinki, and was approved by the Inria internal ethical committee (COERLE).

### 6.2.5 Hypotheses

$H1$: Haptic rendering will not change the path followed by participants through the crowd. Indeed, pedestrians mainly rely on vision to control their locomotion [Patla, 1997, Warren, 1998], and we replicated each crowd configuration across the 3 blocks, resulting into identical visual information for participants to navigate. Therefore the followed path will be similar in the tree blocks of the experiment (*NoHaptic1*, *Haptic* and *NoHaptic2*).

$H2$: Haptic rendering of collisions will make participants aware of collisions and influence their body motion during the navigation through the crowd. Therefore, concerning the *NoHaptic1* and *Haptic* blocks of the experiment, we expect that:

$H2_1$: Participants will navigate in the crowd more carefully in the *Haptic* block in order to avoid collisions. There will be more local avoidance

movements (e.g., increased shoulder rotations) and a difference in participants' speed.

$H2_2$: With theses changes on participants' local body motions, there will be both less collisions, and smaller volumes of interpenetration when a collision occurs.

$H3$: We expect some after-effect due to haptic rendering, i.e., we expect that participants will remain more aware and careful about collisions even after we disabled haptic rendering. Therefore we expect $H2_1$ and $H2_2$ to remain true in the *NoHaptic2* block.

$H4$: Haptic rendering will improve the sense of presence and the sense of embodiment of participants in VR, as they will become more aware of their virtual body dimensions in space with respect to neighbour virtual characters.

## 6.3 Analysis

This section presents the collected data as well as the variables used to evaluate our hypotheses.

### 6.3.1 Collected Data

During the experiment, we recorded at 45 $Hz$ the trajectories of participants, as well as the position and orientation of their limbs in the virtual environment using the Xsens sensors and Unity. We also recorded the body poses over time of each character of the virtual crowd. Then we were able to replay offline the entire trials in order to compute complex operations such as the volume of each collision.

### 6.3.2 Trajectories

To study $H2$, we compared the trajectories formed by participants through the virtual crowd. To this end, we decomposed the environment into cells based on a Delaunay triangulation [Chew, 1989], the vertices of which were the crowd characters. A trajectory is then represented as a sequence of traversed cells. An example is displayed in Figure 6.6, where the displayed trajectory corresponds to the following sequence of cells: $C_{15}C_{18}C_{13}C_{31}C_5C_{30}C_2C_{34}C_4$.

Represented this way, comparison is possible only when the configuration of crowd characters is identical, which is one reason why we ensured to repeat the same configurations through the 3 studied blocks (cf. Section 6.2). In other words,

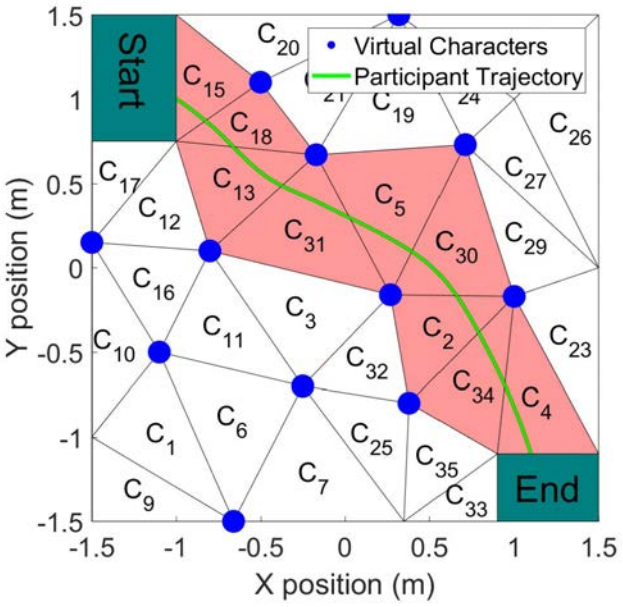



**Figure 6.6** – Illustration of a participant trajectories in a crowd, and the decomposition of the environment in cells using Delaunay triangulation [Chew, 1989].

we first grouped trajectories by crowd configuration, and then compared the set of trajectories performed in the same crowd configuration across different conditions.

Comparison was based on the Dice similarity coefficient ($DSC$) [Sørensen et al., 1948]. The $DSC$ computes the similarity between two sets $A$ and $B$ according to:

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{6.1}$$

Since our trajectories are sets of traversed cells, two trajectories traversing the same set of cells will be 100% similar. Similarity will decrease with the number of different cells traversed by the participant (occurring in one trajectory and not the other).

### 6.3.3 Body Motions

Navigating in cluttered environments, such as studied in this experiment, requires participants to weave with their bodies through the crowd. This section presents the data that will be used to analyse body movements when navigating through the virtual crowd to study $H2_1$.

#### 6.3.3.1 Shoulder Rotation

Turning the shoulders is a known strategy for squeezing through narrow openings [Wilmut and Barnett, 2010], i.e., in our case to get between two close characters. To evaluate the effect of haptic rendering on the emergence of such behaviours we measured the shoulder orientation at certain critical points of the path. These critical points are the crossing points between the Delaunay cell boundaries (cf. Section 6.3.2) and the participant's trajectories. More specifically, we computed the angle $\alpha_{SA}$ between the participants' shoulder-to-shoulder axis and the segment connecting the two considered virtual characters, as shown in Figure 6.7. This angle provides information about the orientation of the shoulders, and thus the trunk, at the narrowest parts of their path when participants passed between two characters. The larger this angle, the more careful — trying to lower their width at the maximum — participants were when traversing the opening between the two virtual characters. .

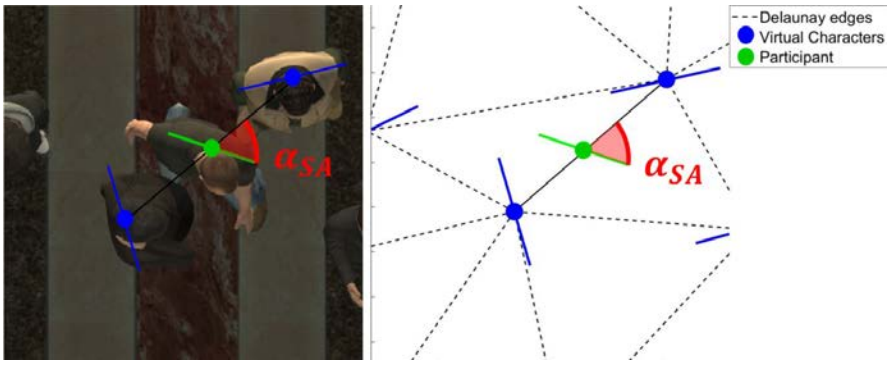



**Figure 6.7** – Shoulder rotation. Angle $\alpha_{SA} \in [0, 90]°$ is defined between the participants' shoulder-to-shoulder axis and the segment connecting the two virtual characters. Left: top view of the scene. Right: diagram with the Delaunay triangles, the virtual characters, and the participant.

#### 6.3.3.2 Walking Speed

Beyond the postural analysis introduced in the previous section, we are also interested in the walking speed to analyse whether participants performed the motion task differently according to conditions. To evaluate this parameter only during the navigation, we removed portions of trials where participants were mostly static (e.g., the time during which they were reading the board). To this end, we computed the minimum distance between the participant and the screen, which corresponds to the moment when participants stopped to read the information. We then removed all the frames when the participant's position was less than one

step from this position (chosen as $0.74m$ for men and $0.67m$ for women [Cho et al., 2004]).

### 6.3.4 Collisions

A collision is the *detected contact* between any part of the participant's virtual body and any part of the mesh of one virtual character. In order to detect them we replayed offline the experiments performed by the participants using the data recorded, as shown in the Figure 6.8. We identify a collision by the pair participant-virtual character as well as the initial time. This means that we separately classify collisions with different characters, even if they are happening at the same time. This also means that we can detect several collisions with the same character but with different initial times. The detection starts at the first contact of any of the limbs of the character involved and the participant's geometry, and it lasts until there is no more contact detected between the two respective meshes.

To analyse the collisions we selected two main values of interest: the *number of collisions* and the *maximum volume of interpenetration* between the participant and the virtual character during a collision:

- *Number of collisions.* We count any collision with an interpenetration volume greater than $10^{-6}$ $m^3$ and lasting more than 250 $ms$.
- *Maximum volume of interpenetration.* The maximum volume of interpenetration between a participant's avatar and a virtual character during a collision is computed at each time stamp through the voxelization of the intersection of their respective meshes, according to the following procedure. Each 250 $ms$ the computation starts from the meshes of the two characters involved. Around those, we build an AABB (axis aligned bounding box), which is then iteratively subdivided in octant where, at each of this octant-iteration, only the voxels in collision are kept. The octant-iteration stops when the target voxel size is reached. In our analysis, it was set to a cube of width 0.01 m. This process is shown in Figure 6.9. At the end we collect all the volumes computed at each time interval of 10 ms and we extract the maximum one.

### 6.3.5 Presence and Embodiment

Another important aspect of our analysis is its perceptual relevance. In accordance with $H4$, we looked for any difference in the users' feelings of presence and

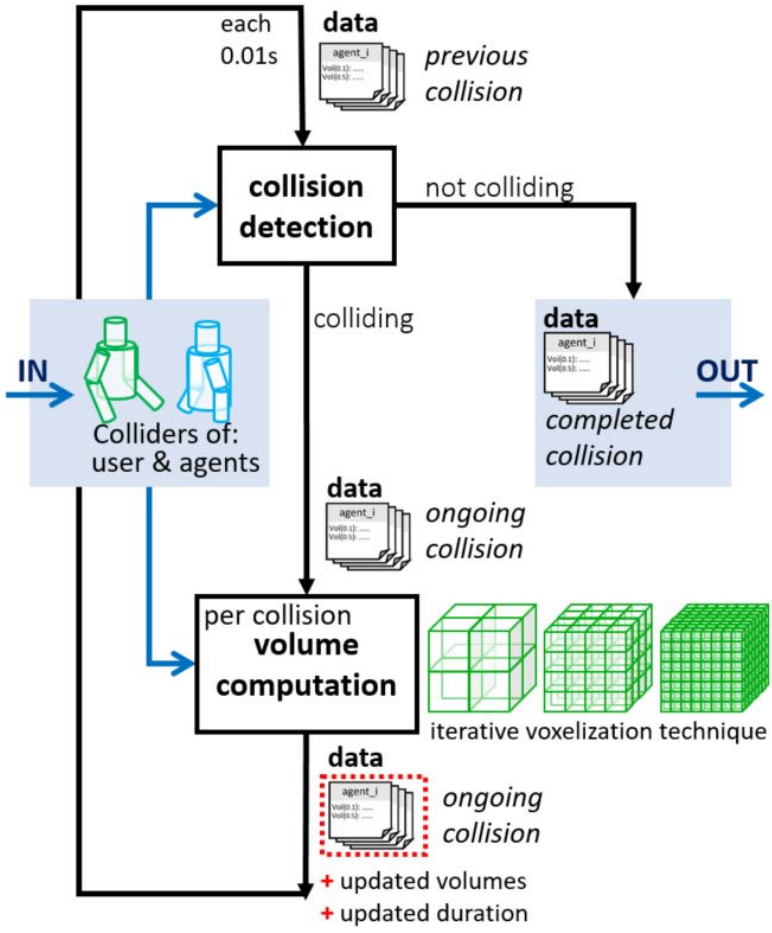



**Figure 6.8** – Collision iteration loop scheme representing one step of the collision offline detection, in which we detect if there is a collision (either a new or an ongoing one) and compute its volume. We add this information to collision's data. When the collision is finished we send out the data.



a) b) c)

**Figure 6.9** – Volume computation using iteration of voxel spaces of decreasing dimensions. (a) Starting from the AABB around the selected geometries, the first voxel space with 8 voxels (green cubes) is created and intersected with the geometries. (b) In the next iteration only the intersecting voxels are kept, and further subdivided in 8 cubes each. (c) The process is iteratively applied until reaching the minimum subdivision size, where the final interpenetration volume is displayed in purple.

embodiment, comparing the registered subjective perception with and without haptic rendering. Participants answered both questionnaires at the end of each block (Embodiment then Presence), answering each question on a 7-point Likert scale.

#### 6.3.5.1 Presence

Using an haptic device is generally expected to increase the user's immersion in the virtual world [Krogmeier et al., 2019], as it adds a new sensorial feedback, even though it does not always lead to an increase of perceived realism [Slater, 2003]. For this reason, we measured Presence using the Slater-Usoh-Steed (SUS) questionnaire [Usoh et al., 2000] (Table 6.5). Each user answered the set of 6 questions, summarized in Table 6.5, at the end of each block.

#### 6.3.5.2 Embodiment

As for Presence, we focused on comparing the sense of embodiment between different blocks to study the influence of the haptic rendering on the perception of the virtual body. We measured embodiment based on the Roth and Latoschik questionnaire [2020]. Participants answered Embodiment questionnaires simultaneously with those about Presence.

### 6.3.6 Statistical Analyses

Our objective is to understand whether and to what extent users change their behaviour in each experimental block. To do so, we analysed the differences across blocks for all the aforementioned variables. For all dependent variables, we set the level of significance to $\alpha = 0.05$. First, a Shapiro-Wilk test was performed to evaluate whether the distribution of our data followed a normal distribution. If the distribution was not normal, a Friedman test was performed to evaluate the effect of the condition on these variables. Post-hoc comparisons were then performed using a Wilcoxon signed rank test with Bonferroni correction. On the other hand, if the distribution was normal, a one-way analysis of variance (ANOVA) with repeated measures was performed. Greenhouse-Geisser adjustments to the degrees of freedom were applied if the data violated the sphericity assumption. Bonferroni post-hoc tests were used to analyse any significant effects between groups.

## 6.4 Results

This section presents the results of our experiment, starting with the study of $H1$ on the trajectories formed by participants through the virtual crowd. We then explore $H2_1$ and $H2_2$ with respect to the analysis of body movements. Finally, we report the results on collision metrics so as to evaluate $H3$, to finish with the answers to the Presence and Embodiment questionnaires related to $H4$.

### 6.4.1 Trajectory Analysis

Table 6.1 shows the results of the Dice similarity measure between all possible pairs of blocks. Similarity ranges from 84.7% (*Nohaptic1* vs. *Haptic* blocks) to 88.5% (*Haptic* vs. *NoHaptic2* blocks). The score is higher for *Haptic* vs. *Nohaptic2* blocks ($88.6 \pm 4.1\%$) and for *Nohaptic1* vs. *Nohaptic2* ($85.9 \pm 4.0\%$).

Because it is difficult to identity from this data only whether the obtained level of similarity is due to natural variety in human behaviours, or to the difference in conditions explored in each block„ we propose to measure similarity between paths belonging to the same block as follows. For each block and each configuration, we randomly divided the trajectories into two subsets and computed the Dice similarity score between them. We repeated this process 30 times (which changes the way trajectories are divided into 2 subsets). Performing this process and computing similarity over the 3 blocks resulted into 90 measures of "intra-block similarity". The obtained average value is $81.2 \pm 3.3\%$, that can be compared with

the "inter-block similarity" scores presented in Table 6.1.

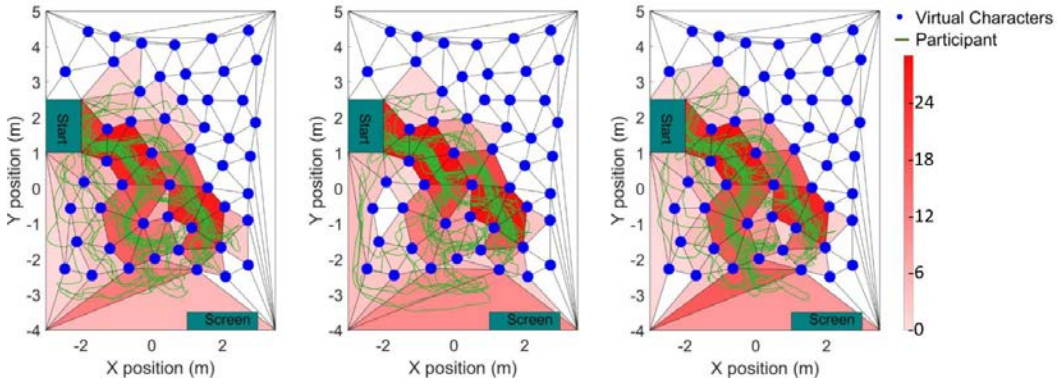



**Figure 6.10** – Participants' trajectories and Delaunay triangulation for trial $T_6$ for blocks *NoHaptic1* (left), *Haptic* (middle) and *NoHaptic2* (right). The color-bar represents the number of times participants walked on a triangle.

Our results show that there is no statistical differences between intra-block and *Nohaptic1* vs. *Haptic* blocks similarity measure ($p > 0.05$). There is however a significant difference between intra-block and *Haptic* vs. *Nohaptic2* blocks ($p < 0.01$), as well as intra-block and *Nohaptic1* vs. *Nohaptic2* ($p < 0.05$), where intra-block similarity measures are always lower. Given that similarity measures between pairs of blocks where either as similar or more similar than intra-block similarities, we can conclude that participants chose their path through the crowd similarly, irrespective of the block condition, which supports $H1$. To better illustrate the similarity in navigation paths, Figure 6.10 displays all the participants trajectories and the triangles used to compute the Dice for the specific $T_6$ configuration.

## 6.4.2 Body Motion

### 6.4.2.1 Shoulders Rotation

The average amplitude of shoulder rotations $\alpha_{SA}$, illustrated in Figure 6.11.a, was influenced by the conditions ($F(2, 44) = 13.0, p < 0.001, \eta_p^2 = 0.37$). In particular, it was significantly higher in the block with haptic rendering ($40.1 \pm 8.2°$), than in the first block without haptic rendering ($34.3 \pm 6.0°$). We remind that a higher $\alpha_{SA}$ angle means that participants made a larger rotation to squeeze between virtual characters, therefore validating the hypotheses $H2_1$. Furthermore, it was also significantly higher in block *NoHaptic2* ($38.7 \pm 3.7°$) than in block *NoHaptic1*, suggesting that participants continued to turn more their shoulders even after haptic rendering was disabled, therefore supporting $H3$.

**Table 6.1** – Similarity measure (Dice) of participant trajectories between all blocks (*NoHaptic1*,*Haptic*,*NoHaptic2*) for all the trials.

| Trials | Blocks | | |
|---|---|---|---|
| | *NoHaptic1* vs. *Haptic* | *Haptic* vs. *NoHaptic2* | *NoHaptic1* vs. *NoHaptic2* |
| $T_1$ | 84.0% | 88.6% | 85.0% |
| $T_2$ | 88.4% | 93.8% | 88.3% |
| $T_3$ | 78.1% | 93.2% | 79.4% |
| $T_4$ | 91.9% | 88.7% | 90.7% |
| $T_5$ | 88.4% | 90.2% | 85.3% |
| $T_6$ | 82.8% | 85.8% | 91.0% |
| $T_7$ | 78.8% | 81.6% | 82.0% |
| $T_8$ | 85.0% | 85.9% | 85.3% |
| $\boldsymbol{T_{all}}$ | $\mathbf{84.7 \pm 4.8\%}$ | $\mathbf{88.6 \pm 4.1\%}$ | $\mathbf{85.9 \pm 4.0\%}$ |

#### 6.4.2.2 Walking Speed

We found an effect of haptic rendering ($F(1.56, 34.2) = 7.14, p = 0.005, \eta_p^2 = 0.245$) on participant's average walking speed (Figure 6.11.b), where participants' walking speed was on average significantly lower in the *Haptic* block ($0.40 \pm 0.1 m.s^{-1}$) than in the *NoHaptic1* ($0.43 \pm 0.1 m.s^{-1}$) and *NoHaptic2* ($0.42 \pm 0.1 m.s^{-1}$) blocks. This result therefore supports hypothesis $H2_1$.

### 6.4.3 Collisions

Figures 6.11.c & 6.11.d illustrate the results regarding collision characteristics, i.e., number of collisions as well as volume of interpenetration.

The average number of collisions per trial was influenced by haptic rendering with a large effect ($F(2, 44) = 7.13, p = 0.002, \eta_p^2 = 0.25$). Post-hoc analysis showed that the number of collisions was higher during the *NoHaptic1* block ($71 \pm 29.2$) than during the *Haptic* ($62.8 \pm 34.6$, $p = 0.018$) and *NoHaptic2* blocks ($60.7 \pm 34.6$, $p = 0.002$), which shows that participants made on average more collisions before they experienced haptic rendering.

The average volume of interpenetration was also influenced by the block ($F(2, 44) = 4.35, p = 0.019, \eta_p^2 = 0.16$), where post-hoc analysis showed that this volume was smaller ($p = 0.016$) in the *Haptic* block ($0.6 \pm 0.3 dm^{-3}$) than during the *NoHaptic1* ($0.8 \pm 0.3 dm^{-3}$).

These results validate our hypothesis $H2_2$, that states that haptic rendering reduces the severity of collisions between participants and virtual characters.

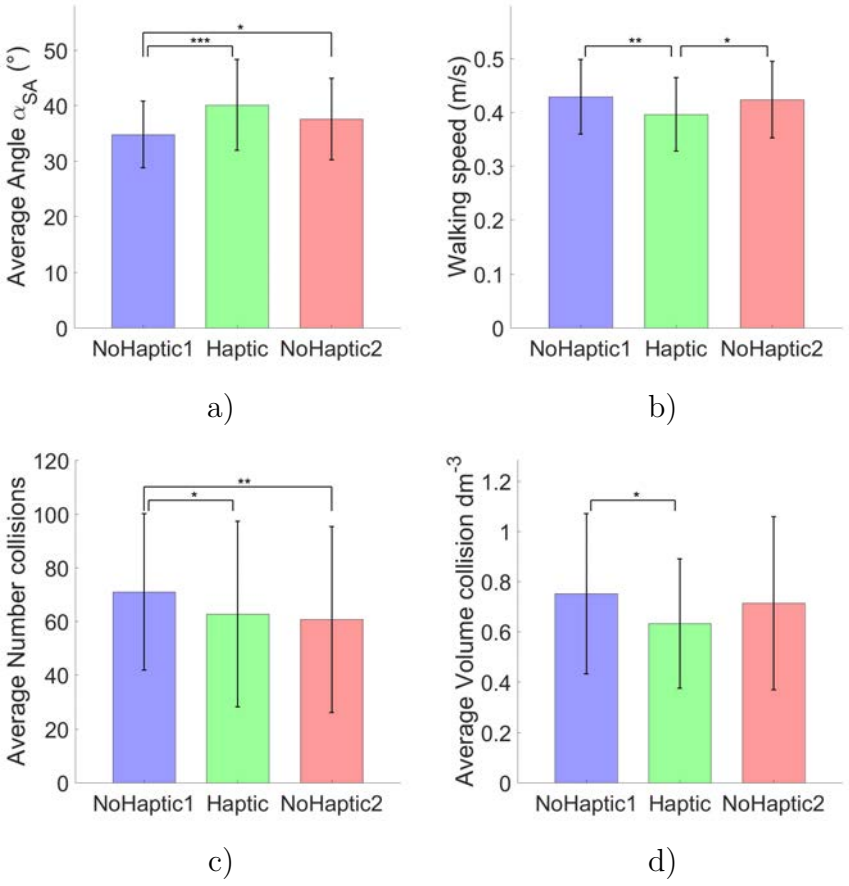



**Figure 6.11** – Main significant differences between the three blocks of the experiment (*NoHaptic1*, *Haptic* and *NoHaptic2*): a) amplitude of shoulder rotations ($\alpha_{SA}$), b) walking speed, c) number of collisions per trial, d) volume of interpenetration. Error bars depict standard deviation of the mean.

Furthermore, as the number of collisions is higher during block *NoHaptic1* than during block *NoHaptic2*, this also supports $H3$ on potential after-effects of haptic rendering.

### 6.4.4 Presence and Embodiment

The average participant ratings and all the questions for **embodiment** are shown in Tables 6.2, 6.3 and 6.4. We did not find any significant effect of the blocks for *Agency* ($p = 0.438$), *Change* ($p = 0.085$) and *Ownership* ($p = 0.753$). Furthermore, Table 6.5 shows the questions and the average participant ratings for **presence**, for which we also did not find a significant effect of the blocks ($p = 0.222$). These results therefore do not support hypothesis $H4$, suggesting that haptic rendering does not improve the sense of presence or the sense of embodiment of participants in VR.

**Table 6.2** – *Agency* questionnaire: average participant ratings for the three blocks.

| **Questions** | **blocks** | | |
|---|---|---|---|
| | *NoHaptic1* | *Haptic* | *NoHaptic2* |
| The movements of the virtual body felt like they were my movements.<br>I felt like I was controlling the movements of the virtual body<br>I felt like I was causing the movements of the virtual body.<br>The movements of the virtual body were in sync with my own movements. | $6.1 \pm 0.9$ | $6.0 \pm 0.8$ | $5.9 \pm 0.7$ |

**Table 6.3** – *Change* questionnaire: average participant ratings for the three blocks.

| **Questions** | **Blocks** | | |
|---|---|---|---|
| | *NoHaptic1* | *Haptic* | *NoHaptic2* |
| I felt like the form or appearance of my own body had changed.<br>It felt like the weight of my own body had changed.<br>I felt like the size (height) of my own body had changed.<br>I felt like the width of my own body had changed. | $3.6 \pm 1.3$ | $3.8 \pm 1.5$ | $3.3 \pm 1.5$ |

**Table 6.4** – *Ownership* questionnaire: average participant ratings for the three blocks.

| **Questions** | **Blocks** | | |
|---|---|---|---|
| | *NoHaptic1* | *Haptic* | *NoHaptic2* |
| It felt like the virtual body was my body. It felt like the virtual body parts were my body parts. The virtual body felt like a human body. It felt like the virtual body belonged to me. | $4.9 \pm 1.4$ | $5.1 \pm 1.2$ | $5.0 \pm 1.2$ |

**Table 6.5** – Slater-Usoh-Steed (SUS) questionnaire [Usoh et al., 2000] and average participant ratings for the three blocks.

| **Questions** | **Blocks** | | |
|---|---|---|---|
| | *NoHaptic1* | *Haptic* | *NoHaptic2* |
| I had a sense of being there in the train station. There were times during the experience when... the train station was the reality for me... The train station seems to me to be more like... I had a stronger sense of... I think of the train station as a place in a way similar to other places that I've been today. During the experience I often thought that I was really standing in the train station. | $5.2 \pm 0.9$ | $5.2 \pm 1.2$ | $5.0 \pm 1.1$ |

## 6.5 Discussion

The main objective of this study was to evaluate the effect of haptic rendering of collisions on participants' behaviour while navigating in a virtual dense crowd. To this end, we designed an experiment where participants had to reach a goal by physically walking in a virtual train station populated with a dense crowd. Participants were equipped with vibrotactile sensors located on their arms and performed this task following 3 blocks: *NoHaptic1*, *Haptic* and *NoHaptic2* for which respectively haptic rendering of collisions with virtual characters was not experienced, experienced, and not experienced again.

### 6.5.1 Trajectories

In Section 6.4.1, the analysis of the Dice similarity measure showed that haptic rendering did not change the way participants selected their path through the crowd, as stated in hypothesis $H1$. We even found that paths accross blocks were "more similar" than within the same block. One possible explanation is given by the way we compose the sets we compare the similarity of, where we assume that paths are independent from participants. Indeed, the intra-block similarity measure required us to split a set of trajectories belonging to the same block and crowd configuration, which resulted into comparing paths performed by different participants. In contrast, the inter-block analysis considered sets that were split according to haptic rendering conditions, thus comparing paths performed by the same group of 23 participants.

In spite of this limitation in our analysis, we consider that paths are similar across blocks. One can describe human motion as a trajectory resulting from a perception-action loop [Gibson, 1958, Warren, 1998]. Depending on the tasks, the loop is a multi-modal one, meaning that different senses are used to control motion. However in the context of walking, several studies [Patla, 1997, Warren, 1998] have shown that vision is the most used perceptual input to navigate to the goal. Such statements hold in our case, where a major difference with previous work is the higher density of obstacles. Nevertheless, assuming that tactile feedback may affect path selection, it would have been probable that some participants reversed their course after a collision has been rendered, which was not observed.

### 6.5.2 Avoidance Behaviour

In this experiment, we demonstrated that haptic rendering had an effect on shoulder rotations, which supports hypothesis $H2_1$. In particular, participants rotated

more their shoulders when traversing the gaps between virtual characters during the *Haptic* block than during the *NoHaptic1* block. This result is consistent with the obsevrations of Mestre *et al.* [2016] with participants passing through a virtual half-open door with or without haptic rendering. More generally, let us remind that the human trunk is most often larger along the transverse axis than along the antero-posterior axis. Thus, the more the participants turn their shoulders the smaller the volume swept by their body motion. Our results therefore suggest that participants might have tried to minimize the risk of collision with virtual characters more in the condition where they experienced haptic rendering than in the first block of the experiment. The slower speed observed in the *Haptic* block also reveals that participants moved more cautiously.

Being more cautious effectively resulted into less collisions as expected in hypothesis $H2_2$. Results presented in section 6.4.3 show that the average number of collisions as well as the average volume interpenetration were significantly lower in the *Haptic* block than in the *NoHaptic1* block. Furthermore, this observation is consistent with previous studies [Louison et al., 2018] where haptic feedback lowered the number of collisions with a static object.

### 6.5.3 Haptic Rendering After-effects

While there were less collisions and more shoulder rotations observed in the *Haptic* block in comparison with the *NoHaptic1* block, there was no difference between the *Haptic* and the *NoHaptic2* blocks. This supports hypothesis $H3$ on potential after-effects of haptic rendering. However, such an after-effect did not equally influence all measurements, such as walking speed that increased again in the *NoHaptic2* block. One possible explanation might be a perceptual calibration of the participants. During the experiment, participants became more familiar with the environment, the task to be performed, but also the virtual representation of their body and the virtual environment, enabling them to move faster and better avoid collisions with the virtual characters in the last block (*NoHaptic2*). Another point to highlight is that participants, at the beginning of the *Haptic* block, did not know that contacts would now trigger a vibrotactile haptic sensation. For this reason, we might expect to see a short learning phase at the beginning of the block, where participants learn to deal with the newly-rendered haptic collisions. Considering this point, we can expect the effect of providing haptic sensations of collisions even stronger than registered. However, to provide a more definitive conclusion on the role of the haptic after-effect would require to add a control group with no haptic rendering throughout the 3 blocks of the experiment, which

could be explored in future work.

These results can also open perspectives regarding the design of new experiments including haptic priming tasks. In a recent study, Krum et al. Krum et al. [2018] showed that haptic priming of collision had no effect on participants' proxemics and more precisely on distances with a virtual character. It is important to note that the task was different: it included an interaction with one virtual character and there were no risks of collision since the virtual character never came very close to the participant. It would be interesting then to re-evaluate such influence when the intimate space is violated by a virtual character.

### 6.5.4 Embodiment & Presence

In contrast with our hypothesis $H4$, we did not find any significant change in terms of user's perceived senses of embodiment and presence when experiencing haptic feedback. This result is quite surprising, as we did find significant effects in other measurements, suggesting that participants took different actions when provided with haptic sensations of contact. An explanation for this result could lie in the fact that users already registered high embodiment and presence levels without experiencing haptic feedback in the first condition (*NoHaptic1*), leaving little room for improvement in the *Haptic* condition. Another possibility is that vibrotactile feedback is not suited to render collisions in crowds, although there are several examples of this type of feedback being used to render similar events [Bimbo et al., 2017, Devigne et al., 2020]. Fututre works should consider the question of the plausibility [Slater, 2009] of such haptic feedback in a situation where participants can bump into virtual agents. For instance, unplausible haptic feedback may indeed brings participants close to the uncanny valley of haptics introduced by Berger *et al.* [2018], and thus would have an impact on the presence and embodiment of participants. Finally, a last explanation could be the location and number of our haptic devices. Employing a higher number of bracelets spread throughout the body might better render the target contact sensations. All these considerations will drive our future work.

### 6.5.5 Limitations

Our study had a few limitations. As explained above, we employed a limited number of haptic rendering devices located on participants arms only. It is quite possible that employing more devices, including some for the legs and hips, would have resulted in stronger effects. However, our setup still revealed significant effects, and the question of nature, number, and location of haptic devices would

probably require a fully dedicated study. Another related issue is the *quality* of the provided haptic sensations. Our devices show high wearability and portability, but can only provide vibrotactile haptic sensations. Other haptic delivery options include the use of arm or full-body exoskeletons, which can provide well-rounded force sensations. However, these devices are significantly more cumbersome and expensive that those employed in this work, severely limiting their applicability and availability.

A second limitation concerns the behaviour of the virtual characters present in the crowd. Indeed, they do not react to collisions, as noticed by some participants in their feedback. This could also be an explanation for the lack of significant differences in embodiment and presence scores. It would therefore be required to have an animation technique capable of reacting to collisions such as, for instance, the virtual character taking a step in the opposite direction of the collision. We could also trigger verbal reactions to express that virtual characters are embarrassed by collisions. Adding such virtual behaviours combined with haptic feedback could improve participants' immersion and feeling of presence.

Finally, one last point concerns the many VR devices (armbands, MSI VR one, HMD, X-Sens, etc.) required to be worn by participants for a significant amount of time. Carrying such equipment can have an effect on participants' motion as well as comfort. In our case, the experience was still relatively short and lasted only for 15 to 20 minutes. However, longer immersion durations might require to use wireless HMD solutions instead, even if this today means decreasing the field of vision.

## 6.6 Conclusion

In this chapter, we designed an experiment to evaluate the effects, as well as the after-effects, of haptic rendering on a motion task in a highly crowded environment. Participants performed a goal-directed navigation task through a dense virtual crowd. Wearable haptic devices provided them with vibrotactile feedback whenever a collision occurred with their arms. Results showed that providing haptic feedback impacted the way participants moved through the virtual crowd. They were more cautious about the collisions they provoked with virtual characters, but they did not change their global trajectories. We also demonstrated the presence of an after-effect of haptic feedback, since changes in their movements remained after haptic feedback was disabled. Finally, quite surprisingly, we did not notice any impact of haptic rendering on the perceived Presence and Embodiment. These results show that visual information is probably the main sense

used for navigation in dense crowds. However, a combination of visual and haptic feedback improves the overall realism of the experience, as participants show a more realistic behaviour: they are more cautious about not touching virtual characters. For this reason, we therefore suggest using haptic rendering to study human behaviour and locomotion interactions that may lead to contacts.

For future work, we are interested in populating our virtual environments with more interactive and reactive virtual characters. This is a crucial aspect since it seems to be a requirement to further improve the feeling of presence of participants. Also, this may increase the effect of haptic rendering, since we may expect stronger participants reactions after a collision is rendered in the case where virtual characters also react. A more detailed analysis that evaluates motion before and after a collision is rendered and a virtual character reacts would then also be relevant to develop. Finally, we plan to use more compelling wearable haptic devices to provide a more realistic sensation of collision while keeping the overall system compact and easy to wear, e.g., skin stretch or tapping devices for the shoulder and upper arm.

Chapter

# 7 Conclusion and Perspectives

## Contents

**7.1 Virtual Reality as a Relevant Tool to Study Human Behaviour . . . . . 126**
7.1.1 Experiments Lessons . . . . . 126
7.1.2 Recommendations to Design Virtual Experiments . . . . 128
**7.2 Future Works . . . . . 131**
7.2.1 Short Term Perspectives . . . . . 131
7.2.2 Long Term Perspectives . . . . . 132

The main objective of this thesis was to design and evaluate a new experimental platform in virtual reality to analyse human behaviour when navigating in crowds. The main purpose was to improve current crowd simulators by the acquisition of new data to understand human behaviour within a crowd and a secondary purpose was to increase the level of immersion of a user in such a complex situation. In particular, in the context of studying collective behaviours, we were interested in understanding the interaction neighbourhood, namely who is influencing ones' walking trajectory. From results of previous works Meerhoff et al. [2018b], We hypothesised that such pedestrians are the ones we are looking at. We were therefore interested in the joint analysis of both walking trajectories and gaze activity. Since the design of such a multiple-interaction study is very challenging in real conditions, we considered VR as a promising tool to study human behaviour

within a crowd as it provides a strong control of the experimental conditions. However, little was known about the ability of VR to reproduce conditions that allow the fidelity of human behaviour, especially regarding gaze activity, when interacting within pedestrians. That was the focus of our first two experiments in this thesis. Subsequently our work has focused on the impact of crowd density on gaze activity. While these two experiments considered mainly the visual input in the control of navigation in a populated environment, our last study investigated the effect of haptic rendering of collisions in VR, in order to simulate the physical contacts that occur when navigating in a dense crowd. Indeed these contacts are key features in social interaction and can help improving participant's immersion.

## 7.1 Virtual Reality as a Relevant Tool to Study Human Behaviour

### 7.1.1 Experiments Lessons

The first contribution of this thesis was to conduct a collision avoidance experiment between a participant and a confederate. This experiment took place in a physical and a virtual environment, with four different VR set-ups being used for the virtual environment. Our results show that the participants' gaze activity and walking are qualitatively similar between these five conditions. These results are consistent with previous studies on locomotion. As for the gaze activity, the elements present in the environment were gazed at in the same proportions by the participants during these five conditions. The participants also behaved in a similar way, with their gaze being directed towards the confederate at the moment when he was visible. There are, however some quantitative differences, especially more head rotation and angular range of eye movements in virtual reality. These results can be explained by a difference in the field of view between tasks performed in a virtual or physical environment where participants have access to their peripheral vision.

However this scenario studied was an artificial situation simulating a standard collision avoidance; and participants were in a relatively simple environment that included a target, walls and a confederate. Therefore, we subsequently focused on a more complex situation during our second contribution. In a first experiment, participants had to navigate through a crowd in a street, thus representing a daily life situation. Our research focused mainly on the analysis of fixations and our results are consistent with our previous study. In particular, we found a

centred distribution of these fixations in real or virtual environments. However, some significant differences are present for the duration and amplitude of these fixations. As in our first study, the reduced VR field of view may contribute to these differences. We also consider that the lack of realism may influence the gaze activity. In contrast with our first study and despite the use of a faithful reproduction of the street, many details were not present in the virtual environment. Indeed, the atmosphere of a commercial street includes a variety of elements. For instance, the sounds of the city (people, cars), the smell coming from restaurants and french bakeries, the different emotions and expressions of the people in the crowd, and also the occasional presence of animals (dogs on a leash, pigeons...). Being able to recreate these details visually could have an impact on the gaze activity and extending to an olfactory simulation using other perceptual systems could influence the attention of the participants. However, adding all these details in a virtual environment can be complex for some of them and may be impossible for others. In addition, these details can also have an impact on the quality of the experiment, as it is recommended to maintain a minimum framerate of $90hz$ in VR.

During a second experiment using the same virtual environment, we manipulated the density of the crowd in order to study its impact on the participant's gaze activity. Our results show that the gaze activity was similar during the whole experiment with a centred distribution, however the area covered by the gaze was larger for low crowd densities. This result could therefore indicate a difference in the shape of the interaction neighbourhood for crowd simulators. For instance, the interaction neighbourhood could be represented by a field of view with a width related to the density of the crowd in front of the agent. Our results also showed that the distribution of fixations could be modeled by a Gaussian. Thus, the contribution of the agents present in this field of view on the motion of the current agent could be weighted using this same model. We also found that the frequency of fixations was similar regardless of crowd density. This result could suggest that the choice of the number of neighbours to consider for interaction neighbourhood is not dependent on the crowd density. Further analysis is therefore needed to pursue these two lines of evidence.

However, as we wanted to continue analyzing human behaviour in virtual dense crowds, we considered using others perceptual systems in VR to improve the immersion in such situation. In particular, our latest contribution involved the introduction of haptic feedback about collisions during navigation in a virtual crowd with the help of vibrotactile armbands to simulate physical contact on the arms with the goal to improve the immersion in VR. During this experiment,

the haptic feedback was successively absent, then present, and again absent. The results show, first of all, that the participants' trajectories are similar in all three conditions. The visual stimulus being the same during the three conditions, this result is therefore consistent with previous work indicating that vision is the main perceptual system used during navigation. However, significant differences were found for the participant's walking speed and shoulder rotation. These results show that the participants demonstrated a more realistic behaviour when navigating through dense crowds where it is necessary to make such movements in order to maneuver through the crowd. Nevertheless, there was no significant difference in the participants' measures of presence and presence. One possible explanation is that the participants recorded high values for both of these measures under all conditions, perhaps because of a virtual representation of their bodies. It is also possible that this kind of haptic feedback is not the most suitable for simulating physical contact when navigating in a crowd.

All of these results therefore indicate that VR is an interesting and appropriate tool to study human behaviour when navigating in crowds. In particular, the first two studies validated this statement for the gaze activity. However, it is important to keep in mind that although human behaviour in a real or virtual environment may be similar, it is not entirely identical. For instance, there are differences in the perception of distances, walking speed, head rotations, etc. This is why it is necessary to take into account the existing differences when transposing the results from a virtual to a real situation.

### 7.1.2 Recommendations to Design Virtual Experiments

During this thesis, we designed several virtual reality experiments with the purpose of analysing human behaviour during interaction with virtual humans. Our past experiences enable us to provide a number of recommendations to implement such experiments. We used a wide range of VR systems: HMD (Fove, Vive, Pimax...), Cave and Desktop, each with their own advantages and disadvantages. The duration of the experiments was rather variable, and the virtual environments chosen were either simple or complex (faithful reconstruction of real places and crowds of varying density). However, these experiments, like the majority of VR experiments, have common constraints. The main one is to guarantee a minimum frame rate of $90 fps$ in order to prevent participants from getting sick. This constraint is directly related to the computer performance used to run the experiment. Indeed, the components of this computer (processor, graphics card,

RAM, etc.) allow a limited number of resources to be allocated. In order to respect this constraint, it is therefore necessary to make a compromise between several categories of elements. In the context of research on navigation in virtual crowds we propose four different categories with different levels of computational costs, illustrated in Figure 7.1.

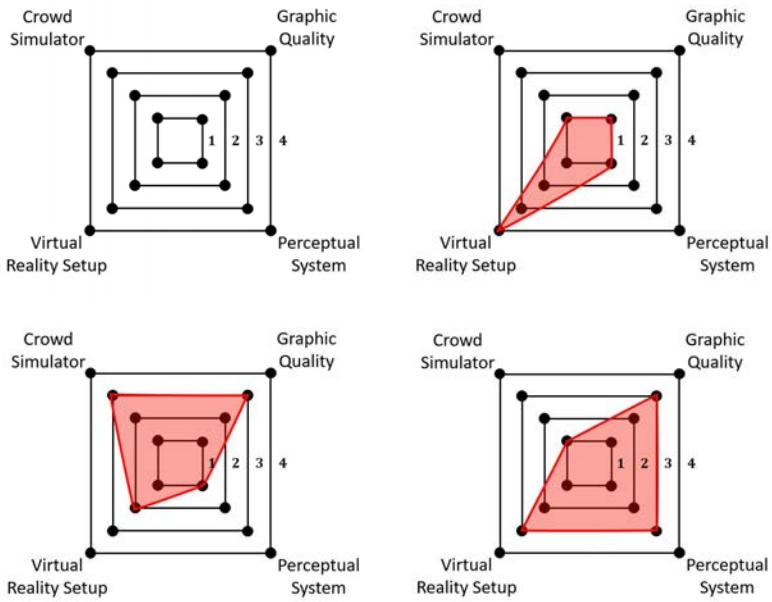



**Figure 7.1** – Top-Left: Plot of the compromise between **Virtual Reality Setup**, **Graphic Quality**, **Crowd Simulator** and **Perceptual System** in order to design an VR experiment, with a scale between low and high computational costs. Illustration of the compromises made for our first (Top-Right), second (Bottom-left) and third (Bottom-Right) experiments

- **Virtual Reality Setup**: As explained in section 2.3.1, there is a wide variety of devices for displaying a virtual environment. Each of these devices has a different display resolution, resulting in different computational costs. In particular, the computation cost is low for a screen (scale of 1), and very high for a Cave (scale of 4). For HMDs, it depends on the HMD's resolution, for instance the Vive (scale of 2) has a resolution per eye of $1200X1080$ pixels while the Pimax 5k (scale of 3) has a resolution of $2560X1440$ pixels per eye.

- **Graphic Quality**: Graphic rendering is an element that can quickly become very expensive in terms of computation costs. In order to have a virtual scene as realistic as possible, it is possible to add a lot of rendering effects. For example, we can add natural light, or add different kinds of

rendering for each element in the scene in order to simulate possible reflections on the surface of the objects. For virtual humans, it is also possible to have a wide range of animations, expressions and clothing. In the end, with each addition to improve the visual and immersive quality of the scene, the general calculation cost will also increase.

- **Crowd Simulator**: There are many different crowd simulation algorithms (see Section 2.1), each with its strengths and weaknesses, but above all with significant differences regarding computation costs. For instance, vision-based models tend to require more computation than physics-based models. In addition, this computation cost is directly related to the number of virtual agents present in the crowd and to parameter choices for crowd simulators.

- **Perceptual System**: Immersing a participant in a virtual world is not just a visual simulation of that environment. It is currently possible to simulate other sensory information, such as the addition of ambient sound or geolocalised 3D sound in the virtual environment. Some work (see section 6.1), such as our third study, focus on the simulation of physical contact Mestre et al. [2016], and there are also studies [Sardo et al., 2017, Harley et al., 2018] on the addition of smell or taste in virtual reality. However, each simulation of sensory information will also increase the computation cost during VR immersion.

When designing a VR experiment to study human behaviour while interacting with virtual agents, it is important to choose the most appropriate compromise between these different categories. This choice is directly dependent on the purpose of the experiment and the factors being studied. For instance, in the case of our second study we were interested in the impact of crowd density on gaze analysis. We therefore decided to allocate more resources to the crowd simulator (Figure 7.1-Bottom-Left) which was an key element for this study. As for our third study, we were interested in haptic feedback in order to simulate physical contact and self-body representation for participants. A static crowd was then sufficient for this experiment, therefore we were able to reduce the resources allocated to the crowd simulator in order allocate more ressources for perceptual systems, as shown in Figure 7.1-Bottom-Right.

## 7.2 Future Works

### 7.2.1 Short Term Perspectives

In our last two experiments, we analysed human behaviour in the complex situation of navigating in a crowd. These studies enabled us to record a large amount of data, thus allowing us to carry out more in-depth analyses in order to improve crowd simulators. Especially in our third experiment, we showed that haptic rendering did not affect on trajectories, giving us a large amount of trajectory data in very dense crowds. It would therefore be interesting to analyse all of these trajectories in order to better understand navigation in a dense crowd. Several studies [Hackney et al., 2013, 2015] have analysed walking strategies when crossing apertures. We could therefore apply this kind of analysis to our situation by considering the space between two virtual agents as an aperture.

In our second contribution on the impact of crowd density on gaze activity, we focused our analysis solely on eye movements and fixations. A first direction for further investigation is to look at the elements that participants looked at. One would expect to find that participants' gaze would fall proportionally more on virtual agents during conditions with high crowd density as they cover a larger area in the field of vision. Moreover, during our study, we showed that the frequency of fixations was not impacted by density. An analysis of the number of different agents looked at would thus enable us to have an estimate of the number of agents to consider in the interaction neighbourhood according to the density. Subsequently, it would also be interesting to analyse in detail the characteristics of the virtual agents looked at (e.g. risk of collision, distance at which it is located), in order to improve the model of this interaction neighbourhood. A second line of investigation is to focus on the coupled analysis of locomotion and gaze activity. In particular, we want to identify when participants performed a maneuver and analyse the gaze activity prior to this maneuver. One of the questions of interest is whether, before performing this maneuver, the participants look at the space that allows them to avoid a collision or they look at the agents with whom they may collide. Furthermore, we are also curious about the patterns in the participants' gaze activity in order to better understand the decision to perform these maneuvers.

### 7.2.2 Long Term Perspectives

In our future work, we intend first of all to continue the analysis of gaze activity when navigating in a crowd. In particular, we intend to extend our studies to different crowd movements (Figure 7.2), with for instance: unidirectional crowd movement in opposite direction or not, crowd crossing, random trajectory... This study will help us to generalise our results for any kind of crowd and thus be used to improve crowd simulators. Subsequently, we want to focus on the impact of pedestrian visual characteristics on human locomotion.

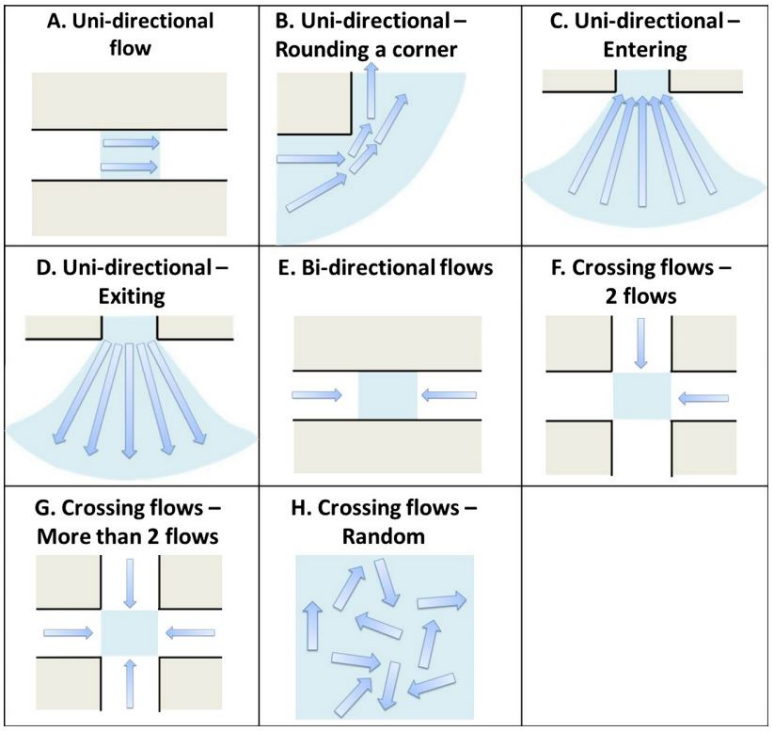



**Figure 7.2** – Illustration of crowd movement base cases from Duives *et al.* 's work [Duives et al., 2013]

In the literature, several works [Bönsch et al., 2018, Huang and Wong, 2018, Volonte et al., 2020] have already been done on the impact of virtual agent emotions on the execution of a VR task. With this approach, we intend to study the influence of crowd agent emotions on locomotion and gaze activity. Indeed, we believe that our behaviour is different if, for instance, we walk in a crowd and meet pedestrians demonstrating aggressive or happy behaviours. Finally, we also intend to make full use of the VR benefits to study situations that are more difficult to study in a real environment. In particular, we intend to study situations such as evacuation scenarios for which we believe that the

analysis of gaze activity will contribute to understand the manoeuvres performed by participants. One of the main issues in the realisation of such experiments is the creation of an atmosphere that will provide a complete immersion of the participants. Several studies on interactions with pedestrians have shown that such atmosphere (fire Ríos and Pelechano [2020], festive activity [Duverne et al., 2020]) has little influence on human behaviour. However, they had only focused on a visual modification of the scene, thus the incorporation of other systems of perception, such as the addition of haptic feedback, could improve the immersion of the participants.



**Figure 7.3** – Pictures of the movie“Ready Player On” where the protagonist is completely immersed in RV

In the end, VR offers an extensive range of possibilities to study and better understand human behaviour in a wide variety of situations. We are still a long way from the technology of a full VR immersion as in the movie “Ready Player On” (Figure 7.3), but each technological advancement improves participant immersion and thus the potential viability of the studies performed. However, VR introduces also some biases into this behaviour which needs to be taken into account before transposing the results to real life situations.

# Acronyms

**CAVE** Cave Automatic Virtual Environment. 135

**HMD** Head-Mounted Display. 135

**MPD** Minimum Predicted Distance. 135

**PS** Personal Space. 135

**RE** Real Environement. 135

**TTCA** Time To Closest Approach. 135

**VE** Virtual Environement. 135

**VR** Virtual Reality. 135

# List of Figures

1.1 Left: battle scene from the Game of Thrones series using Goalem Software's crowd simulator, Right: MassMotion crowd simulator developed by Oasys and used to test the ability of the Arup Liverpool office to comply with physical distancing measures during Covid. . . . . 2

1.2 The Perception and Action loop proposed by Gibson [1958]. . . . 4

2.1 This thesis, on the analysis of human behaviour when navigating in a crowd, is at the intersection of three main domains: **Crowd Simulation**, **Human behaviour Analysis** and **Virtual Reality** 8

2.2 Illustration of the three main approaches for crowd simulation **Flow-based** [Treuille et al., 2006] (left), **Data-driven** [Lee et al., 2007] (middle) and **Agent-based** [Dutra et al., 2017] (right) . . . 9

2.3 Illustration of physics-based model (Left [Helbing and Molnár, 1995] and Right [Karamouzas et al., 2014]) on a agent (red-circle), with the repulsive forces from the other agents (green arrows), the attractive force (red arrow) toward the goal and the final velocity vector (black arrow). . . . . 12

2.4 Illustration of the set of velocities $VO_B^A$ that cause a future collision for an agent A with an agent B (left), and extension to a case with multiple interactions (right) . . . . . 13

2.5 Illustration of the main loop used in Lopez *et al.* [2019] to update and move an agent. . . . . 14

2.6 Proxemics distance . . . . . 16

2.7 Illustration of two systems to record trajectories during experiments about interactions between pedestrians. Motion capture system (e.g, Vicon) are used in the left image from Olivier *et al.* 's experiment [2012], whereas videos and annotations are used in the right image from Boltes and Seyfried [2013] works . . . . . 17

2.8 Setup used during Olivier *et al.* 's experiment [2012]. The participant (red) and the confederate (blue) start to see each other at time *tsee*. During the experiment the minimum distance between them (*mpd*) takes place at the time *ttca*. . . . . 18
2.9 Collision avoidance between two pedestrians can be described as a 3 step process (Observation (blue), reaction (red) and regulation (green)) when considering the evolution of the minimum predicted distance (*mpd*), which is the future distance of closest approach. . 18
2.10 Left: laboratory setup used during Seyfried *et al.* 's experiment[2005a]. Right: Resulting fundamental diagram which represents the relation between $N$ participants velocity and density with $N \in [15, 20, 25, 30, 34]$ . . . . . 20
2.11 Anatomy of the eye . . . . . 21
2.12 Illustration of the different regions of the visual field. With the *fovea* ($0° - 5°$), the *parafovea* ($5° - 10°$), the near-peripheral vision ($10° - 30°$), the mid-peripheral vision ($30° - 60°$) and the far-peripheral vision ($60° - 100°$). . . . . 22
2.13 Illustration the three eye movements: **Pursuit** (left), **Vergence** (middle) and **Saccade** (right). For each illustration, the eye vectors are in red, the gaze points are in green and the rectangles represent the objects being looked at. . . . . 24
2.14 Top: Illustrations of the different indoor (A&D) and outdoor (B&C) types of ground used during the experiment of Thomas *et al.* [2020]. Bottom: Mean angle for eye (red), head (blue) and combination of eye and head (gray) for the four different ground. 26
2.15 Experimental setup (oval corridor) used during the experiment of Jovancevic *et al.* [2009]. The participant is represented by the red dot and the confederates by the black squares and arrows. The top image represents a point of view of the participant during the experiment with his gaze represented by the red crosshair. . . . . 27
2.16 Schematic overview of the experimental setup used for Hessels *et al.* 's study [2020], with participant (light blue arrowhead), static obstacles (yellow) and three groups of pedestrians (orange arrowheads, with group leaders in green arrowheads). The black arrow represent the path participants had to take. . . . . 28
2.17 From left to right several virtual reality devices through time with: Sensorama (a) in 1956, SegaVR (b) in 1991, Vive (c) in 2019 and Immersia Cave (d) in Rennes . . . . . 30

2.18 Plot of the compromise between experimental control and ecological validity for traditional methods and VR according to Loomis *et al.* [1999] . . . 31
2.19 Evaluation framework (center) presented by Cirio *et al.* [2013], with the reference trajectories in the real world (left) and trajectories in the virtual environment (right) . . . 32
2.20 Real (right) and virtual (left) environment in which participants have to estimate the distance between themselves and the building during Plumert *et al.* 's experiment [2005]. . . . 33
2.21 Board of light used during Pfeil *et al.* 's experiment [2018] in order to investigate the differences in gaze activity between real (left) and virtual (right) environment. . . . 34
2.22 Example of several scenarios used during Nummenmaa *et al.* 's experiment [2009]: agent's head and eyes facing forward (left), agent's head and gaze towards participant's right (middle-left), agent's head forward and gaze toward participant's left (middle-right), agent's head and gaze toward participant's left (right) . . . 36
2.23 Illustration of the global (center of gravity represented by a sphere or a cylinder) and local (full human body, trunk or legs) visual cues used during Lynch *et al.* [2017] experiment. . . . 37
2.24 Illustration of positive and negative facial expression of virtual agents used during Volonte *et al.* 's experiment [2020]. . . . 38
2.25 Illustration of initial condition for Bruneau *et al.* 's experiment [2015]: example of group size and interpersonal distance between pedestrians (top row) and group appearance (bottom row). 39
2.26 Illustration of Rio *et al.* [2018] experiment, with participant view in VR (left), aerial view of the virtual environment (middle) and diagram of a heading perturbation (right). . . . 39

3.1 Illustration of participant who performs an experiment in VR with CrowdMP running on the computer located on the participant's back. . . . 44
3.2 Organisation of participant and trial files used in CrowdMP, and example of a trial XML file. . . . 45
3.3 Code organization of the platform . . . 46
3.4 Example of the elements used to designed our second experiment [Berton et al., 2020]. . . . 47
3.5 Workflow during a regular trial main loop . . . 48

3.6 From left to right list of virtual reality devices supported by CrowdMP: HTC Vive (a), Fove (b), Pimax (c) and Immersia (d). 49
3.7 Tobii pro glasses 2, a pair of glasses that record the gaze activity of a person wearing them and what he/she is seeing. . . . . 50
3.8 Tobii Controller Interface . . . . . 51
3.9 Buttons order to start a recording . . . . . 52
3.10 Architecture of the data extracted from the recordings of trials with the TobiiController. . . . . 53
3.11 Illustration of a recording from a participant wearing the Tobii Glasses. The current (red) and precedent (blue) fixations are displayed with also all the objects detected (black bounding boxes) by the algorithm Yolo [Redmon and Farhadi, 2018] with the name of their classes. . . . . 54

4.1 Conditions studied in this chapter. (a) Real situation, (b-e) Participants avoiding a virtual character while (b-c) wearing a HMD and (b) walking or (c) navigating using a game controller, (d) standing in a CAVE and navigating using a game controller, (e) interacting with a computer Screen. . . . . 56
4.2 Participants were asked to navigate towards a target, while avoiding any collision with a real or virtual confederate crossing their path perpendicularly. Walls prevented participants from seeing the real or virtual confederate before reaching their comfort speed. . . . . 59
4.3 Real and virtual environments (a-b) and their respective segmentation (c-d). **Walker** is in green, **Target** in white and the rest is considered as **Environment** . . . . . 62
4.4 Schematic model used to calculate $mpd_{CA}$ . . . . . 63
4.5 Illustration of the 4 angles we computed to relate head and eye movements. . . . . 65
4.6 Average clearance distance (±SD) depending on the experimental condition. Significant post-hoc comparisons are highlighted with stars (***p<0.001) . . . . . 67
4.7 Kinematics adaptations to avoid a collision during the interaction are depicted through minimum predicted distance (mpd) (top) and its temporal derivative (bottom). . . . . 68
4.8 Number of fixations per second during the interaction for each condition . . . . . 68
4.9 Average duration of fixation during the interaction for each condition. 69

4.10 Average (± SD) gaze allocation towards the confederate walker (left), the target (center) and the environment (right) during the interaction period. . . . . 69

4.11 Evolution of the four angles $\text{Angle}_{HH}$ (top left), $\text{Angle}_{GH}$ (bottom left) , $\text{Angle}_{GW}$ (top right) and $\text{Angle}_{GT}$ (bottom right) during the normalized interactions for all the conditions. SPM analysis showed an effect of the condition on each angle. Significant pairwise post hoc results are depicted above each figure where horizontal black lines indicate during which period of time the angle differs between 2 conditions. . . . . 71

5.1 Our objective is to analyse eye-gaze activity within a crowd to better understand walkers' interaction neighbourhood and simulate crowd behaviour. We designed two experiments where participants physically walked both in a real and virtual street populated with other walkers, while we measured their eye-gaze activity (red circle). We evaluated the effect of virtual reality on eye-gaze activity by comparing real and virtual conditions (top row) and investigated the effect of crowd density (bottom row). . . . . 78

5.2 Setup used to collect gaze location for both real (a) and virtual (b) conditions. For each condition, 2D gaze location is displayed in the image recorded by the real or virtual camera (in black). . . . . 82

5.3 Generation of the virtual scenarios. a) Video recording when a participant walked in the real street. b) Density of people seen by participants over the normalized duration of the trial, estimated by tracking people visible in the video recording using deep learning algorithm. c) Virtual scenario, reproducing qualitatively similar situations in the virtual conditions in terms of virtual characters encountered, as seen by the participant. b) Density of individuals actually seen by the participant over the normalized duration in both the real (blue) and the virtual trial (red). . . . . 85

5.4 Average number of people seen by participants across trials for both RE and VE. The black bars represent the standard deviation for each participant and conditions. . . . . 86

5.5 a) Average duration of fixations and b) average amplitude of saccades for RE and VE. . . . . 87

5.6 a) $Peak_{Fixation}$ and b) Coverage for RE and VE. . . . . 87

5.7 Heat-map (log-transformed) of the gaze fixation distribution for both RE and VE . . . . . 88
5.8 a) Average duration of fixations and b) average participant's amplitude of saccades depending on crowd density. . . . . 91
5.9 a) $Peak_{Fixation}$ and b) Coverage depending on crowd density. . . . 92
5.10 Virtual street with all the different crowd densities. For each density, the fixation area (log-transformed) is displayed on top of the image . . . . . 92

6.1 Our objective is to understand whether and to what extent providing haptic rendering of collisions during navigation through a virtual crowd (right) makes users behave more realistically. Whenever a collision occurs (center), armbands worn on the arms locally vibrate to render this contact (left). We carried out an experiment with 23 participants, testing both subjective and objective metrics regarding the user's path planning, body motion, kinetic energy, presence, and embodiment. . . . . 99
6.2 Devices worn by participants during the experiment. . . . . 102
6.3 Wearable vibrotactile armband, composed of four vibrating motors (A). The electronics is enclosed in a 3D-printed case (B) [Scheggi et al., 2016]. . . . . 103
6.4 Male (a) and female (b) avatars used to represent the participants in the virtual environment. For both avatars the capsule around each segment represents the solid used to compute collisions. . . 104
6.5 Snapshots of the environment under two different points of view. Participants started from the blue cross on the floor, and were instructed to reach the screen board. Figure (b) displays an example trajectory in a red doted line. The screen displayed the train information only when participants reached the green area. . . . . 105
6.6 Illustration of a participant trajectories in a crowd, and the decomposition of the environment in cells using Delaunay triangulation [Chew, 1989]. . . . . 108
6.7 Shoulder rotation. Angle $\alpha_{SA} \in [0, 90]^\circ$ is defined between the participants' shoulder-to-shoulder axis and the segment connecting the two virtual characters. Left: top view of the scene. Right: diagram with the Delaunay triangles, the virtual characters, and the participant. . . . . 109

6.8 Collision iteration loop scheme representing one step of the collision offline detection, in which we detect if there is a collision (either a new or an ongoing one) and compute its volume. We add this information to collision's data. When the collision is finished we send out the data. . . . . 111

6.9 Volume computation using iteration of voxel spaces of decreasing dimensions. (a) Starting from the AABB around the selected geometries, the first voxel space with 8 voxels (green cubes) is created and intersected with the geometries. (b) In the next iteration only the intersecting voxels are kept, and further subdivided in 8 cubes each. (c) The process is iteratively applied until reaching the minimum subdivision size, where the final interpenetration volume is displayed in purple. . . . . 112

6.10 Participants' trajectories and Delaunay triangulation for trial $T_6$ for blocks *NoHaptic1* (left), *Haptic* (middle) and *NoHaptic2* (right). The color-bar represents the number of times participants walked on a triangle. . . . . 114

6.11 Main significant differences between the three blocks of the experiment (*NoHaptic1*, *Haptic* and *NoHaptic2*): a) amplitude of shoulder rotations ($\alpha_{SA}$), b) walking speed, c) number of collisions per trial, d) volume of interpenetration. Error bars depict standard deviation of the mean. . . . . 116

7.1 Top-Left: Plot of the compromise between **Virtual Reality Setup**, **Graphic Quality**, **Crowd Simulator** and **Perceptual System** in order to design an VR experiment, with a scale between low and high computational costs. Illustration of the compromises made for our first (Top-Right), second (Bottom-left) and third (Bottom-Right) experiments . . . . . 129

7.2 Illustration of crowd movement base cases from Duives *et al.* 's work [Duives et al., 2013] . . . . . 132

7.3 Pictures of the movie"Ready Player On" where the protagonist is completely immersed in RV . . . . . 133

7.4 A gauche : une scène de bataille de la série Game of Thrones utilisant le simulateur de foule de Goalem. A droite : le simulateur de foule MassMotion développé par Oasys et utilisé pour tester la capacité des bureaux d'Arup à Liverpool à respecter les mesures d'éloignement physique pendant le Covid. . . . . 168

7.5 La boucle de perception et d'action proposée par Gibson [1958]. . 169

# List of Tables

4.1 Summary of the conditions presented in this experiment . . . . . 61
4.2 $mpd_{CA}$, average (±SD) speed of participants, average (±SD) number of collisions, and average (±SD) number of inversions of crossing order with respect to the experimental conditions. . . . . 66

5.1 $\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$ of the gaussian distribution for the gaze location in the image and $R^2$ between this distribution and the initial distribution with respect to the experimental conditions. . . . . 89
5.2 $\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$ of the gaussian distribution for the gaze location in the image and $R^2$ between this distribution and the initial distribution with respect to the density conditions. . . . . 93

6.1 Similarity measure (Dice) of participant trajectories between all blocks (*NoHaptic1*,*Haptic*,*NoHaptic2*) for all the trials. . . . . 115
6.2 *Agency* questionnaire: average participant ratings for the three blocks. . . . . 117
6.3 *Change* questionnaire: average participant ratings for the three blocks. . . . . 117
6.4 *Ownership* questionnaire: average participant ratings for the three blocks. . . . . 118
6.5 Slater-Usoh-Steed (SUS) questionnaire [Usoh et al., 2000] and average participant ratings for the three blocks. . . . . 118

# Bibliography

P. Agethen, V. S. Sekar, F. Gaisbauer, T. Pfeiffer, M. Otto, and E. Rukzio. Behavior analysis of human locomotion in the real world and virtual reality for the manufacturing industry. *ACM Transactions on Applied Perception (TAP)*, 15(3):20, 2018.

M. Aggravi, F. Pausé, P. R. Giordano, and C. Pacchierotti. Design and evaluation of a wearable haptic device for skin stretch, pressure, and vibrotactile stimuli. *IEEE Robotics and Automation Letters*, 3(3):2166–2173, 2018.

A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016.

J. Amirian, W. van Toll, J.-B. Hayet, and J. Pettré. Data-driven crowd simulation with generative adversarial networks. In *Proc. Int. Conf. Computer Animation and Social Agents*, pages 7–10, 2019.

J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.

S. Baron-Cohen. *Mindblindness: An essay on autism and theory of mind*. MIT press, 1997.

P. Basili, M. Sağlam, T. Kruse, M. Huber, A. Kirsch, and S. Glasauer. Strategies of locomotor collision avoidance. *Gait & posture*, 37(3):385–390, 2013.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

C. C. Berger, M. Gonzalez-Franco, E. Ofek, and K. Hinckley. The uncanny valley of haptics. *Science Robotics*, 3(17):Art–No, 2018.

A. Berthoz. *The brain's sense of movement*, volume 10. Harvard University Press, 2000.

F. Berton, A.-H. Olivier, J. Bruneau, L. Hoyet, and J. Pettré. Studying gaze behaviour during collision avoidance with a virtual walker: Influence of the virtual reality setup. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 717–725. IEEE, 2019.

F. Berton, L. Hoyet, A.-H. Olivier, J. Bruneau, O. Le Meur, and J. Pettré. Eye-gaze activity in crowds: Impact of virtual reality and density. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 322–331. IEEE, 2020.

J. Bimbo, C. Pacchierotti, M. Aggravi, N. Tsagarakis, and D. Prattichizzo. Teleoperation in cluttered environments using wearable haptic feedback. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3401–3408, 2017.

E. Birmingham and A. Kingstone. Human social attention. *Progress in brain research*, 176:309–320, 2009.

J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological inquiry*, 13(2):103–124, 2002.

R. W. Bohannon. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. *Age and ageing*, 26(1):15–19, 1997.

M. Boltes and A. Seyfried. Collecting pedestrian trajectories. *Neurocomputing*, 100:127–133, 2013.

A. Bönsch, S. Radke, H. Overath, L. M. Asché, J. Wendt, T. Vierjahn, U. Habel, and T. W. Kuhlen. Social vr: How personal space is affected by virtual agents' emotions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 199–206. IEEE, 2018.

S. M. Bourgaize, B. J. McFadyen, and M. E. Cinelli. Collision avoidance behaviours when circumventing people of different sizes in various positions and locations. *Journal of Motor Behavior*, pages 1–10, 2020.

A. T. Brady and M. B. Walker. Interpersonal distance as a function of situationally induced anxiety. *British Journal of Social and Clinical Psychology*, 17(2):127–133, 1978.

J. Bruneau and J. Pettré. Eacs: Effective Avoidance Combination Strategy. *Computer Graphics Forum*, 36(8):108–122, 2017.

J. Bruneau, A.-H. Olivier, and J. Pettré. Going through, going around: A study on individual avoidance of groups. *IEEE transactions on visualization and computer graphics*, 21(4):520–528, 2015.

M. A. Bühler and A. Lamontagne. Circumvention of pedestrians while walking in virtual and physical environments. *IEEE transactions on neural systems and rehabilitation engineering*, 26(9):1813–1822, 2018.

S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool. One-shot video object segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.

A. J. Calder, A. D. Lawrence, J. Keane, S. K. Scott, A. M. Owen, I. Christoffels, and A. W. Young. Reading the mind from eye gaze. *Neuropsychologia*, 40(8): 1129–1138, 2002.

F. Capozzi, A. P. Bayliss, and J. Ristic. Gaze following in multiagent contexts: Evidence for a quorum-like principle. *Psychonomic bulletin & review*, 25(6): 2260–2266, 2018.

R. Challenger, C. Clegg, M. Robinson, and M. Leigh. Understanding crowd behaviours. *Supporting Theory and Evidence*, 2, 2009.

P. Charalambous and Y. Chrysanthou. The pag crowd: A graph based approach for efficient data-driven crowd simulation. In *Computer Graphics Forum*, volume 33, pages 95–108. Wiley Online Library, 2014.

U. Chattaraj, A. Seyfried, and P. Chakroborty. Comparison of pedestrian fundamental diagram across cultures. *Advances in complex systems*, 12(03):393–405, 2009.

S. Chenney. Flow tiles. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 233–242, 2004.

L. P. Chew. Constrained delaunay triangulations. *Algorithmica*, 4(1-4):97–108, 1989.

F. Chinello, C. Pacchierotti, M. Malvezzi, and D. Prattichizzo. A three revolute-revolute-spherical wearable fingertip cutaneous device for stiffness rendering. *IEEE Transactions on Haptics*, 11(1):39–50, 2017.

S. Cho, J. Park, and O. Kwon. Gender differences in three dimensional gait analysis data from 98 healthy korean adults. *Clinical biomechanics*, 19(2):145–152, 2004.

Y.-h. Chou, R. C. Wagenaar, E. Saltzman, J. E. Giphart, D. Young, R. Davidsdottir, and A. Cronin-Golomb. Effects of optic flow speed and lateral flow asymmetry on locomotion in younger and older adults: a virtual reality study. *Journals of Gerontology: Series B*, 64(2):222–231, 2009.

M. E. Cinelli, A. E. Patla, and F. Allard. Behaviour and gaze analyses during a goal-directed locomotor task. *The Quarterly Journal of Experimental Psychology*, 62(3):483–499, 2009.

G. Cirio, A.-H. Olivier, M. Marchal, and J. Pettré. Kinematic evaluation of virtual walking trajectories. *IEEE transactions on visualization and computer graphics*, 19(4):671–680, 2013.

R. Collier. An experimental study of form perception in indirect vision. *Journal of Comparative Psychology*, 11(3):281, 1931.

M. Costa. Interpersonal distances in group walking. *Journal of Nonverbal Behavior*, 34(1):15–26, 2010.

A. Coutrot, N. Guyader, G. Ionescu, and A. Caplier. Influence of soundtrack on eye movements during video exploration. 2012.

J. L. Croft and D. Panchuk. Watch where you're going? interferer velocity and visual behavior predicts avoidance strategy during pedestrian encounters. *Journal of Motor Behavior*, 0(0):1–11, 2017.

J. E. Cutting, P. M. Vishton, and P. A. Braren. How we avoid collisions with stationary and moving objects. *Psychological review*, 102(4):627, 1995.

G. C. Dachner and W. H. Warren. Behavioral dynamics of heading alignment in pedestrian following. *Transportation Research Procedia*, 2:69–76, 2014.

L. Devigne, M. Aggravi, M. Bivaud, N. Balix, S. Teodorescu, T. Carlson, T. Spreters, C. Pacchierotti, and M. Babel. Power wheelchair navigation assistance using wearable vibrotactile haptics. *IEEE Transactions on Haptics*, 2020.

P. Dickinson, K. Gerling, K. Hicks, J. Murray, J. Shearer, and J. Greenwood. Virtual reality crowd simulation: effects of agent density on user experience and behaviour. *Virtual Reality*, 23(1):19–32, 2019.

M. Dicks, C. Button, and K. Davids. Examination of gaze behaviors under in situ and video simulation task constraints reveals differences in information pickup for perception and action. *Attention, Perception, & Psychophysics*, 72 (3):706–720, 2010.

M. Dicks, C. Clashing, L. O'Reilly, and C. Mills. Perceptual-motor behaviour during a simulated pedestrian crossing. *Gait & posture*, 49:241–245, 2016.

H. Dietrich and M. Wuehr. Strategies for gaze stabilization critically depend on locomotor speed. *Neuroscience*, 408:418–429, 2019.

M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth. Variability of eye movements when viewing dynamic natural scenes. *Journal of vision*, 10(10): 28–28, 2010.

D. C. Duives, W. Daamen, and S. P. Hoogendoorn. State-of-the-art crowd motion simulation models. *Transportation research part C: emerging technologies*, 37: 193–209, 2013.

T. B. Dutra, R. Marques, J. B. Cavalcante-Neto, C. A. Vidal, and J. Pettré. Gradient-based steering for vision-based crowd simulation algorithms. *Comput. Graph. Forum*, 36(2):337–348, 2017.

T. Duverne, T. Rougnant, F. Le Yondre, F. Berton, J. Bruneau, K. Zibrek, J. Pettré, L. Hoyet, and A.-H. Olivier. Effect of social settings on proxemics during social interactions in real and virtual conditions. In *International Conference on Virtual Reality and Augmented Reality*. Springer, 2020.

C. Faure. Vers des environnements virtuels plus écologiques: étude des modifications du comportement moteur en réalité virtuelle lors de l'ajout d'informations haptiques par un mécanisme parallèle entraîné par câbles. 2019.

J. M. Findlay, J. M. Findlay, I. D. Gilchrist, et al. *Active vision: The psychology of looking and seeing*. Number 37. Oxford University Press, 2003.

P. W. Fink, P. S. Foo, and W. H. Warren. Obstacle avoidance during walking in real and virtual environments. *ACM Transactions on Applied Perception (TAP)*, 4(1):2, 2007.

P. Fiorini and Z. Shiller. Motion planning in dynamic environments using velocity obstacles. *The International Journal of Robotics Research*, 17(7):760–772, 1998.

A. Frischen, A. P. Bayliss, and S. P. Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4):694, 2007.

K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny. *Statistical Parametric Mapping*. Academic Press, 2007. ISBN 9780123725608.

P. Fuchs. *Le traité de la réalité virtuelle*, volume 2. Presses des MINES, 2006.

A. C. Gallup, A. Chong, and I. D. Couzin. The directional flow of visual information transfer between pedestrians. *Biology letters*, 8(4):520–522, 2012.

Y. Gan. *Continuum Mechanics: Progress in Fundamentals and Engineering Applications*. BoD–Books on Demand, 2012.

M. Gérin-Lajoie, C. L. Richards, and B. J. McFadyen. The negotiation of stationary and moving obstructions during walking: anticipatory locomotor adaptations and preservation of personal space. *Motor control*, 9(3):242–269, 2005.

M. Gérin-Lajoie, C. L. Richards, J. Fung, and B. J. McFadyen. Characteristics of personal space during obstacle circumvention in physical and virtual environments. *Gait & posture*, 27(2):239–247, 2008.

J. J. Gibson. Visually controlled locomotion and visual orientation in animals. *British journal of psychology*, 49(3):182–194, 1958.

J. Godoy, I. Karamouzas, S. J. Guy, and M. Gini. Anytime navigation with progressive hindsight optimization. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 730–735. IEEE, 2014.

A. Golas, R. Narain, S. Curtis, and M. C. Lin. Hybrid long-range collision avoidance for crowd simulation. *IEEE transactions on visualization and computer graphics*, 20(7):1022–1034, 2013.

A. Gorrini, S. Bandini, M. Sarvi, C. Dias, and N. Shiwakoti. An empirical study of crowd and pedestrian dynamics: the impact of different angle paths and grouping. *Transportation Research Record*, 41(42):10, 2013.

A. Gorrini, S. Bandini, and G. Vizzari. Empirical investigation on pedestrian crowd dynamics and grouping. In *Traffic and Granular Flow'13*, pages 83–91. Springer, 2015.

A. Gorrini, L. Crociani, C. Feliciani, P. Zhao, K. Nishinari, and S. Bandini. Social groups and pedestrian crowds: experiment on dyads in a counter flow scenario. *arXiv preprint arXiv:1610.08325*, 2016.

R. Grundberg, M. Cinelli, S. Bourgaize, A. Crétual, and A. Olivier. Collision avoidance in older adults.

D. Guitton and M. Volle. Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of neurophysiology*, 58(3):427–459, 1987.

A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2255–2264, 2018.

A. L. Hackney, L. A. Vallis, and M. E. Cinelli. Action strategies of individuals during aperture crossing in nonconfined space. *Quarterly Journal of Experimental Psychology*, 66(6):1104–1112, 2013.

A. L. Hackney, M. E. Cinelli, and J. S. Frank. Does the passability of apertures change when walking through human versus pole obstacles? *Acta psychologica*, 162:62–68, 2015.

E. T. Hall. A system for the notation of proxemic behavior. *American anthropologist*, 65(5):1003–1026, 1963.

E. T. Hall. *The hidden dimension*, volume 609. Garden City, NY: Doubleday, 1966.

D. Harley, A. Verni, M. Willis, A. Ng, L. Bozzo, and A. Mazalek. Sensory vr: Smelling, touching, and eating virtual reality. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 386–397, 2018.

M. L. Heilig. Sensorama simulator, Aug. 28 1962. US Patent 3,050,870.

D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51(5):4282–4286, 1995.

R. S. Hessels, D. C. Niehorster, M. Nyström, R. Andersson, and I. T. Hooge. Is the eye-movement field confused about fixations and saccades? a survey among 124 researchers. *Royal Society open science*, 5(8):180502, 2018.

R. S. Hessels, A. J. van Doorn, J. S. Benjamins, G. A. Holleman, and I. T. Hooge. Task-related gaze control in human crowd navigation. *Attention, Perception, & Psychophysics*, pages 1–20, 2020.

S. P. Hoogendoorn and P. H. Bovy. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, 38(2):169–190, 2004.

P.-H. Huang and S.-K. Wong. Emotional virtual crowd on task completion in virtual markets. *Computer Animation and Virtual Worlds*, 29(3-4):e1818, 2018.

M. Huber, Y.-H. Su, M. Krüger, K. Faschian, S. Glasauer, and J. Hermsdörfer. Adjustments of speed and path when avoiding collisions with another pedestrian. *PloS one*, 9(2), 2014.

R. L. Hughes. A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodological*, 36(6):507–535, 2002.

R. L. Hughes. The flow of human crowds. *Annual review of fluid mechanics*, 35 (1):169–182, 2003.

R. J. Itier and M. Batty. Neural bases of eye and gaze processing: the core of social cognition. *Neuroscience & Biobehavioral Reviews*, 33(6):843–863, 2009.

X. Jin, J. Xu, C. C. Wang, S. Huang, and J. Zhang. Interactive control of large-crowd navigation in virtual environments using vector fields. *IEEE Computer Graphics and Applications*, 28(6):37–46, 2008.

J. Jovancevic, B. Sullivan, and M. Hayhoe. Control of attention and gaze in complex environments. *Journal of Vision*, 6(12):9–9, 2006.

J. Jovancevic-Misic and M. Hayhoe. Adaptive gaze control in natural environments. *Journal of Neuroscience*, 29(19):6234–6238, 2009.

A. Kar and P. Corcoran. A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access*, 5:16495–16519, 2017.

I. Karamouzas, B. Skinner, and S. J. Guy. Universal power law governing pedestrian interactions. *Physical review letters*, 113(23):238701, 2014.

W. Kerr and D. Spears. Robotic simulation of gases for a surveillance task. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2905–2910. IEEE, 2005.

K. Kitazawa and T. Fujiyama. Pedestrian vision and collision avoidance behavior: Investigation of the information process space of pedestrians using an eye tracker. In *Pedestrian and evacuation dynamics 2008*, pages 95–108. Springer, 2010.

C. Kmiecik, P. Mausar, and G. Banziger. Attractiveness and interpersonal space. *The Journal of Social Psychology*, 108(2):277–278, 1979.

A. Knorr, L. Willacker, J. Hermsdörfer, S. Glasauer, and M. Krüger. Influence of Person- and Situation-Specific Characteristics on Collision Avoidance Behavior in Human Locomotion. *J. of Exp. Psy.: Human Perception and Performance*, 42(9):1332–1343, 2016.

A. Koilias, M. G. Nelson, C.-N. Anagnostopoulos, and C. Mousas. Immersive walking in a virtual crowd: The effects of the density, speed, and direction of a virtual crowd on human movement behavior. *Computer Animation and Virtual Worlds*, page e1928, 2020.

O. V. Komogortsev and A. Karpov. Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior research methods*, 45(1):203–215, 2013.

C. Krogmeier, C. Mousas, and D. Whittinghill. Human, virtual human, bump! a preliminary study on haptic feedback. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1032–1033. IEEE, 2019.

D. M. Krum, S.-H. Kang, and T. Phan. Influences on the elicitation of interpersonal space with virtual humans. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 223–9. IEEE, 2018.

K. E. Laidlaw, T. Foulsham, G. Kuhn, and A. Kingstone. Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, 108(14):5548–5553, 2011.

S. R. Langton, A. S. Law, A. M. Burton, and S. R. Schweinberger. Attention capture by faces. *Cognition*, 107(1):330–342, 2008.

K. H. Lee, M. G. Choi, Q. Hong, and J. Lee. Group behavior from video: A data-driven approach to crowd simulation. In *Proc. ACM SIGGRAPH/Eurographics Symp. Computer Animation*, pages 109–118, 2007.

R. J. Leigh and D. S. Zee. *The neurology of eye movements*. OUP USA, 2015.

D. Leonardis, M. Solazzi, I. Bortone, and A. Frisoli. A 3-rsr haptic wearable device for rendering fingertip contact forces. *IEEE transactions on haptics*, 10 (3):305–316, 2016.

A. Lerner, Y. Chrysanthou, and D. Lischinski. Crowds by example. *Comput. Graph. Forum*, 26(3):655–664, 2007.

S. Licence, R. Smith, M. P. McGuigan, and C. P. Earnest. Gait pattern alterations during walking, texting and walking and texting during cognitively distractive tasks while negotiating common pedestrian obstacles. *PLoS one*, 10 (7):e0133281, 2015.

T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

J. Llobera, B. Spanlang, G. Ruffini, and M. Slater. Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception (TAP)*, 8(1):1–12, 2010.

J. M. Loomis, J. J. Blascovich, and A. C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods, instruments, & computers*, 31(4):557–564, 1999.

J. M. Loomis, J. M. Knapp, et al. Visual perception of egocentric distance in real and virtual environments. *Virtual and adaptive environments*, 11:21–46, 2003.

C. Louison, F. Ferlay, and D. R. Mestre. Spatialized vibrotactile feedback improves goal-directed movements in cluttered virtual environments. *International Journal of Human–Computer Interaction*, 34(11):1015–1031, 2018.

S. Lynch, J. Pettré, J. Bruneau, R. Kulpa, A. Cretual, and A.-H. Olivier. Effect of virtual human gaze behaviour during an orthogonal collision avoidance walking task. In *IEEE Virtual Reality*, 2018.

S. D. Lynch, R. Kulpa, L. A. Meerhoff, J. Pettre, A. Cretual, and A. H. Olivier. Collision avoidance behavior between walkers: global and local motion cues. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2017. ISSN 1077-2626. doi: 10.1109/TVCG.2017.2718514.

A. López, F. Chaumette, E. Marchand, and J. Pettré. Character navigation in dynamic environments based on optical flow. *Comput. Graph. Forum*, 38(2): 181–192, 2019.

D. S. Marigold. Role of peripheral visual cues in online visual guidance of locomotion. *Exercise and sport sciences reviews*, 36(3):145–151, 2008.

D. S. Marigold and A. E. Patla. Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience*, 144(1):302–313, 2007.

M. F. Mason, E. P. Tatkow, and C. N. Macrae. The look of love: Gaze shifts and person perception. *Psychological science*, 16(3):236–239, 2005.

L. Meerhoff, J. Bruneau, A. Vu, A.-H. Olivier, and J. Pettré. Guided by gaze: Prioritization strategy when navigating through a virtual crowd can be assessed through gaze activity. *Acta psychologica*, 190:248–257, 2018a.

L. A. Meerhoff, J. Pettré, S. D. Lynch, A. Crétual, and A.-H. Olivier. Collision avoidance with multiple walkers: Sequential or simultaneous interactions? *Frontiers in psychology*, 9:2354, 2018b.

L. Meli, C. Pacchierotti, G. Salvietti, F. Chinello, M. Maisto, A. De Luca, and D. Prattichizzo. Combining wearable finger haptics and augmented reality: User evaluation using an external camera and the microsoft hololens. *IEEE Robotics and Automation Letters*, 3(4):4297–4304, 2018.

D. R. Mestre, C. Louison, and F. Ferlay. The contribution of a virtual self and vibrotactile feedback to walking through virtual apertures. In *International Conference on Human-Computer Interaction*, pages 222–232. Springer, 2016.

C. Mousas, A. Koilias, D. Anastasiou, B. Hekabdar, and C.-N. Anagnostopoulos. Effects of self-avatar and gaze on avoidance movement behavior. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 726–734. IEEE, 2019.

M. Moussaïd, D. Helbing, and G. Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888, 2011a.

M. Moussaïd, D. Helbing, and G. Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888, 2011b.

R. Narain, A. Golas, S. Curtis, and M. C. Lin. Aggregate dynamics for dense crowd simulation. *ACM Trans. Graph.*, 28(5):122, 2009.

L. Nummenmaa and A. J. Calder. Neural mechanisms of social attention. *Trends in cognitive sciences*, 13(3):135–143, 2009.

L. Nummenmaa, J. Hyönä, and J. K. Hietanen. I'll walk this way: Eyes reveal the direction of locomotion and make passersby look and go the other way. *Psychological Science*, 20(12):1454–1458, 2009.

A.-H. Olivier, A. Marin, A. Crétual, and J. Pettré. Minimal predicted distance: A common metric for collision avoidance during pairwise interactions between walkers. *Gait and Posture*, 36(3):399–404, 2012. ISSN 09666362.

A.-H. Olivier, A. Marin, A. Crétual, A. Berthoz, and J. Pettré. Collision avoidance between two walkers: Role-dependent strategies. *Gait & Posture*, 38(4):751–756, 2013. ISSN 09666362.

A. H. Olivier, J. Bruneau, R. Kulpa, and J. Pettré. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2017. ISSN 1077-2626. doi: 10.1109/TVCG.2017.2714665.

C. Pacchierotti, S. Sinclair, M. Solazzi, A. Frisoli, V. Hayward, and D. Prattichizzo. Wearable haptic systems for the fingertip and the hand: taxonomy, review, and perspectives. *IEEE Transactions on Haptics*, 10(4):580–600, 2017.

R. Parasuraman and M. Rizzo. *Neuroergonomics: The brain at work*. Oxford University Press, 2008.

S. Paris, J. Pettré, and S. Donikian. Pedestrian reactive navigation for crowd simulation: a predictive approach. *Comput. Graph. Forum*, 26:665–674, 2007.

S. Patil, J. Berg, S. Curtis, M. C. Lin, and D. Manocha. Directing crowd simulations using navigation fields. *IEEE Trans. Vis. Comput. Graphics*, 17:244–254, 2010.

A. E. Patla. Understanding the roles of vision in the control of human locomotion. *Gait & Posture*, 5(1):54–69, 1997.

S. Pellegrini, J. Gall, L. Sigal, and L. van Gool. Destination flow for crowd simulation. In *Proc. European Conf. Computer Vision*, pages 162–171, 2012.

J. Pettré, J. Ondřej, A.-H. Olivier, A. Cretual, and S. Donikian. Experiment-based modeling, simulation and validation of interactions between virtual walkers. In

*Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 189–198, 2009.

K. Pfeil, E. M. Taranta II, A. Kulshreshth, P. Wisniewski, and J. J. LaViola Jr. A comparison of eye-head coordination between virtual and physical realities. In *Proceedings of the 15th ACM Symposium on Applied Perception*, page 18. ACM, 2018.

L. C. Pimenta, N. Michael, R. C. Mesquita, G. A. Pereira, and V. Kumar. Control of swarms based on hydrodynamic models. In *2008 IEEE international conference on robotics and automation*, pages 1948–1953. IEEE, 2008.

J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker. Distance perception in real and virtual environments. *ACM Transactions on Applied Perception (TAP)*, 2(3):216–233, 2005.

E. Pöppel and L. O. Harvey. Light-difference threshold and subjective brightness in the periphery of the visual field. *Psychologische Forschung*, 36(2):145–161, 1973.

V. Rapos, M. Cinelli, N. Snyder, A. Crétual, and A.-H. Olivier. Minimum predicted distance: Applying a common metric to collision avoidance strategies between children and adult walkers. *Gait & posture*, 72:16–21, 2019.

J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

R. Reed-Jones, J. Reed-Jones, L. A. Vallis, and M. Hollands. The effects of constraining eye movements on visually evoked steering responses during walking in a virtual environment. *Experimental brain research*, 197(4):357–367, 2009.

M. S. Remland, T. S. Jones, and H. Brinkman. Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age. *The Journal of social psychology*, 135(3):281–297, 1995.

C. W. Reynolds. Steering behaviors for autonomous characters. In *Game developers conference*, volume 1999, pages 763–782. Citeseer, 1999.

K. Rio and W. H. Warren. The visual coupling between neighbors in real and virtual crowds. *Transportation Research Procedia*, 2:132–140, 2014.

K. W. Rio, C. K. Rhea, and W. H. Warren. Follow the leader: Visual control of speed in pedestrian following. *Journal of vision*, 14(2):4–4, 2014.

K. W. Rio, G. C. Dachner, and W. H. Warren. Local interactions underlying collective motion in human crowds. *Proceedings of the Royal Society B: Biological Sciences*, 285(1878):20180611, 2018.

A. Ríos and N. Pelechano. Follower behavior under stress in immersive vr. *Virtual Reality*, pages 1–12, 2020.

A. Ríos, M. Palomar, and N. Pelechano. Users' locomotor behavior in collaborative virtual reality. In *Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games*, page 15. ACM, 2018.

D. Roth and M. E. Latoschik. Construction of the virtual embodiment questionnaire (veq). *IEEE Transactions on Visualization and Computer Graphics*, 2020.

M. Rubo and M. Gamer. Virtual reality as a proxy for real-life social attention? In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, page 81. ACM, 2018.

M. R. Saeedpour-Parizi, S. E. Hassan, T. Baniasadi, K. J. Baute, and J. B. Shea. Hierarchical goal effects on center of mass velocity and eye fixations during gait. *Experimental Brain Research*, pages 1–11, 2020.

D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.

F. A. Sanz, A.-H. Olivier, G. Bruder, J. Pettré, and A. Lécuyer. Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments. In *2015 IEEE Virtual Reality (VR)*, pages 75–80. IEEE, 2015.

J. Sardo, J. Semião, J. Monteiro, J. A. Pereira, M. A. de Freitas, E. Esteves, and J. M. Rodrigues. Portable device for touch, taste and smell sensations in augmented reality experiences. In *International Congress on Engineering and Sustainability in the XXI Century*, pages 305–320. Springer, 2017.

S. Scheggi, M. Aggravi, and D. Prattichizzo. Cooperative navigation for mixed human–robot teams using haptic feedback. *IEEE Transactions on Human-Machine Systems*, 47(4):462–473, 2016.

S. B. Schorr and A. M. Okamura. Three-dimensional skin deformation as force substitution: Wearable device design and performance during haptic explo-

ration of virtual environments. *IEEE transactions on haptics*, 10(3):418–430, 2017.

A. Seyfried, B. Steffen, W. Klingsch, and M. Boltes. The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10):P10002, 2005a.

A. Seyfried, B. Steffen, W. Klingsch, and M. Boltes. The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10):P10002, 2005b.

M. Shimizu, A. Ishiguro, T. Kawakatsu, Y. Masubuchi, and M. Doi. Coherent swarming from local interaction by exploiting molecular dynamics and stokesian dynamics methods. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, volume 2, pages 1614–1619. IEEE, 2003.

W. S. Silva, G. Aravind, S. Sangani, and A. Lamontagne. Healthy young adults implement distinctive avoidance strategies while walking and circumventing virtual human vs. non-human obstacles in a virtual environment. *Gait & posture*, 61:294–300, 2018.

M. J. Simpson. Mini-review: far peripheral vision. *Vision research*, 140:96–105, 2017.

M. Slater. A note on presence terminology. *Presence connect*, 3(3):1–5, 2003.

M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

M. Slater, P. Khanna, J. Mortensen, and I. Yu. Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications*, 29(3):76–84, 2009.

J. Snape, S. J. Guy, D. Vembar, A. Lake, M. C. Lin, and D. Manocha. Reciprocal collision avoidance and navigation for video games. In *Game Developers Conf., San Francisco*, 2012.

T. Sørensen, T. Sørensen, T. Sørensen, T. SORENSEN, T. Sorensen, T. Sorensen, and T. Biering-Sørensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. 1948.

R. H. Spector. Visual fields. *Clinical methods: The history, physical, and laboratory examinations*, 1990.

M. Startsev, I. Agtzidis, and M. Dorr. 1d cnn with blstm for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods*, 51(2):556–572, 2019.

T. A. Stoffregen, M. A. Schmuckler, and E. J. Gibson. Use of central and peripheral optical flow in stance and locomotion in young walkers. *Perception*, 16(1): 113–119, 1987.

H. Strasburger. Seven myths on crowding and peripheral vision. *i-Perception*, 11 (3):2041669520913052, 2020.

D. J. Sumpter, R. P. Mann, and A. Perna. The modelling cycle for collective animal behaviour. *Interface focus*, 2(6):764–773, 2012.

M. J. Tarr and W. H. Warren. Virtual reality in behavioral neuroscience and beyond. *nature neuroscience*, 5(11):1089–1092, 2002.

N. D. Thomas, J. D. Gardiner, R. H. Crompton, and R. Lawson. Look out: an exploratory study assessing how gaze (eye angle and head angle) and gait speed are influenced by surface complexity. *PeerJ*, 8:e8838, 2020.

A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014.

A. Treuille, S. Cooper, and Z. Popović. Continuum crowds. *ACM Transactions on Graphics (TOG)*, 25(3):1160–1168, 2006.

K. A. Turano, D. Yu, L. Hao, and J. C. Hicks. Optic-flow and egocentric-direction strategies in walking: Central vs peripheral visual field. *Vision research*, 45(25-26):3117–3132, 2005.

M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503, 2000.

J. Van Den Berg, S. Patil, J. Sewall, D. Manocha, and M. Lin. Interactive navigation of multiple agents in crowded environments. In *Symp. on Interactive 3D graphics and games*, pages 139–147, 2008.

J. P. van den Berg, M. C. Lin, and D. Manocha. Reciprocal Velocity Obstacles for real-time multi-agent navigation. In *Proc. IEEE Int. Conf. Robotics and Automation*, pages 1928–1935, 2008.

J. P. van den Berg, S. J. Guy, M. C. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Proc. Int. Symp. Robotics Research*, pages 3–19, 2011.

W. Van Toll, R. Triesscheijn, M. Kallmann, R. Oliva, N. Pelechano, J. Pettré, and R. Geraerts. A comparative study of navigation meshes. In *Proceedings of the 9th International Conference on Motion in Games*, pages 91–100, 2016.

W. van Toll, F. Grzeskowiak, A. L. Gandía, J. Amirian, F. Berton, J. Bruneau, B. C. Daniel, A. Jovane, and J. Pettré. Generalized microscropic crowd simulation using costs in velocity space. In *Symposium on Interactive 3D Graphics and Games*, pages 1–9, 2020.

K. Varma, S. J. Guy, and V. Interrante. Assessing the relevance of eye gaze patterns during collision avoidance in virtual reality. 2017.

J. N. Vickers. *Perception, cognition, and decision training: The quiet eye in action*. Human Kinetics, 2007.

M. Volonte, Y. C. Hsu, K.-y. Liu, J. Mazer, S.-K. Wong, and S. Babu. Effects of interacting with a crowd of emotional virtual humans on users' affective and non-verbal behaviors. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020.

N. Wang, S. Xu, S. Zhang, Y. Luo, and H. Geng. Erp evidence on how gaze convergence affects social attention. *Scientific reports*, 9(1):7586, 2019.

W. Warren. Visually controlled locomotion: 40 years later. *Ecological Psychology*, 10(3-4):177–219, 1998.

W. H. Warren. The dynamics of perception and action. *Psychological review*, 113 (2):358, 2006.

W. H. Warren and B. R. Fajen. Behavioral dynamics of visually guided locomotion. In *Coordination: neural, behavioral and social dynamics*, pages 45–75. Springer, 2008.

W. H. Warren Jr, B. A. Kay, W. D. Zosh, A. P. Duchon, and S. Sahuc. Optic flow is used to control human walking. *Nature neuroscience*, 4(2):213, 2001.

D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063, 1988.

N. Wijermans. Understanding crowd behaviour. *University of Groningen, Groningen*, 2011.

P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanics on distance judgments in virtual environments. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 35–38, 2004.

K. Wilmut and A. L. Barnett. Locomotor adjustments when navigating through apertures. *Human Movement Science*, 29(2):289–298, 2010.

N. Wojke, A. Bewley, and D. Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3645–3649. IEEE, 2017. doi: 10.1109/ICIP.2017.8296962.

D. S. Wooding. Eye movements of large populations: Ii. deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments, & Computers*, 34(4):518–528, 2002.

A. L. Yarbus. *Eye movements and vision*. Springer, 2013.

H. Yeh, S. Curtis, S. Patil, J. van den Berg, D. Manocha, and M. Lin. Composite agents. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 39–47, 2008.

B. Yersin, J. Maïm, J. Pettré, and D. Thalmann. Crowd patches: populating large-scale virtual environments for real-time applications. In *Proceedings of the 2009 symposium on Interactive 3D graphics and games*, pages 207–214, 2009.

X. Zhao, L. Xia, J. Zhang, and W. Song. Artificial neural network based modeling on unidirectional and bidirectional pedestrian flow at straight corridors. *Physica A: Statistical Mechanics and its Applications*, page 123825, 2019.

J. Zhong, W. Cai, L. Luo, and M. Zhao. Learning behavior patterns from video for agent-based crowd modeling and simulation. *Autonomous Agents and Multi-Agent Systems*, 30(5):990–1019, 2016.

# Résumé en français

## Context

Au cours de cette thèse, nous nous sommes intéressés à approfondir nos connaissances sur le comportement collectif dans le cas de foules en mouvement. Pour ce faire, nous avons travaillé sur la conception et l'évaluation d'une plateforme expérimentale en réalité virtuelle pour étudier le comportement des piétons, qui est au cœur du comportement collectif. Dans notre cas, notre recherche s'applique au domaine de la simulation de foule, qui vise à calculer le mouvement d'une foule de personnes afin de reproduire leur comportement. Ces simulations présentent un intérêt pour certains domaines d'application, tels que le divertissement (Figure 1.1-Gauche) ou la gestion de foule (Figure 1.1-Droite), chacun d'entre eux ayant ses propres critères de performance. Par exemple, les agents d'une foule virtuelle dans un jeu vidéo peuvent être amenés à interagir avec le joueur. Il est donc nécessaire de disposer d'une simulation qui permette une interactivité en temps réel, induisant ainsi des performances élevées en termes de temps de calcul. Ce n'est pas le cas pour le cinéma, où l'on peut être amené à simuler une foule de plusieurs milliers d'agents avec des mouvements très réalistes visuellement et la possibilité de retoucher ces animations. Bien entendu, les critères de performances sont implicites dans ces cas-là car elles affectent directement le budget, mais la qualité visuelle reste l'élément principal. Dans le cas d'un scénario d'évacuation dans le cadre de la planification d'un événement, ou de la construction d'une infrastructure, il est important de disposer de simulations prédictives réalistes pouvant être adaptées à toute situation, ce qui nécessite d'évaluer tous les effets possibles des paramètres sur le modèle du simulateur de foule. Cette catégorie de modèle est utilisée par exemple pour choisir les meilleurs emplacements pour les sorties ou pour évaluer divers critères de sécurité tels que le temps d'évacuation ou la distance physique minimale maintenue entre les personnes dans le contexte de Covid (Figure 1.1-Droit).

Pour satisfaire ces applications et respecter ces critères de performance, plusieurs méthodes de simulation ont été développées. Elles peuvent être divisées en

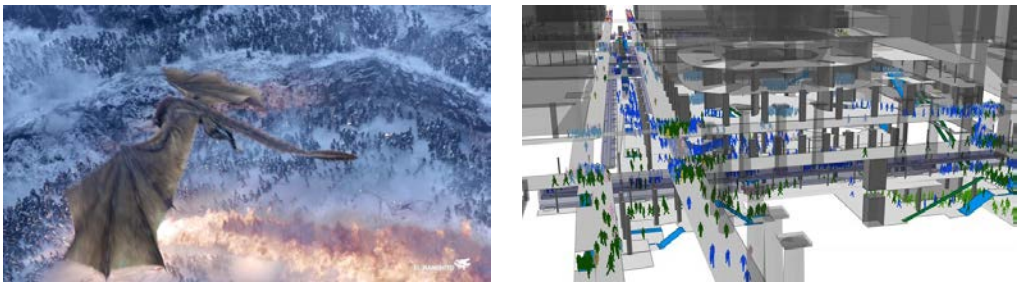


**Figure 7.4** – A gauche : une scène de bataille de la série Game of Thrones utilisant le simulateur de foule de Goalem. A droite : le simulateur de foule MassMotion développé par Oasys et utilisé pour tester la capacité des bureaux d'Arup à Liverpool à respecter les mesures d'éloignement physique pendant le Covid.

trois grandes catégories :

1. *"Flow-based"* : ces méthodes visent à simuler de très grandes foules de personnes en considérant la foule comme une seule entité.

2. *"Data-driven"* : ces méthodes visent à mimer les mouvements de foule à partir de données réelles.

3. *"Agent-based"* : ces méthodes simulent le comportement de chaque agent virtuel au sein de la foule. Le mouvement global de la foule émerge alors de la combinaison des interactions locales entre tous les agents.

Dans cette thèse, nous visons à améliorer les méthodes *"Agent-based"* en fournissant de nouvelles méthodes pour analyser le comportement des piétons localement. Les interactions locales lors de la navigation au sein d'une foule sont diverses et comprennent les tâches suivantes : suivi d'une personne, marche en groupe, atteindre quelqu'un ainsi qu'éviter une collision. Notre travail s'est spécifiquement concentré sur les tâches d'évitement des collisions avec d'autres piétons. Dans ce cadre, les chercheurs ont proposé plusieurs modèles pour simuler de tels interactions, basés sur les principes de la physique (par exemple, les forces répulsives étant donné la distance interpersonnelle entre les marcheurs) ou tenant compte des vitesses relatives des piétons. La complexité d'une cette modélisation consiste à comprendre quel agent au sein d'une foule influence le mouvement du piéton. C'est ce qu'on appelle le " voisinage d'interaction ". Il va sans dire que lorsque nous marchons dans la rue, nos mouvements et nos actions ne tiennent pas compte de toutes les personnes présentes dans la rue mais d'un sous-ensemble autour de nous. La définition d'un tel voisinage d'interaction a été peu étudiée dans la littérature, et représente un défi de taille. En effet, comprendre et modéliser ce voisinage

équivaut à inverser un processus injectif par lequel de multiples sources d'interactions se combinent pour influencer une seule trajectoire de dimension inférieure. Cependant, nous pensons que l'amélioration des modèles actuels de simulation de type *"Agent-based"* repose sur une meilleure définition de ce voisinage. Ceci requiert une analyse approfondie du comportement humain à l'échelle locale lors de la navigation dans des environnements à forte densité de population.

## Notre Approche

Cette thèse vise à améliorer les modèles actuels de simulation de foule en comprenant comment le mouvement des piétons est contrôlé. Dans ce contexte, notre approche s'appuie sur la théorie écologique de la perception visuelle développée par Gibson [1958]. Selon lui : *"Nous devons percevoir pour nous déplacer, mais nous devons aussi nous déplacer pour percevoir".* Gibson considère le système agent - environnement, où les interactions peuvent être décrites comme une boucle perception - action, illustrée Figure 1.2. Dans cette approche, la perception est considérée comme directe, ce qui signifie que l'agent a une perception immédiate des variables de haut niveau, directement disponibles dans le flux sensoriel. L'être humain va donc percevoir l'environnement à travers ses systèmes perceptifs (vision, toucher, audition...) et va ensuite initier une action (un mouvement) qui va à son tour modifier l'environnement perçu. Ce modèle a été développé par Warren [1998, 2006] qui introduit dans cette boucle perception-action la notion de loi de comportement afin de décrire les interactions entre l'environnement et l'agent.

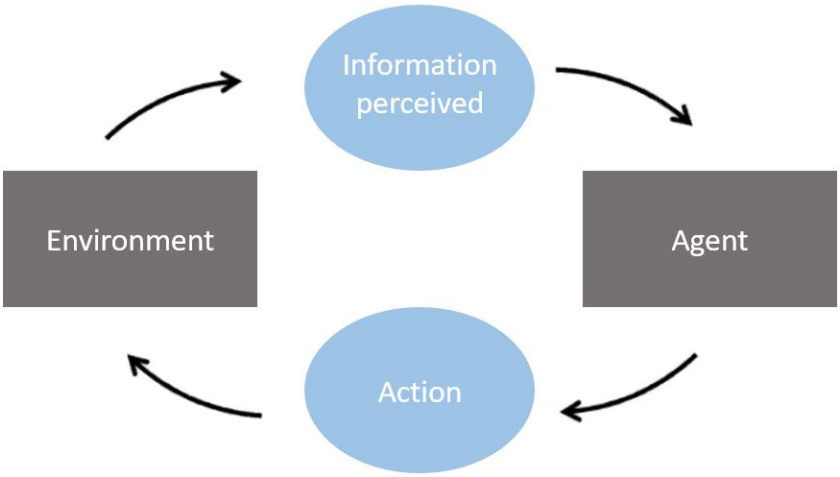



**Figure 7.5** – La boucle de perception et d'action proposée par Gibson [1958].

Afin de répondre à notre problématique, nous nous intéressons à une analyse couplée des trajectoires d'un piéton et de la perception de son environnement au cours de cette trajectoire. Nous voulons en particulier analyser l'activité du regard, car il a été montré que la vision est le principal système perceptif utilisé lors de la marche afin de collecter des informations sur l'environnement [Warren, 1998, Patla, 1997].

Dans cette optique, nous avons conçu nos expériences en réalité virtuelle (RV), dans la mesure où elle permet de contrôler efficacement les conditions expérimentales, ce qui est très difficile à réaliser dans un environnement réel rempli de personnes. Il est intéressant de noter qu'elle permet de reproduire exactement les mêmes stimuli visuels pour plusieurs essais sur un même participant et entre participants, ce qui est un élément essentiel pour les études expérimentales. Ce contrôle du stimuli visuel est crucial dans notre cas, car une différence dans celui-ci peut avoir un impact significatif sur l'activité du regard. Par ailleurs, la RV permet également d'accéder directement à l'environnement perçu par le participant, de sorte que nous pouvons savoir exactement ce que les participants ont regardé sur la base des coordonnées de leur regard et sans avoir à faire de calculs supplémentaires. Toutefois, il est nécessaire d'étudier les différents impacts que la RV peut avoir sur l'activité du regard avant de pouvoir mener nos études expérimentales. En effet, les différences de comportement humain entre un environnement réel et virtuel doivent être identifiées afin de pouvoir par la suite de transférer les résultats observés en réalité virtuelle à des situations en environnement réel.

## Contributions

Cette thèse propose trois contributions principales basées sur trois études expérimentales qui s'appuient sur le développement de deux plateformes techniques.

Dans notre première contribution, nous nous sommes donc concentrés sur l'impact de la réalité virtuelle sur l'activité du regard. Dans ce but, nous avons réalisé une expérience dans un environnement réel et virtuel avec quatre types différents d'installation de RV. Au cours de cette expérience, les participants devaient marcher vers une cible tout en évitant un autre piéton. Nous avons choisi cette situation car elle représente l'une des interactions les plus fréquentes lorsque l'on marche dans une foule. Nos résultats ont montré que l'activité du regard était qualitativement similaire dans toutes les conditions. Cependant, des différences quantitatives ont été constatées, notamment plus de rotations de la tête en RV et une plus grande amplitude pour l'angle du regard. En conclusion, la RV semble être un outil adéquat pour étudier l'activité du regard lors d'une interaction avec

un piéton. Ces travaux ont été présentés à la conférence IEEE VR 2019[Berton et al., 2019].

Suite à cette contribution, nous avons ensuite décidé, dans une deuxième expérience, d'étendre nos recherches à l'analyse de l'activité du regard pendant la navigation dans une rue très fréquentée. Cependant, cette situation étant plus complexe qu'une simple interaction avec un piéton, nous avons d'abord validé que nos résultats précédents étaient similaires pour cette situation. Nous avons ensuite étudié l'impact de la densité de la foule sur l'activité du regard afin d'obtenir de nouvelles informations sur le voisinage de l'interaction. Nos résultats suggèrent qu'à mesure que la foule augmente, le balayage de l'environnement par les yeux devient plus étroit sans pour autant changer sa fréquence de fixation. Nous avons aussi montré que le regard se concentre sur les piétons en face des participants. Ces résultats peuvent indiquer, par exemple, de considérer un nombre constant de piétons dans le *"voisinage d'interaction"*, quelle que soit la densité de la foule. Ce travail a été présenté lors de la conférence IEEE VR 2020[Berton et al., 2020].

Cependant, lorsque l'on marche dans une foule dense, il est récurrent de se heurter à d'autres piétons, ce qui peut être difficile à reproduire en RV. Dans une troisième contribution, nous avons donc étudié l'effet de la simulation de tels contacts physiques sur le comportement humain en RV. À cette fin, nous avons mené une expérience dans laquelle les participants devaient naviguer à travers une foule virtuelle dense avec ou sans rendu haptique. Nos résultats démontrent que le rendu haptique ne modifie pas les trajectoires prises par les participants, ce qui est conforme aux études [Warren, 1998, Patla, 1997] indiquant que la vision est le principal système perceptif utilisé lors de la marche. Cependant, l'introduction du rendu haptique a modifié les mouvements des participants en augmentant la rotation des épaules et en diminuant la vitesse de marche afin de pouvoir se faufiler dans la foule dense. Ce travail a été soumis à la revue TVGC et est actuellement en cours de révision.

En outre, j'ai également participé à des travaux de collaboration sur des sujets liés à la simulation de foule et à l'étude du comportement humain lors de l'interaction avec les piétons en RV. En particulier, nous nous sommes intéressés à la transgression de l'espace personnel dans une gare et dans une zone de supporters sportifs, cette étude [Duverne et al., 2020] a en outre été menée dans un environnement réel et virtuel. Dans ce travail, actuellement soumis à la conférence EuroVR 2020, j'ai été impliqué dans l'analyse des données de RV et la rédaction du plan d'expérience. Nos résultats suggèrent que les normes proxémiques varient en fonction de la relation subjective de l'individu avec le contexte social dans l'environnement réel. Toutefois, si nous avons pu montrer que les normes sociales

existent toujours dans la RV, nos résultats n'ont pas montré d'effet principal des environnements sur la tolérance des participants à la transgression des normes proxémiques. Une autre collaboration [van Toll et al., 2020] s'est concentrée sur la simulation de foule où nous avons proposé une méthode visant à reproduire plusieurs modèles de types *"Agent-based"* en utilisant l'optimisation d'une fonction de coût. Ma contribution dans cette collaboration a principalement porté sur la représentation des résultats. Finalement, j'ai participé activement au développement d'un logiciel (Chaos) permettant de visualiser les mouvements de foule à partir de trajectoires numériques. Ce logiciel permet également d'enregistrer plusieurs types de données qui peuvent être utilisées comme base de données pour des algorithmes de *deeplearning*.

UNIVERSITÉ DE RENNES 1

**Titre :** Foules virtuelles immersives : Évaluation du comportement des piétons en réalité virtuelle

**Mots clés :** Réalité virtuelle, Interaction avec les piétons, Activité du regard, Foule

**Résumé :** La réalité virtuelle (RV) est devenue un outil de plus en plus utilisé afin d'étudier le comportement humain. En effet, son utilisation permet d'avoir un contrôle absolu sur les conditions expérimentales et de reproduire le même stimulus pour tous les participants. Dans cette thèse, nous utilisons la RV pour étudier le comportement piétons dans les foules afin par la suite d'améliorer les simulateurs de foules. En particulier nous nous intéressons à l'analyse couplée de la marche et du regard pour pouvoir comprendre et modéliser le voisinage d'interaction lors de la navigation. Dans nos premiers travaux, nous avons évalué l'impact de la RV sur l'activité du regard au cours d'une interaction entre deux piétons, lors d'une expérience où les participants réalisaient une tâche d'évitement de collision dans un environnement réel et virtuel. Par la suite nous nous sommes intéressés à une situation plus complexe qui est la navigation dans une rue peuplée. Nous avons de nouveau évalué l'impact de la RV sur l'activité du regard, puis nous avons analysé l'impact de la densité de la foule sur cette activité. Finalement, dans une troisième étude nous avons simulé, en utilisant un rendu haptique, les collisions se produisant lors de la navigation dans une foule dense, et nous avons évalué l'influence d'un tel rendu sur la navigation des participants. En conclusion, nos résultats montrent que la réalité virtuelle est un outil pertinent pour l'étude du comportement des piétons dans les foules. En particulier, avec les récentes innovations technologiques, cet outil est adapté à l'étude de l'activité du regard, qui a d'ailleurs été peu explorée jusqu'à présent pour ce type de situation.

**Title :** Immersive Virtual Crowds: Evaluation of Pedestrian Behaviours in Virtual Reality

**Keywords:** Virtual reality, Pedestrian interaction, Gaze activity, Crowd

**Abstract:** Virtual Reality (VR) has become more and more used as a tool to study human behaviour. Indeed, its use provides absolute control over experimental conditions and can reproduce the same stimulus for all participants. In this thesis, we use VR to investigate pedestrian behaviour in crowds in order to subsequently improve crowd simulators. In particular we are interested in a coupled analysis of locomotion and gaze in order to understand and model the interaction neighbourhood during navigation. In our first work, we evaluated the impact of VR on gaze activity during an interaction between two pedestrians, in a study where participants performed a collision avoidance task in a real and virtual environment. We then studied a more complex situation which is the navigation in a crowded street. We again evaluated the impact of VR on gaze activity and then explored the impact of crowd density on this activity. Finally, in a third study we simulated the collisions that occur when navigating in a dense crowd using haptic rendering, and evaluated the influence of such rendering on participants' locomotion. In conclusion, our results show that VR is a relevant tool to study pedestrian behaviour in crowds. In particular, with recent technological innovations, this tool is appropriate for the study of gaze activity, which to date has been little explored for this kind of situation.