



**HAL**  
open science

# With or Without You: Effect of Contextual and Responsive Crowds on VR-based Crowd Motion Capture

Tairan Yin, Ludovic Hoyet, Marc Christie, Marie-Paule Cani, Julien Pettré

## ► To cite this version:

Tairan Yin, Ludovic Hoyet, Marc Christie, Marie-Paule Cani, Julien Pettré. With or Without You: Effect of Contextual and Responsive Crowds on VR-based Crowd Motion Capture. IEEE Transactions on Visualization and Computer Graphics, 2024, pp.1-11. 10.1109/TVCG.2024.3372038 . hal-04511956

**HAL Id: hal-04511956**

**<https://inria.hal.science/hal-04511956>**

Submitted on 19 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# With or Without You: Effect of Contextual and Responsive Crowds on VR-based Crowd Motion Capture

Tairan Yin , Ludovic Hoyet , Marc Christie , Marie-Paule Cani , and Julien Pettré 

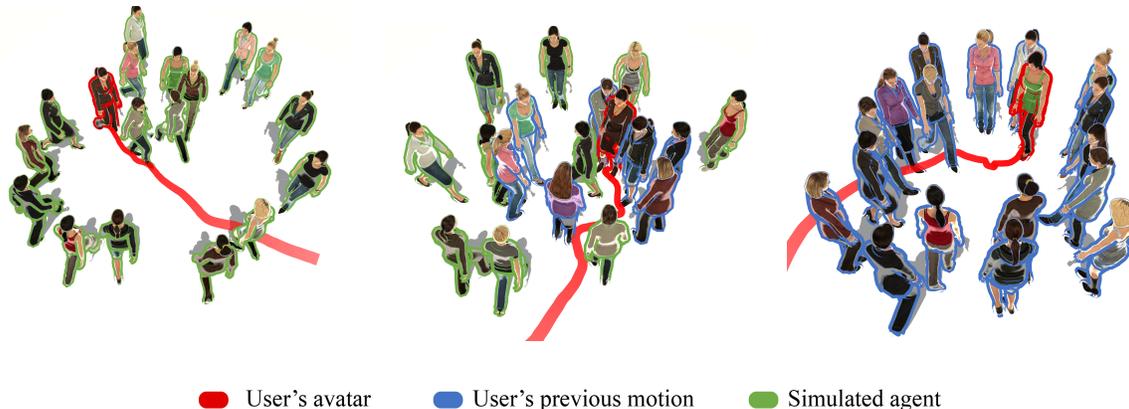


Fig. 1: Illustration of our proposed 3R paradigm (Replace-Record-Replay), where a single user (red outline) is initially immersed into a simulated *contextual crowd* whose autonomous agents (green outline) are successively replaced by the user's captured data (blue outline). The consistency of the user's behavior in this 3R scenario is compared with 4R (Replace-Record-Replay-Responsive), where the blue agents are made locally responsive.

**Abstract**—While data is vital to better understand and model interactions within human crowds, capturing real crowd motions is extremely challenging. Virtual Reality (VR) demonstrated its potential to help, by immersing users into either simulated virtual crowds based on autonomous agents, or within motion-capture-based crowds. In the latter case, users' own captured motion can be used to progressively extend the size of the crowd, a paradigm called Record-and-Replay (2R). However, both approaches demonstrated several limitations which impact the quality of the acquired crowd data. In this paper, we propose the new concept of contextual crowds to leverage both crowd simulation and the 2R paradigm towards more consistent crowd data. We evaluate two different strategies to implement it, namely a Replace-Record-Replay (3R) paradigm where users are initially immersed into a simulated crowd whose agents are successively replaced by the user's captured-data, and a Replace-Record-Replay-Responsive (4R) paradigm where the pre-recorded agents are additionally endowed with responsive capabilities. These two paradigms are evaluated through two real-world-based scenarios replicated in VR. Our results suggest that the behaviors observed in VR users with surrounding agents from the beginning of the recording process are made much more natural, enabling 3R or 4R paradigms to improve the consistency of captured crowd datasets.

**Index Terms**—Crowd Simulation, Human Interaction, Virtual Reality

## 1 INTRODUCTION

Crowd motion data plays a critical role in understanding and modeling crowd dynamics and local pedestrian behaviors. While comprehensive datasets that include body movements, eye gaze direction, and hand gestures would be valuable, most of the existing crowd datasets primarily document pedestrian positional data, due to technological constraints [22]. However, with the emergence of data-driven crowd modeling approaches, the need for rich and large datasets is paramount. As traditional approaches struggle to find a good compromise between data reconstruction and time-money cost efficiency, Virtual Reality (VR) has recently been explored as a data collection tool [3, 8, 48]. Indeed, performing data acquisition in an immersive virtual environ-

ment allows for improved scene control and repeatability, a wider range of acquired modalities, as well as more efficient data processing with guaranteed synchronization.

However, applying VR-based motion capture to crowds is extremely challenging. Let us describe the two alternative strategies that have been explored so far, and explain why each of them has partially failed. The first strategy is to immerse the captured participant in a simulated, virtual crowd, consisting of autonomous – and therefore reactive – agents (e.g., [10, 44]). This approach has proven to be very appropriate for closely observing the reaction of one (at a time) individual to a given situation within a crowd. Repeating the same experiment with numerous participants allows obtaining various possible responses to the same situation, and to understand the average behavior. Let us stress that this first strategy is however not applicable for recording full crowd datasets (i.e., the behavior of multiple participants at once), since everything but the VR user is simulated. The second strategy alleviates the above limitation by enabling users to create a full crowd dataset on their own, thanks to an iterative process enabling them to record successive trajectories (e.g., [67]). The final dataset thus consists entirely of captured data. Unfortunately, the successive iterations may alter the realism of the users' behavior, since they start recording in an empty scene and react only to their own previously recorded data.

- Tairan Yin, Ludovic Hoyet, Marc Christie, and Julien Pettré are with Inria, Univ Rennes, CNRS, IRISA, France. E-mail: {tairan.yin | ludovic.hoyet | julien.pettre}@inria.fr, marc.christie@irisa.fr
- Marie-Paule Cani is with Ecole Polytechnique/CNRS (LIX), IP Paris, France. E-mail: marie-paule.cani@polytechnique.edu

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: [reprints@ieee.org](mailto:reprints@ieee.org). Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxxx

Therefore, the objective of this work is to propose and validate a new paradigm for creating fully VR-based crowd datasets that are consistent with natural behavior. To achieve this, the insight is to combine the two approaches described above, namely – to always expose subjects to a realistic surrounding crowd, but also to allow them to compose a complete crowd dataset where the final data results only from the actual actions of a participant. To this end, we build on the recent idea of recording a crowd dataset through an iterative process [67] – which we call here a **Record-and-Replay (2R)** process – and extend this process by immersing participants in *contextual crowds*, defined as numerous characters (agents) that provide participants with realistic surroundings from the beginning.

Specifically, we propose two versions of this new paradigm: **Replace-Record-Replay (3R)**, where, instead of an empty scene as in 2R, users move among a contextual crowd and progressively replace, one at a time, each of its virtual agents; **Replace-Record-Replay-Responsive (4R)** where, extending 3R, a motion correction strategy is added to agents by locally editing their motion when the user is close. This makes user-recorded characters responsive, e.g., by enabling them to show collision avoidance behaviors. Our exploration of these two paradigms leads to the understanding of: (1) the effect of being immersed in a contextual crowd on the user’s behavior during the crowd generation process, and (2) the effect of having responsive characters deviate from their original trajectories on the user’s behavior, which indeed impacts the quality of the generated crowd data. The exploration of these effects relies on behavioral metrics to compare the properties of trajectories recorded using different paradigms, or with real data.

Our contributions are threefold:

1. We introduce the concept of **contextual crowds** to combine the advantages of immersion in a simulated crowd with those of Record-and-Replay methods, towards consistent crowd motion capture in VR. We propose two alternative setups (3R and 4R), and compare their results on two well-studied crowd scenarios.
2. Using the 3R paradigm, we validate the use of contextual crowds to induce interactions during the crowd generation procedure. We further demonstrate that simulated contexts have similar effects as reusing real contexts.
3. Using the 4R paradigm, we demonstrate that the use of locally responsive agents improves the quality of user collision avoidance, at the cost of locally deviating from user-captured-only data.

After discussing related work in Section 2, we introduce the key concepts in Section 3, and detail our experiments in Section 4. In Section 5 we present our analysis methodology, and discuss the corresponding results in Section 6. We finally discuss our method and the main findings in Section 7, before concluding.

## 2 RELATED WORK

In this section, we explore the connections between crowd data, crowd capture in VR and crowd simulation, and position our work in the context of modern VR-based crowd dataset capture.

### 2.1 Human Crowd Datasets

Human crowd datasets are of paramount importance in understanding [22], predicting [1, 65] and modeling [68] crowd behavior. Such datasets allowed to study both individual [39] and collective [28] behavior, under unidirectional [19, 35, 54], multi-directional [12], and some specific navigation scenarios [26]. They are of great interest in observing rich interactions among pedestrians [57], and have been used to evaluate simulations [64], to develop knowledge-driven [47] and data-driven models [13], as well as to infer implicit causality in crowd motion [36].

Among this body of work, early-stage research focused on the analysis of trajectory data and dissected crowd movements at both the individual and collective levels. For instance, multidirectional scenarios were shown to provoke emergent patterns such as lane and stripe formations [41], i.e., they enabled the identification of emerging self-organization phenomena resulting from the individual behaviors of

pedestrians [57]. Similarly, it was shown that the aggregation of individual behaviors leads to a statistical relationship between walking speed and density [55].

Recent research has extended the study to the understanding of low-level individual features such as gaze [30, 31], gait and posture [56], and other specific interactions observed in crowds [20, 21]. These explorations are, however, facing the limits of data. Existing crowd datasets were mainly captured through field observations [11, 14, 25, 69] or laboratory controlled experiments [2, 18], and therefore often contain only trajectory data. In addition, obtaining additional information from real crowds is quite challenging, as existing data construction pipelines, including detailed individual motion capture, suffer from either technical or ethical challenges [22, 58].

### 2.2 VR Datasets based on Virtual Crowds

Recent research has leveraged VR to create human-related datasets. Unlike real-world capture, VR allows for cost-efficient immersive experiences with multimodal data recording (possibly including full-body motion, gaze, sound, etc.). Since the experimenter has full control over the 3D environment, VR data capture provides high controllability, repeatability, and low risk. Therefore, VR was used to capture both human-environment [3] and human-crowd [7] interactions. In particular, VR-based data collection enabled the exploration of avoidance behavior [38, 42], speed-density perception and adaptation [17, 59], path planing [49], herding effects [33, 40], and response to stress [40, 52]. Additional data captured during these experiments, e.g., body motion [67] or eye gaze [8, 50], also contributed to a better understanding of crowds.

To be effective, such VR experiments require virtual crowds to provide an appropriate sense of immersion, and induce correct user behaviors. Since these experiments typically involve a single user at a time, the virtual crowd surrounding them must be carefully designed. A common approach is to use microscopic simulation of autonomous agents capable of responding to the user’s movement in real-time [63]. Previous research in crowd simulation has explored force-based [29], velocity-based [60, 61], data-driven [66] and deep-learning-based [1] approaches. Their performance depends heavily on the choice of parameters, especially the preferred speed [60]. Thus, parameter calibration [64] is mandatory to ensure a good quality of simulation [15].

Given that these simulation methods have limited capabilities to reproduce human behavior, the alternative of reusing real MoCap data for the surrounding crowd, at the expense of the agents’ interactivity, has been explored in an attempt to maximize realism [46]. Yin et al. combined MoCap with VR to propose the One-Man-Crowd (OMC) [67], a specific instance of the aforementioned Record-and-Replay paradigm, which enables to generate virtual crowd data without any simulation. In brief, single users are embodied into an avatar and move in a VR scene, while their body motion is recorded using MoCap. The system then creates a virtual character that replays the recorded motion while the user successively acts out another character. This process is repeated until the entire crowd is recorded. While this paradigm provides valuable insights on how to generate crowd data in VR, it has two major limitations, as mentioned in the original paper: 1) At the beginning of the generation process, users face an empty scene, and are therefore unlikely to initiate interaction behaviors with the future crowd to be generated. 2) During the process, more and more passive characters are created. They do not respond to the user’s actions, since they are only replaying recorded data. As a result, it becomes increasingly difficult for the user to avoid collisions and to behave naturally.

When using VR to generate crowd datasets, it is of paramount importance to evaluate the similarity between user behaviors in real and virtual situations. Unfortunately, VR-based experiments have been shown to introduce a number of perceptual and behavioral biases, such as affected depth perception [4, 51], reduced walking speed [6, 23], increased social distance [5, 10, 27, 37, 53], and increased eye gaze and head movements [8, 9]. Due to this, considerable work has been done to assess how these biases affect behavior in VR. The general finding is that VR motion capture leads to measurable changes in a majority of spatial variables, although the qualitative nature of participant behavior may remain consistent [67].

### 3 KEY CONCEPTS & NOTATIONS

This work aims to propose a novel approach for VR-based crowd motion capture that goes beyond the limitations of a traditional Record-and-Replay method. While inspired by Yin et al.'s work [67], we explore the benefits of using contextual and responsive crowds to improve the consistency of users' behavior, towards the generation of more realistic crowd datasets. The remainder of this section defines the fundamental concepts underlying our study, i.e., the different models we use for virtual agents, the alternative VR motion capture processes, and the experimental scenarios we choose for validation.

#### 3.1 Agents

Agents are the animated virtual humans used to populate a virtual scene. We use several types of agents, which differ by the method that drives their **motion**, by their **role** in the data collection process or finally by their **interaction capabilities**.

**Motion.** Agents can be animated using different methods:

- **OMC agents** refer to agents whose movement was recorded by the participant in a previous iteration (see Section 2).
- **Captured agents** also move from pre-existing motion data (e.g., crowd datasets), but not recorded by the participant. When 2D-only crowd datasets are used, 2D trajectories are augmented into 3D full-body animations using animation techniques.
- **Simulated agents** refer to agents whose motion is fully computed using a real-time simulator (e.g., some microscopic crowd simulation method), enabling them to make real-time decisions to reach a predefined destination in the scene while avoiding collisions. For captured agents, the computed trajectory is combined on the fly with other animation techniques to produce 3D full-body animated agents.

**Role.** Our goal is to enable a single user to create new crowd motion datasets. For such datasets to originate from the motion of participants, **OMC agents** data are stored (and can then be edited by the 4R mechanism, as explained below). The other two types of agents (captured or simulated) may be used as **Contextual agents**, i.e., agents intended to provide users with contextual information, while they successively act to capture motion for the OMC agents. Given our new concept of contextual crowds, Contextual agents are to be present in the virtual scene from the first motion capture iteration.

**Interaction Capabilities.** From the agent descriptions given above, we understand that, at any given moment, we can immerse the participant among agents whose movements are pre-recorded and who ignore the participant's actions. Since collision avoidance behaviors are expected to be reciprocal, this may cause the user to fail avoiding collisions. Therefore, we investigate the possibility of providing both OMC and Captured agents with collision avoidance capabilities, and define:

- **Non-responsive agents** as Captured or OMC agents, non-responsive as they only move by following pre-recorded data and cannot interact with the user.
- **Responsive agents** as either Simulated (i.e., inherently responsive), or modified Captured or OMC agents that locally adjust their motion upon the detection of a risk of collision with the user.

**Agents implementation details.** In our experimental setup, the ORCA crowd simulation technique [60] is used to generate the global trajectory for Simulated agents as well as the reactive portions of the trajectory for Responsive agents, using the implementation provided by van Toll et al. [62]. ORCA is always activated for Simulated agents, but needs to be activated and deactivated for the other types of Responsive agents according to the risk of collision with the participant. To this end, we simply check the distance  $d$  between any agent and the participant, and trigger activation when  $d < 1 m$ . ORCA is active for at least  $0.2 s$  to avoid flickering effects due to frequent activation and deactivation.

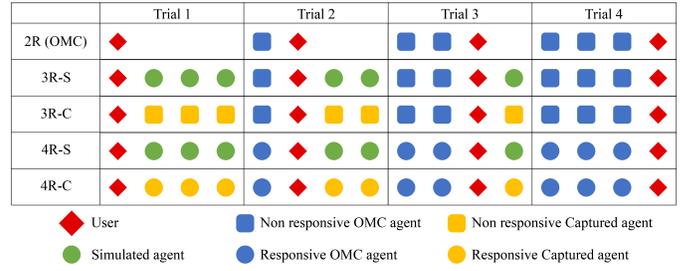


Fig. 2: Illustration of the different paradigms in an example with 4 agents. The figure illustrates the behavior of each agent before and after being acted out by the user, resulting in a total of 4 trials to record the exemplar scenario. Since Simulated agents are always responsive, they come in a single version (disk shape).

The preferred velocity of Simulated agents is calculated to steer them towards their final goal (see Section 3.3). The preferred velocity of the other categories of reactive agents is calculated to recover their pre-recorded trajectory when ORCA is further deactivated. 2D trajectories are augmented into 3D animated characters using a Unity animator module driven by walking speed, in combination with a set of motion-captured walking animations.

#### 3.2 Crowd-motion-capture process

The main objective of our work is to evaluate the importance of adding contextual agents when a single user records a crowd dataset. To this end, we propose and compare a number of variations of the OMC paradigm (illustrated in Figure 2):

- **2R** stands for “Record and Replay”. This paradigm is our baseline, and corresponds to the “Vanilla” OMC technique. In this version, the user is initially alone in the virtual scene, which is iteratively populated with OMC agents (see Figure 2).
- **3R** stands for “Replace, Record and Replay”. In this version, the virtual scene is initially populated with **contextual agents**. The participant iteratively embodies each contextual agent, to record a new movement and replace it with a **non-responsive** OMC agent. We distinguish 2 versions of the **3R** paradigm: i) **3R-C** where contextual agents are **captured agents**, thus all **non-responsive**, and ii) **3R-S** where contextual agents are **simulated agents**, thus all **responsive**.
- **4R** stands for “Replace, Record, Replay and Responsive”. In this version, the virtual scene is initially populated with **responsive contextual agents**. The participant iteratively embodies each contextual agent, to replace it with a **responsive** OMC agent. We distinguish 2 versions of the **4R** paradigm: i) **4R-C** where contextual agents are **responsive captured agents**, and ii) **4R-S** where contextual agents are **simulated agents**, thus **responsive**.

In summary, the **2R** condition is our baseline. Since one of its flaws is the lack of contextual information, the **3R** condition allows us to evaluate the importance of adding contextual agents. In addition, the **4R** condition enables us to evaluate the importance of also preventing collisions between agents and participants. Finally, the **XR-S** conditions allow us to evaluate the use of simulation-based contextual agents instead of those based on captured data (**XR-C**).

#### 3.3 Experimental scenarios

To evaluate and compare our paradigms in the challenging task of capturing crowd datasets with many multi-directional interactions, we replicated two existing laboratory crowd scenarios.

**Circle.** In this scenario, the starting positions are homogeneously distributed around a circle. The goal of the agents is to walk to the antipodal point of the circle, which maximizes the number of interactions [45]. Such a scenario is challenging to crowd simulation, as

congestion may occur as the agents converge to the center. Previous research has used this scenario to validate simulation algorithms [32, 34] and calibrate simulation parameters [16, 64]. In our experiments, we replicated the setup presented in [64] using a set of captured data with 18 pedestrians, as depicted in Figure 4-a (top).

**Crossing Flow.** This scenario, representative of regular daily crowd movements, corresponds to a flow of pedestrians walking in two orthogonal directions and sharing a crossing area. It is complementary to the Circle scenario, as it involves a continuous human flow in each direction, and therefore continuous interactions. We adapted the experimental setup presented in [41], using a set of captured data with 18 pedestrians as depicted in Figure 4-a (bottom). To adapt to the physical constraints of our VR experiment, we retained the first 9 pedestrians of each group, and selected the moment when the first pedestrian enters the crossing area, along with the positions of these 18 individuals at that time, as the starting point of our experiment.

## 4 EXPERIMENT

In this section, we present the experiment we designed to evaluate the performance of the 3R and 4R paradigms. In particular, our experiment was designed to assess whether these two novel VR crowd motion capture methods can address the limitations of the 2R method, where users 1) freely walk in an empty scene during the first trials instead of walking among a moving crowd, and 2) struggle to avoid collisions as more and more non-reactive agents are progressively added to the scene. This has been shown to lead to biased participant behavior, as they are 1) unlikely to initiate interactions when faced with an empty experimental environment, and 2) likely to experience unavoidable collisions when faced with a dense non-responsive crowd. Our general hypotheses are

- **H1:** By initializing scenarios with contextual crowds, we can mitigate the bias caused by the fact that users initially walk alone in an empty scene.
- **H2:** In the 3R paradigm, the type of contextual crowd, whether it is created by simulation or reanimation from previously captured data, does not significantly impact the results.
- **H3:** The local responsiveness of the surrounding agents can reduce the number of collisions experienced by users and thus improve the consistency of the dataset. However, frequent changes to the trajectories already recorded by the participant may gradually decrease its similarity to human behavior.

### 4.1 Apparatus

Participants experienced the virtual environment using a Pimax 5K Head-Mounted Display (HMD) with the following characteristics: 90 Hz refresh rate, 200° field of view, resolution of 2560×1440. The HMD’s wide field of view was ideal for situations requiring proximity to other characters. The environment was created using Unity 2021.3. Four SteamVR 2.0 base stations were used to provide a tracking area of approximately 10 m × 10 m. For the experiment, the participants’ physical movements were captured using an Xsens motion capture system, displayed in real-time on their avatar, and used to animate the corresponding character’s motions in subsequent trials. Participants were also equipped with an HP Z VR G2 backpack (Intel Core i7-8850H, NVIDIA RTX 2080, 32 GB RAM). All the necessary devices were physically connected to the backpack, allowing participants to move freely in the real world while simultaneously interacting with the virtual crowd and observing their own movements reflected in their co-located avatar.

### 4.2 Experimental Design

Participants first read and signed the experiment consent form, and were presented with the task to be performed. They were then equipped with the Xsens motion capture system, the VR backpack, and the Pimax 5K HMD. The motion capture system was then calibrated under the guidance of the experimenter to ensure the quality of the motion capture and to ensure that the size of the avatar matched the dimensions of

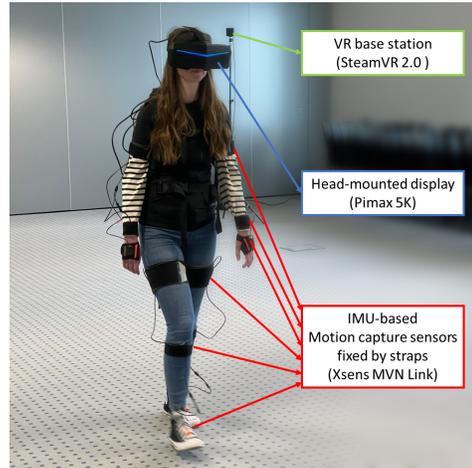


Fig. 3: Illustration of the apparatus. During the experiment, participants are equipped with a motion capture system to capture their body motions. They also wear an HMD with a large field of view. These devices are connected to a VR backpack worn on the participant’s back.

the participants. Participants were then immersed in an empty tutorial room and asked to walk six times back and forth from one side of the room to the other, to familiarize themselves with walking in a virtual environment. Meanwhile, the participants’ comfort walking speed was measured during this process, and used to adjust the preferred walking speed of the simulated characters (see details below).

During the experiment, participants went through the 5 experimental conditions: 2R, 3R-S, 3R-C, 4R-S, 4R-C. We informed the participants that different agents might have different behavior during the entire experiment, but did not provide information about our record and replay paradigm, neither any specific information about the various conditions we studied. The order in which each participant completed these 5 blocks was randomized. In each condition, participants were always immersed successively in all the different virtual characters of the sequence in a fixed order. To ensure a smooth and realistic experience, we adjusted the speed of the contextual agents, which prevented participants from falling behind and losing contact with the contextual crowd due to a reduced walking speed. For *captured agents*, we adjusted the speed of the real trajectories using the factor  $f = V_U/V_R$ , where  $V_U$  represents the participant’s comfort speed, measured during the tutorial and unique to each individual, and  $V_R = 1.4 m/s$  represents the empirical comfort walking speed of pedestrians in real life [9]. Similarly, we randomly generated a preferred walking speed for each *simulated agent* within  $1.4 \pm 0.2 m/s$  as a constant default speed for each experience, and then varied their preferred walking speed by the factor  $f$ .

In each trial, participants walked from one side of the scene to the opposite side, until they reached a green dot (in the Circle scenario) or green line (in the Crossing Flow scenario) on the ground, located 8 m in front of the starting position. In the case of green line, participants could pass through any point on the line or its extension on either end. After reaching the goal, the trial finished and participants were guided back to the starting position to begin the next trial.

### 4.3 Participants

Thirty participants took part in our experiment (average age of 31.6 ± 8.4 y.o., ranging from 22 to 51 y.o.). They were recruited via internal student and staff emailing lists. All participants were naive to the purpose of the experiment and had normal or corrected-to-normal vision and gave written informed consent. The study followed the Declaration of Helsinki, and was approved by the Inria internal ethics committee (COERLE). Each participant performed only one scenario (but all five conditions for that scenario). In total,  $15 \times 5 \times 2 = 150$  sets of crowd motion were generated, containing 2700 individual trajectories, with an average duration of 14.8 s.

## 5 ANALYSIS

We divide our analysis of the experiment into two parts. The first analysis focuses on the effect of contextual crowds, which corresponds to the *Replace* component in both paradigms. The second analysis focuses on the effect of the local correction of user-generated trajectories, corresponding to the additional *Responsive* component of the 4R paradigm.

### 5.1 Contextual Crowds

To compare trajectories of different lengths and durations across participants, we normalize the captured trajectories over the distance covered. This normalization leads to relative walking progress from 0% (start) to 100% (end) of each trajectory. For trajectories with different shapes and lengths, the *relative progress* provides us with a universal “semantic” reference. These normalized trajectories are used to compute the following characteristics. We then leverage Statistical Parametric Mapping (SPM) [24] to study the differences and similarities of these characteristics under different conditions.

**Average Trajectories.** Since the participants act and replace the contextual agents in a fixed order, we perform statistical analysis on the generated movements with respect to each agent. This is done by overlapping the same agent’s relative walking trajectories of different participants (see above). For each condition  $c$  and each agent  $a$ , we compute the mean trajectory  $\bar{\mathbf{x}}_{c,a}^*$  of all the participants:

$$\bar{\mathbf{x}}_{c,a}^* = \frac{1}{M} \sum_{j=1}^M \mathbf{x}_{c,a,j}^* \quad (1)$$

where  $\mathbf{x}_{c,a,j}^*$  is the trajectory of participant  $j$  for agent  $a$  in condition  $c$ , and  $M$  is the number of participants. The average trajectory reveals the mean behavior of participants when acting out each specific agent.

**Normalized Speed.** In general, participants experience three distinct phases during walking: an initial acceleration phase from the starting point to just before entering the crossing area, a deceleration phase upon entering the crossing area due to interactions and avoidance maneuvers with other agents, and a re-acceleration phase after disengaging from the agents. We thus study the velocity profiles under each condition. Since each set of crowd motions was generated by a single participant, the crowd’s moving speed depended on the participant’s preferred walking speed. This speed is unique for each participant. Thus, we normalized each participant’s data by their comfort walking speed  $v_p$  before performing statistical analysis. With a similar idea, we normalized real data’s speed when making comparisons between the real data and our results. The real data, being a set of trajectories from different individuals, cannot be normalized by a particular person’s preferred speed. We therefore use an average speed value  $v_p^* = 1.4 \text{ m/s}$  provided by previous research [9] to compute the normalized speed  $v_n = v_r/v_p$ , where  $v_r$  is the speed from the real data.

**Lateral Movement.** To estimate the amount of effort participants put into avoiding collisions, we compute their lateral movement for each trial. In an empty space, participants would ideally walk straight from their start point to their destination, so we assume that lateral movement is the result of avoidance behavior. It is computed by projecting the participants’ positions onto the line segment between their start point and destination, and computing the distance to the projected points.

**Local Density.** The local density describes the participant’s surrounding density, and relates to the number of agents who are in their proximity at each given time. This metric reveals both the difficulty of collision avoidance and the size of the personal zone with which participants are comfortable. To compute the local density around each character at each time frame, it is common to perform a Voronoi tessellation of the space, and then compute the density as the inverse of the area of the character’s Voronoi cell. To avoid artifacts caused by edge effects, we adopted a modified Voronoi tessellation [43], in which an agent is

considered to be on the edge if its Voronoi cell shares an intersection with the convex hull of all the agents’ positions. Nicolas et al. [43] define the angle  $\theta$  that describes the fraction of the non-empty space in each cell, such that the local density can be computed as:

$$\rho_i = \frac{\theta}{2\pi A_i}, \quad (2)$$

where  $A_i$  is the area of the Voronoi cell intersecting with the convex hull of the agents. For agents that are not on the edge,  $\theta = 2\pi$  and thus the local density is the inverse of the area of the Voronoi cell.

**Distance to Nearest Neighbor.** We aim to study participants’ tolerance of proximity to other agents while they are walking in virtual crowds. By assuming that they do not allow any virtual agent to enter their minimal personal zone, we focus only on the positional relationship between participants and the closest virtual agent. To do so, we compute the distance to the nearest neighbor of the participant in each frame.

**First vs. Last trajectories.** We further investigate whether participants behaved differently between the beginning and the end of the experiment. To do so, we analyze the aforementioned normalized walking speed and lateral movement. These two metrics focus on the individual characteristics exhibited by each crowd member, and thus can separately reflect the participants’ behavior during each trial. For each trial, we compute a single mean value of both the normalized walking speed and lateral movement, then aggregate the results across all participants.

### 5.2 Local Corrections due to responsive agents

To evaluate the effect of local corrections of the user-captured trajectories due to reactive agents, we take a multi-faceted approach. Due to the novel nature of the problem (i.e., ideally, evaluating the reduction in realism due to these local corrections), there is no established analysis framework to rely on. Therefore, in order to provide a comprehensive understanding, we selected several metrics for our analysis. Initially, we investigate the effectiveness of local corrections in reducing user-agent collisions. Subsequently, we delve into its impact on previously generated trajectories. Specifically, we seek to answer the following questions: How frequently is the local correction invoked? In which parts of the agent trajectories do local corrections occur? How do the adjusted trajectories differ from the original user-generated trajectories?

**User Collision.** We examine the average number of collisions between the participant and the OMC agents in each condition. Collisions are determined based on interpersonal distance (0.5 m threshold).

**Correction Frequency.** To examine the frequency of changes in OMC agent trajectories, we count the number of responsive OMC agents that changed in each trial. As the agent count increases with each trial, so does the likelihood of local correction. This makes each trial individually valuable for analysis, leading us to perform evaluations on a per-trial basis rather than aggregating all trials.

**Correction Location.** We examine the trajectory modifications of all OMC agents within each trial to understand the occurrence of local corrections. To this end, we measure the position along the agent’s original trajectory – expressed as the relative walking progress – where the local correction occurs. To discern overarching patterns, we aggregate the data from all participants and calculate the average position of local correction occurrences for each OMC agent in each trial. This approach allows us to scrutinize the relationship between trial number, agent index, and the triggering of local correction.

**Trajectory Dissimilarities.** We further shift our focus to quantifying the dissimilarities between the OMC agents’ modified trajectories and the original ones within each individual trial. To accomplish this, we use Dynamic Time Warping (DTW) as a metric to measure these differences. DTW offers the advantage of nonlinear matching capabilities, allowing us to align sequences that may differ in time but have similar shapes. It also provides a balanced integration of both local and global

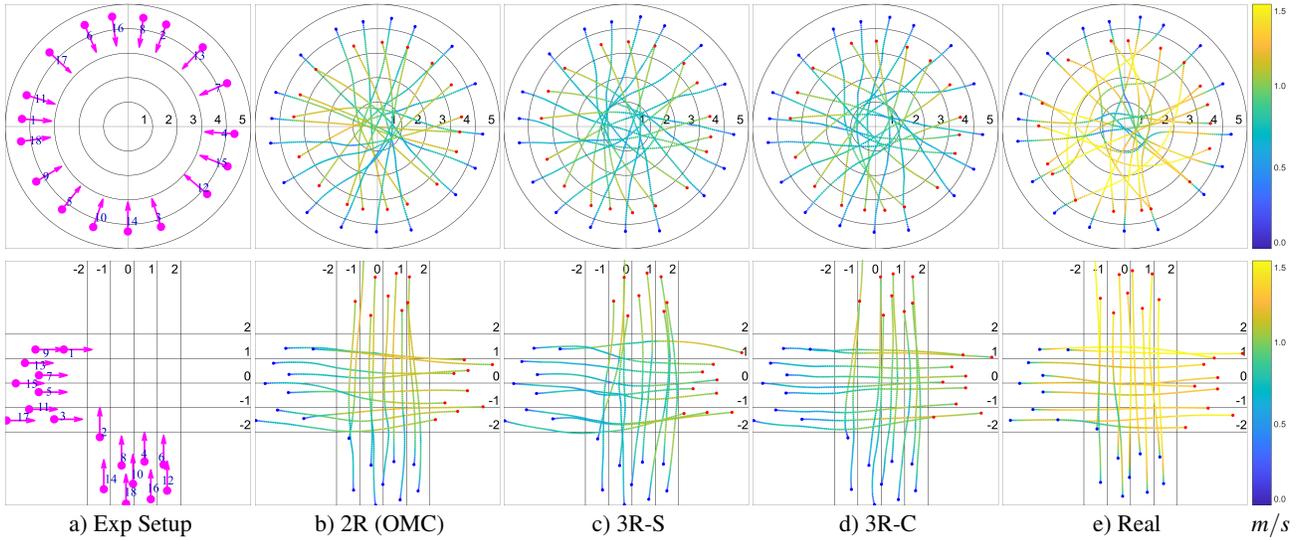


Fig. 4: a) Initial conditions of our experimental scenarios (top: Circle; bottom: Crossing Flow). Dots depict initial positions, arrows initial moving directions, and numbers the order participants followed to act the virtual characters. b-d) Mean trajectories under different conditions (aggregated over participants), colored by average speed; Blue and red discs indicate respectively starting and ending locations. e) Real data used to replicate the experiment. Equidistant 1 m guidelines (top: concentric; bottom: horizontal and vertical) provide visual information about the crossing areas.

information, enhancing the robustness of our analysis. Again, we aggregate the data by averaging results across all participants to identify possible systematic effects related to scenarios, trial number, and agent index.

## 6 RESULTS

As in the previous section, we present our results in two subsections, dedicated to the effect of contextual crowds and the effect of agent responsiveness, respectively. Since the two experimental scenarios were analyzed in the same way, we present their results together, while highlighting their similarities and differences.

### 6.1 Contextual Crowds

To isolate the effects of contextual crowds, we only present in this section comparisons between the 2R and 3R-X conditions, as well as real data. Additional comparisons to the 4R-X conditions are provided in the supplemental material for completeness.

**Average trajectories.** We start by reporting the average trajectories in each of these conditions, shown in Figure 4. This figure is useful for visual inspection and understanding. For instance, we see that the 2R and 3R-X (X represents either S or C) conditions present each unique features, with, for 3R-X, lower speeds and more gradual speed changes within the crossing area (the central region of each scenario). Further quantitative comparisons are provided by Figure 5. Note that the figure reports quantities as a function of the progression on the trajectory, normalized by its length. Thus, 0% is the start position, 100% corresponds to the end position, and the 40% to 60% portion of the trajectory is where the interactions are most intense.

**Normalized Speed.** Figure 5-a illustrates the average evolution of the normalized speed in each condition. In the Circle scenario, we observe that all results follow an accelerate-decelerate-re-accelerate pattern, but the location of the deceleration phase differs: the 3R-X results are closer to the real data, while the 2R results are significantly different from the real data and 3R-X. SPM analysis confirms these observations by detecting significant differences between the 2R and real data (47%-100%), and between the 2R and 3R-X conditions (2R/3R-S: 44%-88% and 2R/3R-C: 45%-80%). In comparison, the 3R results differ significantly from the real data only at the end of the trajectory, suggesting that participants did not come to a complete stop upon reaching the endpoint. We observe similar trends in the Crossing Flow

scenario, although the deceleration phase is less pronounced. SPM analysis also reveals significant differences between the 2R condition and the real data (51%-84%), and between the 2R and 3R-X conditions (2R/3R-S: 34%-87% and 2R/3R-C: 36%-100%). Again, few significant differences were found between the real and 3R-X results.

**Lateral Movement.** Figure 5-b depicts the average lateral movement in each condition. The 3R-X results show higher lateral movement near the crossing area compared to both the 2R and real results. For the Circle scenario, while SPM analysis revealed a 100% significant difference of 2R compared to 3R-X, it did not show any significant difference with the real data. In contrast, for the Crossing Flow scenario, SPM shows that the 3R-X results have a 100% significant difference from the real data, while the 2R condition is only significantly different in the 0%-70% range. Other partial differences were observed between the 2R and 3R-X conditions. We highlight that 3R-S shows larger lateral movements compared to 3R-C in the 63% to 100% phase, suggesting that participants tend to make larger lateral movements and significant deviations from their goal when faced with simulated contextual crowds.

**Density.** While we can observe in Figure 5-c that local density is slightly higher for the real data than for any other condition, SPM analysis shows significant differences only in extremely small areas. These differences occur briefly in the Circle scenario (about 50% of the walking progress for 2R and 3R-X conditions), and earlier in the Crossing Flow scenario (2R/Real: 25%-48%; 3R-C/Real: 40%-45%). For the Crossing Flow scenario, SPM analysis also showed some significant differences between the 2R and 3R-X conditions (2R/3R-S: 83%-97%; 2R/3R-C: 12%-30% and 64%-98%). Although the significant differences in density within the crossing area decrease in the Crossing Flow scenario, real data still maintains the highest average density.

**Distance to nearest agent.** Figure 5-d shows that the analysis of the distance to the nearest neighbor did not reveal any meaningful significant difference between the different experimental conditions.

**First vs. Last trajectories.** The iterative process set by the 2R or 3R-X paradigms results in an order for each recorded trajectory. Figure 6 shows that the 2R paradigm introduces more differences between first and last recorded trajectories, both in normalized speed and lateral movement. In particular, the Circle scenario exhibits a noticeable trend

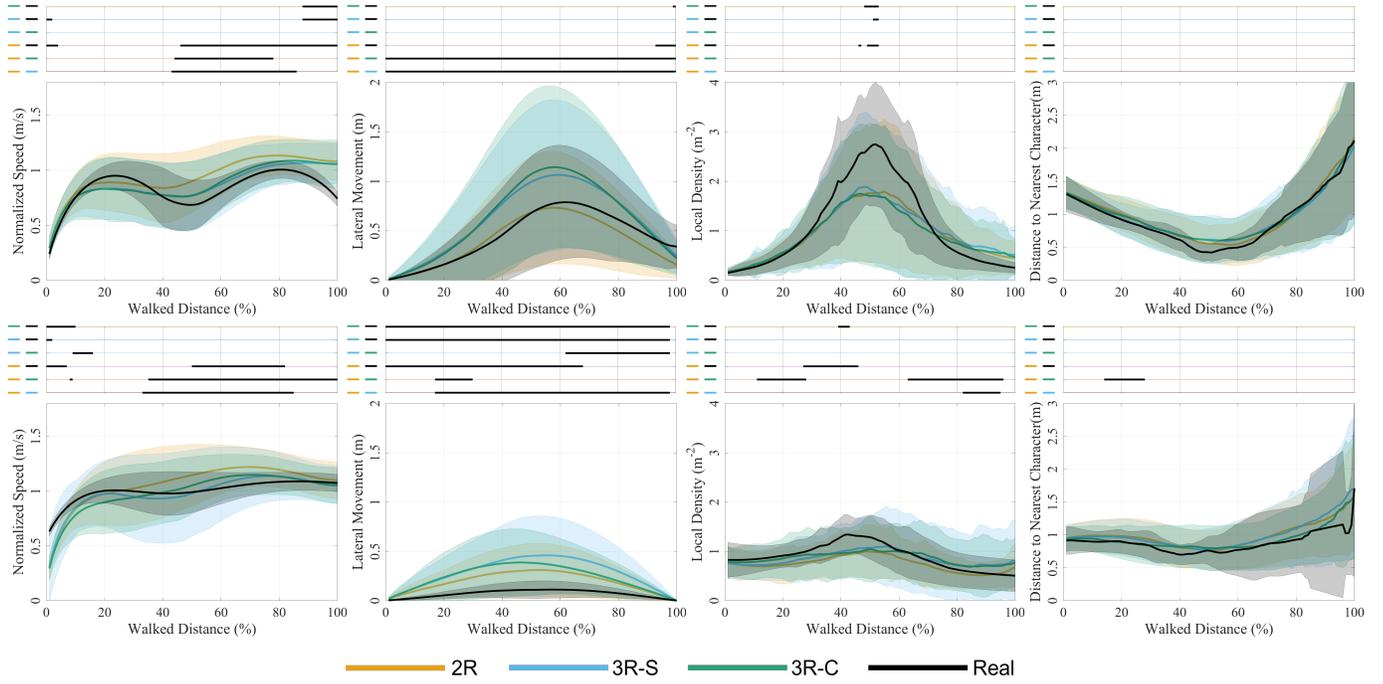


Fig. 5: Evolution along the distance that participants walked (in percentage) of participants' normalized speed, lateral movement, local density, and distance to the nearest character. Top row: results of the Circle scenario. Bottom row: results of Crossing Flow scenario. The legends are depicted beneath. In each row, we present the corresponding SPM significant differences above the figures. In the SPM results, each horizontal black line indicates during which interval of walked distance significant differences were detected between two conditions. The compared conditions are indicated by the left side of each line.

of higher speed and lower lateral movement during the first 6 trials, indicating that participants tended to walk straight forward across the scene, as the scene was not very populated in the beginning. The 3R-X results, on the other hand, show consistency among different trials. This implies that participants tended to make more similar efforts across trials. The observation remains similar in the Crossing Flow scenario, although the difference between conditions is smaller.

## 6.2 Local Corrections from Responsive OMC Agents

The effect of local corrections is now studied by comparing 4R-X conditions with others. While we report the results aggregated over 4R-S and 4R-C, as we found similar results during our analyses, the detailed analyses are available in supplementary material for completeness.

**User Collision.** As shown in Table 1, the local correction strategy effectively reduced the number of collisions between participants and OMC agents. Although there are still some collisions, our method generally reduced the number by half in both scenarios. This observation suggests that participants had less difficulty avoiding in the 4R conditions than the 3R conditions. However, since the OMC agents were only set to be responsive when the human user was nearby, they became non-responsive afterward, leading to the final generated crowd showing more collisions (4R-S vs. 3R-S, 4R-C vs. 3R-C), especially in the Crossing Flow scenario. This suggests the presence of irregular trajectories in the 4R-X results, which is discussed in Section 7.2.

**Correction Frequency.** Figure 7 shows the number of OMC agents whose trajectories were modified with respect to the trial number during the 4R process. As the recording process progresses, participants encounter more OMC agents, leading to an increased number of OMC agents that trigger local corrections. However, it is important to note that participants do not induce local corrections in all OMC agents in every trial. Using a linear regression, we found that on average 30% of OMC agents experienced local corrections in each trial, a finding that holds across both experimental scenarios.

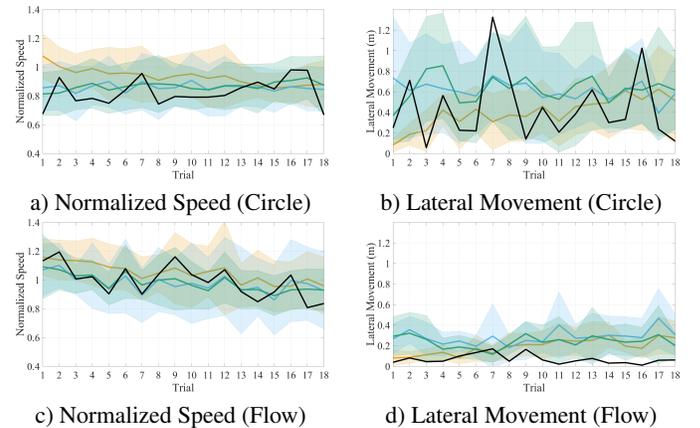


Fig. 6: Mean normalized speed and lateral movement over all participants when acting different characters under different conditions, sorted by trial number. While participants walked together in the real scenarios, and therefore the real data had no order, it is overlapped on the figure for illustrative purposes.

**Correction Location.** Figure 8 shows the position of local correction occurrences across responsive OMC agents, depending on their relative walking progress, for both scenarios. However, the two experimental scenarios exhibit different patterns. In the Circle scenario, most local corrections occur around the middle of the trajectory (30 to 50%), suggesting that encounters with users occur mostly near the center of the interaction area. Instead, a large number of local corrections in the Crossing Flow scenario occur very close to the start of trajectory, suggesting that numerous OMC agents trigger local corrections almost immediately as the trial begins (approximately 40% of OMC agents

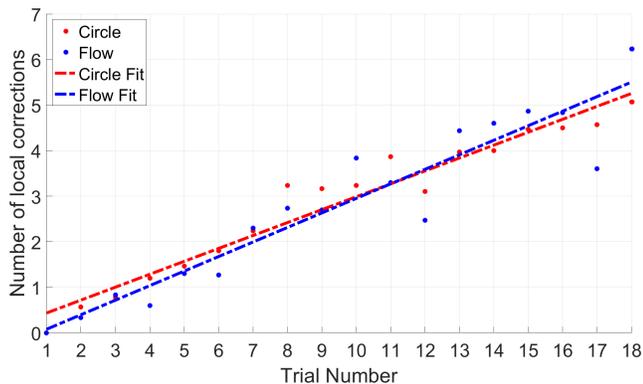


Fig. 7: Number of OMC agents whose trajectories are modified with respect to the trial number, given that  $i - 1$  OMC agents are present in trial  $i$ . Linear regressions show that participants triggered local corrections on approximately 30% of the OMC agents in each trial, in both scenarios. Results are averaged over 4R-S and 4R-C.

Table 1: Average number of collisions between participants and OMC agents across trials and participants (top), and total number of collisions between all individuals in the final generated crowd, averaged over participants (bottom).

Scenario	2R (OMC)	3R-S	3R-C	4R-S	4R-C
<b>Number of collisions with OMC agents (mean <math>\pm</math> SD over participants and trials)</b>					
Circle	1.10 $\pm$ 0.85	1.30 $\pm$ 1.07	1.10 $\pm$ 0.96	0.50 $\pm$ 0.47	0.47 $\pm$ 0.43
Crossing Flow	0.23 $\pm$ 0.19	0.37 $\pm$ 0.30	0.30 $\pm$ 0.19	0.18 $\pm$ 0.15	0.17 $\pm$ 0.17
<b>Number of collisions in the final generated crowd (mean <math>\pm</math> SD over participants)</b>					
Circle	13.2 $\pm$ 11.9	14.5 $\pm$ 14.9	11.5 $\pm$ 11.2	16.6 $\pm$ 10.1	12.7 $\pm$ 9.2
Crossing Flow	1.9 $\pm$ 1.7	3.6 $\pm$ 3.8	2.7 $\pm$ 2.5	6.5 $\pm$ 3.1	10.1 $\pm$ 4.2

become reactive in the first 10% of their performed trajectory). This is potentially caused by the close initial proximity of characters in the Crossing Flow scenario, which is further discussed in Section 7.2.

**Trajectory Dissimilarities.** As OMC agents become responsive when upcoming collisions with users are detected (4R-X), we measured after each trial the difference between each OMC agent and its original user-recorded trajectory using DTW. Figure 9 shows these differences averaged over participants for each OMC agent and trial, using a matrix-based visualization that reveals a gradual amplification of this difference as the recording process progresses. In addition, the maximum trajectory difference in the Circle scenario is approximately twice that of the Crossing Flow scenario.

## 7 DISCUSSION

In this section, we first discuss the results of our analyses of the effects of contextual crowds, as well as of local correction strategies. Then, we further discuss the limitations of our work, and highlight possible directions for future work.

### 7.1 Effect of Contextual Crowds

We first hypothesized (**H1**) that we could mitigate the bias caused by the sequential generation of trajectories inherent to the Record-and-Replace (2R) paradigm by initializing scenarios with contextual crowds, while simultaneously generating trajectories exhibiting characteristics more similar to real data. The results presented in Section 6.1 validate **H1**, despite some differences. They demonstrated that trajectories generated with our Replace-Record-Replace (3R) paradigm are more similar to real trajectories in terms of speed (both scenarios) and lateral deviations (Circle scenario) than those generated using a simpler 2R paradigm. However, other characteristics, such as local density and distance to the nearest character, were similar between all the conditions.

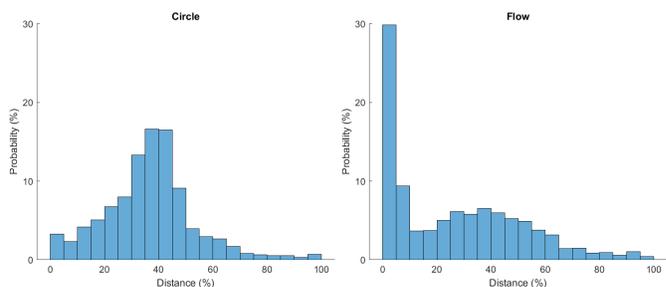


Fig. 8: Histograms of the position at which local corrections occur for responsive OMC agents, across the 4R-S and 4R-C conditions. In the Circle scenario (left), most local corrections occur around 30% to 50% of the agents' trajectories, suggesting that most local corrections occur in the crossing area. In the Crossing Flow scenario, nearly half of the local corrections occur at the beginning of the agents' trajectories (first 10%), suggesting that local corrections occur mostly with the agents of the same flow than the participants.

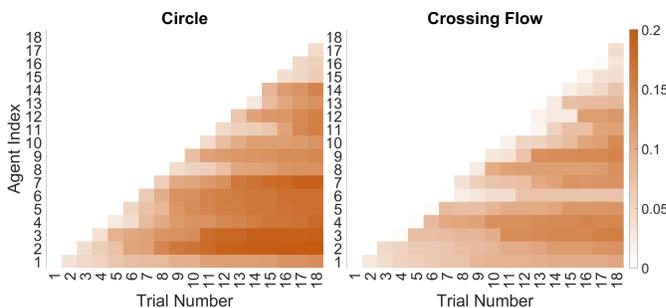


Fig. 9: Average trajectory differences (computed using DTW) compared to each OMC agent's original trajectory. The color scale represents the error in meters.

While some of the characteristics appear to be more similar to real data for the 2R condition, in particular for lateral movements in the Crossing Flow scenario, our results suggest that these similarities may be caused by biases typically observed when relying on the 2R paradigm. Figure 6 illustrates such biases: in early trials, when few characters are present in the scene, walking speed is typically high and lateral movement is low, as users rarely interact with other characters. As the process proceeds, the walking speed of the generated characters decreases, while lateral movement increases significantly. While the average lateral movement in the 2R paradigm is visually more similar to real data, this higher similarity appears to be due to extremely low deviations in the initial trials, biasing the results. In comparison, both walking speed and lateral movement are more consistent throughout the generation process using 3R paradigms. These results suggest that contextual crowds enable us to maintain constant user interactions throughout the generation process, therefore ensuring that participants experience pedestrian interactions that are more similar to those that can be experienced in real life.

Let us also emphasize that previous research has already demonstrated an increase in users' avoidance distance in Virtual Reality [27, 53], compared to real situations. This raises the question of how to evaluate the characteristics of crowd datasets generated in VR. Ideally, the evaluation should be based on a comparison of our 3R/4R approaches with  $n$  simultaneously immersed participants to jointly record a  $n$ -trajectory dataset, which is technically difficult. Nevertheless, a very interesting related question is to search for the number  $m$  of simultaneously immersed participants to record such an  $n$ -trajectory dataset in a 3R manner that maximizes the dataset quality tradeoff between efficiency ( $m \ll n$ ) and realism ( $m \approx n$ ).

Second, we hypothesized that the type of contextual crowd, whether it is created by simulation or reanimation from existing data, would

not significantly affect the results (**H2**). This is important because the scarcity of crowd datasets would make the use of XR-C paradigms difficult in practice, and defeat the purpose of capturing complex crowd datasets with only a limited number of people. Indeed, our results show that similar results are obtained between the 3R-C and 3R-S versions we studied, as well as between 4R-C and 4R-S: no significant difference was found between these pairs in the Circle scenario, and only slight differences were found in the Crossing Flow scenario, where participants showed smaller lateral movements and slower speeds between captured than simulated agents. Our explanation is that participants took advantage of the greater adaptation capabilities of the simulated agents to adjust their strategy accordingly. These differences, however, remain subtle. Our general conclusion about **H2** is that XR-S paradigms are more relevant in practice because the simulated contextual crowds offer greater flexibility. An interesting direction for future work would be to explore more deeply how simulation parameters can actually affect participants' self-behaviors. For example, a contextual crowd of hurried agents would probably influence a participant differently than a crowd of lazy wandering agents.

## 7.2 Effects of the Local Corrections

As participants are likely to experience unavoidable collisions when faced with non-responsive crowds generated using 2R, our last hypothesis was that providing recorded agents with responsive capabilities would reduce the number of collisions experienced by users, while introducing some differences in the recorded trajectories (**H3**).

First, the results presented in Table 1 demonstrate that using responsive agents indeed reduces the number of collisions experienced by users by approximately 50%, which partially validates **H3**. Interestingly, the location of these collisions is influenced by the scenario, as can be seen in Figure 8. In the Circle scenario, most of the collisions occurred around the middle of the trajectory, i.e., the virtual agents started to avoid users more towards the center of the interaction area where they all converged. In comparison, most of the collisions in the Crossing Flow scenario occurred early while recording new movements, which suggests that responsive capacities were activated for agents initially close to the user, and probably among the same walking group. Our explanation is that, as characters started very close in the Crossing Flow scenario, collision avoidance was activated from the start of the trial. This suggests that the parameters chosen for local corrections (distance of 1 m to activate the Responsive component), might have been too conservative for this scenario. This raises the question of whether responsive simulation should be tailored to each scenario, vs. using fixed parameters, which is a question left open for future studies.

In addition, results illustrated in Figure 9 show increasing differences between user-recorded and responsive-agent trajectories. This indicates that local corrections degrade progressively the similarity to the original user-generated trajectories, which fully validates **H3**. At the moment, our understanding of what actually happens when local corrections are triggered is still limited, which is something that we plan to further study in the future. In particular, it would be interesting to more finely quantify such differences, potentially relying on existing trajectory metrics (e.g. [16]), or by including more body-movement characteristics, which are, to our knowledge, currently rarely accounted for in the literature for such questions.

As combining collision avoidance approaches with recorded agent trajectories is still a challenging problem, responsive agents sometimes displayed unrealistic behaviors compared to what we would expect from a real person, such as sudden stops instead of more natural reactions (e.g., slight detours or side-steps). Such behaviors impacted the final trajectories, as well as the studied characteristics. For instance, as responsive agents came to a halt to give priority to participants, thereby reducing the number of potential collisions, this resulted in delays relative to the agents' original trajectory. While the agents proceeded to catch up with their recorded trajectory at their maximum permitted speed, this led to increased average speed in the 4R-X. Simultaneously, only the vicinity of participants activated local corrections (towards both participants and other agents), which were deactivated when agents

were catching-up with the recording trajectory (if participants were sufficiently far). This led to an increased number of collisions in the final dataset, as can be seen in Table 1. However, as we believe that solving such local collisions is important for interactivity in such a VR-based crowd motion capture context, it would be interesting to explore whether more complex simulation models would provide better local corrections in responsive OMC agents, or to explore the benefits of using post-processing local corrections. Overall, such observations raise novel challenging questions about the general evaluation of VR-based crowd motion capture datasets, their comparison to baseline examples, and how to most appropriately use or adapt to their corresponding real contexts.

## 8 CONCLUSION

In this work, we proposed and compared several new solutions for recording increasingly complex and realistic crowd datasets using VR. To this end, we proposed a new paradigm in which a participant, who initially evolves in a contextual crowd of autonomous agents, gradually replaces each of them to build a complete dataset from their own locomotion data. We have proposed and evaluated different versions of this paradigm (3R-C, 3R-S, 4R-C, 4R-S). Combining flexibility and fidelity, the 3R-S paradigm seems, in the light of our results, the most relevant one to implement and further explore.

Of course, our study has limitations, which we discussed in Section 7. In particular, we are not yet in the ideal configuration to finely evaluate the artifacts introduced by the iterative recording of individual behaviors, which would require comparison with simultaneous VR recordings of multiple participants. This represents technical difficulties that led us to prefer comparisons with real-world datasets. We also limited our evaluation metrics to recorded global trajectories, whereas future evaluations could include more complex details, such as full-body motion or gaze, which are relevant in the context of studying collective movements.

However, our results are promising, and already allow us to recommend the use of the 3R-S paradigm in practice, which opens up new research directions. One of our most important and immediate follow-ups is to further explore the impact that a contextual crowd might have depending on the simulated context. For example, with the same starting instructions and task, could we induce different behaviors, and therefore different datasets, by starting the recording process with different contextual crowds, e.g., a scared running crowd or a more relaxed one? Such an approach would facilitate the study of the propagation of certain behavioral traits in collective movements, which is extremely challenging to achieve in real situations.

More generally, we believe that the paradigm we are proposing for recording crowd data can make it possible to make significant progress in understanding the coupling that exists between individuals and crowds. Until now, all experimental studies have focused on understanding how individual behavior and local social interactions between individuals can give rise to specific collective patterns of behavior. But conversely, the influence of collective movement on local interactions is very difficult to study. Following the example given in the paragraph above, the solution proposed in this paper opens up new opportunities to control situational patterns in a precise and repeatable way to enable new types of crowd experiments.

## ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860768 (CLIFE project).

## REFERENCES

- [1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 961–971, 2016. 2
- [2] C. Appert-Rolland, J. Pettré, A.-H. Olivier, W. H. Warren, A. Duigou-Majumdar, E. Pinsard, and A. Nicolas. Experimental study of collective pedestrian dynamics. *Collective Dynamics*, 2018. 2
- [3] J. P. Araujo, J. Li, K. Vetrivel, R. G. Agarwal, D. E. Gopinath, J. Wu, A. Clegg, and C. K. Liu. Circle: Capture in rich contextual environments. *ArXiv*, abs/2303.17912, 2023. 1, 2
- [4] C. Armbrüster, M. Wolter, T. Kuhlen, W. Spijkers, and B. Fimm. Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extraperpersonal Space. *CyberPsychology & Behavior*, 11(1):9–15, Feb. 2008. doi: 10.1089/cpb.2007.9935 2
- [5] J. N. Bailenson, J. Blasovich, A. C. Beall, and J. M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments*, 10(6):583–598, 2001. 2
- [6] T. Banton, J. Stefanucci, F. Durgin, A. Fass, and D. Proffitt. The perception of walking speed in a virtual environment. *Presence*, 14(4):394–406, 2005. 2
- [7] F. Berton, F. Grzeskowiak, A. Bonneau, A. Jovane, M. Aggravi, L. Hoyet, A.-H. Olivier, C. Pacchierotti, and J. Pettré. Crowd navigation in vr: Exploring haptic rendering of collisions. *IEEE Transactions on Visualization and Computer Graphics*, 28:2589–2601, 2020. 2
- [8] F. Berton, L. Hoyet, A.-H. Olivier, J. Bruneau, O. Le Meur, and J. Pettre. Eye-gaze activity in crowds: Impact of virtual reality and density. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 322–331, 2020. doi: 10.1109/VR46266.2020.00052 1, 2
- [9] F. Berton, A.-H. Olivier, J. Bruneau, L. Hoyet, and J. Pettre. 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)*, pp. 717–725, 2019. doi: 10.1109/VR.2019.8798204 2, 4, 5
- [10] J. Bruneau, A.-H. Olivier, and J. Pettre. Going through, going around: A study on individual avoidance of groups. *IEEE transactions on visualization and computer graphics*, 21(4):520–528, 2015. 1, 2
- [11] D. Bršćić, T. Kanda, T. Ikeda, and T. Miyashita. Person tracking in large public spaces using 3-d range sensors. *IEEE Transactions on Human-Machine Systems*, 43(6):522–534, 2013. doi: 10.1109/THMS.2013.2283945 2
- [12] S. Cao, A. Seyfried, J. Zhang, S. Holl, and W. Song. Fundamental diagrams for multidirectional pedestrian flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(3):033404, 2017. 2
- [13] P. Charalambous and Y. Chrysanthou. The pag crowd: A graph based approach for efficient data-driven crowd simulation. In *Computer Graphics Forum*, vol. 33, pp. 95–108. Wiley Online Library, 2014. 2
- [14] D. Conigliaro, P. Rota, F. Setti, C. Bassetti, N. Conci, N. Sebe, and M. Cristani. The s-hock dataset: Analyzing crowds at the stadium. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 2
- [15] B. C. Daniel, R. Marques, L. Hoyet, J. Pettré, and J. Blat. A perceptually-validated metric for crowd trajectory quality evaluation. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 4:1 – 18, 2021. 2
- [16] B. C. Daniel, R. Marques, L. Hoyet, J. Pettré, and J. Blat. A perceptually-validated metric for crowd trajectory quality evaluation. *Proc. ACM Comput. Graph. Interact. Tech.*, 4(3), sep 2021. doi: 10.1145/3480136 4, 9
- [17] 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 Real.*, 23(1):19–32, mar 2019. doi: 10.1007/s10055-018-0365-0 2
- [18] I. Echeverría-Huarte, A. Garcimartín, R. C. Hidalgo, C. Martín-Gómez, and I. Zuriguel. Estimating density limits for walking pedestrians keeping a safe interpersonal distancing. *Scientific Reports*, 11, 2021. 2
- [19] T. Ezaki, K. Ohtsuka, M. Chraibi, M. Boltes, D. Yanagisawa, A. Seyfried, A. Schadschneider, and K. Nishinari. Inflow process of pedestrians to a confined space. *Collective Dynamics*, 1:1–18, 2016. doi: 10.17815/CD.2016.4 2
- [20] S. Feldmann and J. Adrian. Forward propagation of a push through a row of people. *Safety Science*, 164:106173, 2023. doi: 10.1016/j.ssci.2023.106173 2
- [21] C. Feliciani, H. Murakami, K. Shimura, and K. Nishinari. Efficiently informing crowds – experiments and simulations on route choice and decision making in pedestrian crowds with wheelchair users. *Transportation Research Part C-emerging Technologies*, 114:484–503, 2020. 2
- [22] Y. Feng, D. C. Duijves, W. Daamen, and S. P. Hoogendoorn. Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment*, 2021. 1, 2
- [23] 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–es, 2007. 2
- [24] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny. Statistical parametric mapping. *Academic Press*, 2007. 5
- [25] W. Garcia, S. Mendez, B. Fray, and A. Nicolas. Model-based assessment of the risks of viral transmission in non-confined crowds. *Safety Science*, 144:105453, 2021. doi: 10.1016/j.ssci.2021.105453 2
- [26] P. Georg, J. Schumann, S. Holl, and A. Hofmann. The influence of wheelchair users on movement in a bottleneck and a corridor. *Journal of Advanced Transportation*, 2019. 2
- [27] 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. 2, 8
- [28] D. Helbing, L. Buzna, A. Johansson, and T. Werner. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation science*, 39(1):1–24, 2005. 2
- [29] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51:4282–4286, May 1995. doi: 10.1103/PhysRevE.51.4282 2
- [30] R. S. Hessels, J. S. Benjamins, D. C. Niehorster, A. van Doorn, J. J. Koenderink, G. A. Holleman, Y. de Kloe, N. V. Valtakari, S. van Hal, and I. T. C. Hooge. Eye contact avoidance in crowds: A large wearable eye-tracking study. *Attention, Perception & Psychophysics*, 84:2623 – 2640, 2022. 2
- [31] R. S. Hessels, A. van Doorn, J. S. Benjamins, G. A. Holleman, and I. T. C. Hooge. Task-related gaze control in human crowd navigation. *Attention, Perception & Psychophysics*, 82:2482 – 2501, 2020. 2
- [32] I. Karamouzas, N. Sohre, R. Narain, and S. J. Guy. Implicit crowds. *ACM Transactions on Graphics (TOG)*, 36:1 – 13, 2017. 4
- [33] M. Kinateder and W. H. Warren. Exit choice during evacuation is influenced by both the size and proportion of the egressing crowd. *Physica A*, 569, 2021. 2
- [34] A. Kwiatkowski, V. Kalogeiton, J. Pettré, and M.-P. Cani. Understanding reinforcement learned crowds. *Computers & Graphics*, 110:28–37, 2023. doi: 10.1016/j.cag.2022.11.007 4
- [35] S. Lemercier, A. Jelic, R. Kulpa, J. Hua, J. Fehrenbach, P. Degond, C. Appert-Rolland, S. Donikian, and J. Pettré. Realistic following behaviors for crowd simulation. *COMPUTER GRAPHICS FORUM*, 31:489–498, 2012. doi: 10.1111/j.1467-8659.2012.03028.x 2
- [36] Y. Liu, R. Cadei, J. Schweizer, S. Bahmani, and A. Alahi. Towards robust and adaptive motion forecasting: A causal representation perspective. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17060–17071, 2022. doi: 10.1109/CVPR52688.2022.01657 2
- [37] 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. 2
- [38] S. D. Lynch, J. Pettré, J. Bruneau, R. Kulpa, A. Crétual, and A.-H. Olivier. Effect of virtual human gaze behaviour during an orthogonal collision avoidance walking task. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 136–142, 2018. 2
- [39] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz. Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences*, 276(1668):2755–2762, 2009. 2
- [40] M. Moussaïd, M. Kapadia, T. Thrash, R. W. Sumner, M. Gross, D. Helbing, and C. Hölscher. Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of The Royal Society Interface*, 13(122):20160414, 2016. 2
- [41] P. Mullick, S. Fontaine, C. Appert-Rolland, A.-H. Olivier, W. H. Warren, and J. Pettré. Analysis of emergent patterns in crossing flows of pedestrians reveals an invariant of ‘stripe’ formation in human data. *PLoS computational biology*, 18(6):e1010210, 2022. 2, 4
- [42] M. Nelson, A. Koiliias, S. Gubbi, and C. Mousas. Within a virtual crowd:

- Exploring human movement behavior during immersive virtual crowd interaction. In *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI '19*. Association for Computing Machinery, New York, NY, USA, 2019. doi: [10.1145/3359997.3365709](https://doi.org/10.1145/3359997.3365709) 2
- [43] A. Nicolas, M. Kuperman, S. Ibañez, S. Bouzat, and C. Appert-Rolland. Mechanical response of dense pedestrian crowds to the crossing of intruders. *Scientific reports*, 9(1):105, 2019. 5
- [44] A.-H. Olivier, J. Bruneau, G. Cirio, and J. Pettré. A virtual reality platform to study crowd behaviors. *Transportation Research Procedia*, 2:114–122, 2014. The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands. doi: [10.1016/j.trpro.2014.09.015](https://doi.org/10.1016/j.trpro.2014.09.015) 1
- [45] J. Ondřej, J. Pettré, A.-H. Olivier, and S. Donikian. A synthetic-vision based steering approach for crowd simulation. *ACM Transactions on Graphics (TOG)*, 29(4):1–9, 2010. 3
- [46] Y. Patotskaya, L. Hoyet, A.-H. Olivier, J. Pettré, and K. Zibrek. Avoiding virtual humans in a constrained environment: Exploration of novel behavioural measures. *Comput. Graph.*, 110:162–172, 2023. 2
- [47] 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*, pp. 189–198, 2009. 2
- [48] I. Podkosova and H. Kaufmann. Mutual collision avoidance during walking in real and collaborative virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D '18*, 2018. doi: [10.1145/3190834.3190845](https://doi.org/10.1145/3190834.3190845) 1
- [49] P. Raimbaud, A. Jovane, K. Zibrek, C. Pacchierotti, M. Christie, L. Hoyet, J. Pettré, and A.-H. Olivier. The stare-in-the-crowd effect when navigating a crowd in virtual reality. In *ACM Symposium on Applied Perception (SAP)*, 2023. 2
- [50] P. Raimbaud, A. Jovane, K. Zibrek, C. Pacchierotti, M. Christie, L. Hoyet, J. Pettré, and A.-H. Olivier. The stare-in-the-crowd effect in virtual reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 281–290, 2022. doi: [10.1109/VR51125.2022.00047](https://doi.org/10.1109/VR51125.2022.00047) 2
- [51] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments—a review. *ACM Computing Surveys (CSUR)*, 46(2):1–40, 2013. 2
- [52] A. Ríos and N. Pelechano. Follower behavior under stress in immersive vr. *Virtual Reality*, 24(4):683–694, 2020. 2
- [53] 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)*, pp. 75–80, 2015. doi: [10.1109/VR.2015.7223327](https://doi.org/10.1109/VR.2015.7223327) 2, 8
- [54] A. Seyfried, O. Passon, B. Steffen, M. Boltes, T. Rupperecht, and W. Klingsch. New insights into pedestrian flow through bottlenecks. *Transportation Science*, 43(3):395–406, 2009. 2
- [55] A. Seyfried and A. Schadschneider. Fundamental diagram and validation of crowd models. In *International Conference on Cellular Automata for Research and Industry*, 2008. 2
- [56] A. Sharma, V. Rai, M. Calvert, Z. Dai, Z. Guo, D. Boe, and E. Rombokas. A non-laboratory gait dataset of full body kinematics and egocentric vision. *Scientific Data*, 10, 2023. 2
- [57] X. Shi, Z. Ye, N. Shiwakoti, and O. Grembek. A state-of-the-art review on empirical data collection for external governed pedestrians complex movement. *Journal of Advanced Transportation*, 2018. 2
- [58] R. Sundararaman, C. De Almeida Braga, E. Marchand, and J. Pettré. Tracking pedestrian heads in dense crowd. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3865–3875, 2021. 2
- [59] H. Trivedi and C. Mousas. Human-virtual crowd interaction: Towards understanding the effects of crowd avoidance proximity in an immersive virtual environment. *Computer Animation and Virtual Worlds*, p. e2169, 2023. 2
- [60] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, pp. 3–19. Springer, 2011. 2, 3
- [61] J. P. van den Berg, M. C. Lin, and D. Manocha. Reciprocal velocity obstacles for real-time multi-agent navigation. *2008 IEEE International Conference on Robotics and Automation*, pp. 1928–1935, 2008. 2
- [62] W. van Toll, F. Grzeskowiak, A. L. Gandía, J. Amirian, F. Berton, J. Bruneau, B. C. Daniel, A. Jovane, and J. Pettré. Generalized microscopic crowd simulation using costs in velocity space. In *Symposium on Interactive 3D Graphics and Games*, pp. 1–9, 2020. 3
- [63] W. van Toll and J. Pettré. Algorithms for microscopic crowd simulation: Advancements in the 2010s. *Computer Graphics Forum*, 40, 2021. 2
- [64] D. Wolinski, S. J. Guy, A.-H. Olivier, M. C. Lin, D. Manocha, and J. Pettré. Parameter estimation and comparative evaluation of crowd simulations. *Computer Graphics Forum*, 33, 2014. 2, 4
- [65] P. Xu, J.-B. Hayet, and I. Karamouzas. Socialvae: Human trajectory prediction using timewise latents. In *European Conference on Computer Vision*, 2022. 2
- [66] B. Yersin, J. Maïm, J. Pettré, and D. Thalmann. Crowd patches: populating large-scale virtual environments for real-time applications. In *ACM Symposium on Interactive 3D Graphics and Games*, 2009. 2
- [67] T. Yin, L. Hoyet, M. Christie, M.-P. Cani, and J. Pettré. The one-man-crowd: Single user generation of crowd motions using virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2245–2255, 2022. doi: [10.1109/TVCG.2022.3150507](https://doi.org/10.1109/TVCG.2022.3150507) 1, 2, 3
- [68] J. Zhong, D. Li, Z. Huang, C. Lu, and W. Cai. Data-driven crowd modeling techniques: A survey. *ACM Trans. Model. Comput. Simul.*, 32(1), jan 2022. doi: [10.1145/3481299](https://doi.org/10.1145/3481299) 2
- [69] B. Zhou, X. Wang, and X. Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2871–2878, 2012. doi: [10.1109/CVPR.2012.6248013](https://doi.org/10.1109/CVPR.2012.6248013) 2