BIOMECHANICAL LOCOMOTION HETEROGENEITY IN SYNTHETIC CROWDS

MICHAEL BRANDON HAWORTH

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN DEPARTMENT OF ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

SEPTEMBER 2019

Abstract

Synthetic crowd simulation combines rule sets at different conceptual layers to represent the dynamic nature of crowds while adhering to basic principles of human steering, such as collision avoidance and goal completion. In this dissertation, I explore synthetic crowd simulation at the steering layer using a critical approach to define the central theme of the work, the impact of model representation and agent diversity in crowds. At the steering layer, simulated agents make regular decisions, or actions, related to steering which are often responsible for the emergent behaviours found in the macro-scale crowd. Because of this bottom-up impact of a steering model's defining rule-set, I postulate that biomechanics and diverse biomechanics may alter the outcomes of dynamic synthetic-crowds-based outcomes. This would mean that an assumption of normativity and/or homogeneity among simulated agents and their mobility would provide an inaccurate representation of a scenario. If these results are then used to make real world decisions, say via policy or design, then those populations not represented in the simulated scenario may experience a lack of representation in the actualization of those decisions.

A focused literature review shows that applications of both biomechanics and diverse locomotion representation at this layer of modelling are very narrow and often not present. I respond to the narrowness of this representation by addressing both biomechanics and heterogeneity separately. To address the question of performance and importance of locomotion biomechanics in crowd simulation, I use a large scale comparative approach. The industry standard synthetic crowd models are tested under a battery of benchmarks derived from prior work in comparative analysis of synthetic crowds as well as new scenarios derived from built environments. To address the question of the importance of heterogeneity in locomotion biomechanics, I define tiers of impact in the multi-agent crowds model at the steering layer–from the action space, to the agent space, to the crowds space. To this end, additional models and layers are developed to address the modelling and application of heterogeneous locomotion biomechanics in synthetic crowds. The results of both studies form a research arc which shows that the biomechanics in steering models provides important fidelity in several applications and that heterogeneity in the model of locomotion biomechanics directly impacts both qualitative and quantitative synthetic crowds outcomes. As well, systems, approaches, and pitfalls regarding the analysis of steering model and human mobility diversity are described.

Acknowledgements

I find it difficult to choose the words with which to acknowledge those closest to this dissertation and to my work. Their impact on my work, growth, and development is immeasurable.

My supervisor, advisor, and mentor Dr. Petros Faloutsos is responsible for altering the course of my life. I can not thank him enough for sticking with me (despite my own protests!). I did not understand the world I now have the opportunity and privilege to join. Because of his mentorship and the hand he extended to me, I will spend the rest of my career extending mine to others. There can be no words to accurately describe his impact, and I am ever grateful for that.

My master's co-supervisor and doctoral supervisory committee member, Dr. Melanie Baljko has in no small measure opened my mind and practice to what computer science can be. Before meeting Melanie my conceptualization of Computer Science was limited in scope and depth. Her work and critical practice brings life to the field. Her attention to detail, recapitulation of knowledge, and ability to see problem spaces from multiple standpoints has been invaluable in the development of my work and professional practice.

My supervisory committee members together, Dr. Petros Faloutsos, Dr. Melanie Baljko, and Dr. Matthew Kyan have been supportive and encouraging throughout my doctoral studies. Their mentorship and support have ensured the successful completion of this dissertation. Their time and input have immeasurable value, and they gave me so much of both. Additionally, I must extend my many thanks to my defence committee members, Dr. Burton Ma, Dr. Petros Faloutsos, Dr. Melanie Baljko, Dr. Sotirios Liaskos, and Dr. Karan Singh for seeing this dissertation through to the end and for their thoroughness in evaluation.

I have had the immense privilege to work with some of the best colleagues a person can ask for over the last several years. My colleague and mentor Dr. Mubbasir Kapadia has been so close to my work for so long. His drive, passion, and advice have been invaluable in my growth and development. As well to my friends and colleagues Dr. Glen Berseth and Muhmamad Usman who have enriched my work and life countless times– there will always be room for Vitamin P (pastry). I look forward to all of our long careers working together. My friends and colleagues through the University of Toronto, Dr. Yana Yunusova, Dr. Elaine Kearney and Dr. Aravind Namasivayam. You all inspired me to keep going when I was not so sure about my place in academia, and you played key roles in my development as a researcher.

To my love and partner, Ro, thank you for putting up with my ramblings about science and math. Your support and care through this journey has meant so much to me.

In conclusion, I wish to recognize the various sources of financial support which made this work possible: the Department of Electrical Engineering and Computer Science at York University; the NSERC Discovery Grant Program; the NSERC Create Grant Program (and the whole CreateDAV team); and the Intelligent Systems for Sustainable Urban Mobility (ISSUM) group.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Synthetic crowd simulation is an enabling technology across a broad spectrum of fields. By simulating crowds of synthetic humans, or agents, graphics and media artists can generate rich and complex scenes; event planners can predict the impacts of their event staging decisions; architects can build predictively, making buildings more usable and safe; and many more applications currently drive human-centric decisions while many more are possible in the future.

In this dissertation, I examine the fidelity of crowd agent modelling and the importance of diverse simulation on synthetic crowds-based outcomes. Much of synthetic crowd agent modelling work is focused on one of two bins of research production: methods which afford large scale (> 1000 agents) simulations; and methods that reproduce observable qualitative features of real crowds. In crowds simulation, and automata at large, it is often the low level rules–those that govern how an individual agent moves or makes decisions–that produce the emergent behaviours we see at the macro scale. Thus, when constructing these synthetic crowd agent models and simulations from base rules it is important to ask, "what are we missing?" Here I examine the specific question, "by looking at the large macro scale what are we overlooking at the micro scale?", I propose that what we overlook in synthetic crowd simulation at the lowest levels is our uniqueness, our diversity, as humans. While human bipedal locomotion has common features amongst normative bodies–often grouped by sex and age–this is not true of humans at large. The normative model is too simplistic to cover the range of humanity's interesting gaits and mobilities, and in the spaces of representation in media and safety-critical design, it is too important to continue to overlook.

1.1 Crowd Simulation

Crowd simulation refers to a system of generating, controlling, and directing multiple entities, actors, or agents which represent people using computational techniques. Typically, crowd simulation handles the movements of these agents on a large scale, in terms of the number of agents and/or the size and complexity of the environment. That is, given a basic description of an environment, crowd simulation provides the estimated movement of agents in the environment based on, but not limited to, one or more of agent initial conditions, goals, collision avoidance, inter-agent interactions, agent-environment interactions, individual behaviours, and group behaviours.

This system can be decomposed as a set of systems responsible for particular aspects of the simulation. While this decomposition has many subtleties confounding the division of labour, it serves as a starting point for a complicated and constantly growing field of techniques. This hierarchy can be seen in Figure 1.1 alongside Maslow's hierarchy of needs (Maslow, 1943)–a common analogy made in the field. At the highest level of abstraction, all entities exist as some physical simplification, such as an agent as a particle (position and radius) or a grid cell. These agents are driven by some goal which may be manually defined or may be derived from a behavioural model that merges a combination of perception, desire, emotion, and reactions to stimuli derived form the agent's context. This goal may be a destination, a leader to follow, a group to stay with, or any series/combination of these. Finding a path towards this goal, which may be in a very different area of the environment, is known as global navigation. Global navigation imitates the mental model one builds when trying to reach some place in the real world. Typically, the output of global navigation is a series of short-term goals at all major turning points, or corners, along the shortest path to the final goal. The local navigation required to reach these short-term goals while avoiding collisions with the environment and other agents is known as local steering and collision avoidance. Local steering is responsible for immediate movement, or steering, decisions and solves inter-agent and agent-environment interactions in some ideal way (e.g. reducing collisions and/or energy expenditure). The actualization of crowd simulation typically exists in two modes. The first being purely data or metrics that measure some aspect of the crowd. This is useful for crowd and environment analysis such as evacuation studies. The second is visual rendering and animation of the crowd. This approach is important for the film and

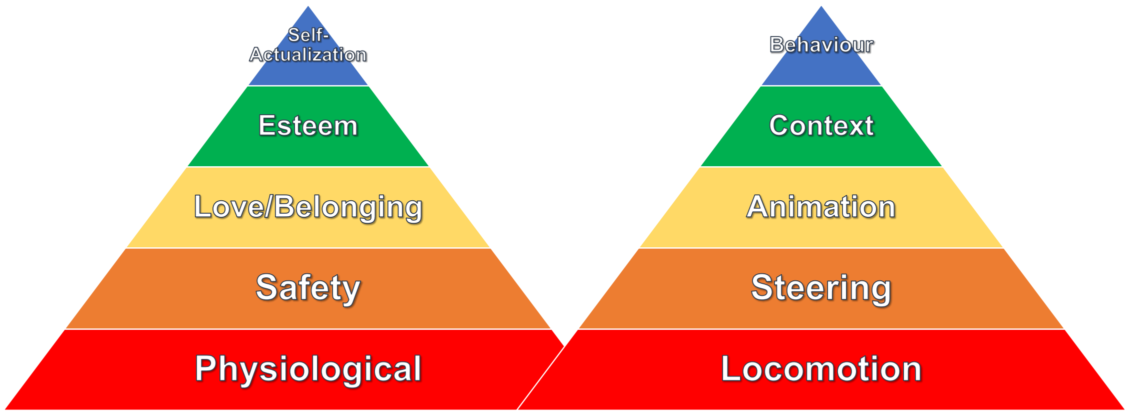


Figure 1.1. This figure represents an (right) abstract overview of a simplified model of the hierarchy of needs in synthetic crowds. The model is purposefully simplified and compared to (left) Maslow's hierarchy of needs. The model has several shortcomings, such as Behaviour and Context being fundamentally intertwined, Animation being driven directly by Locomotion, and Locomotion and Steering being part of the same biological system which makes up locomotion in humans. Often, in synthetic crowds, animation and locomotion are the same thing, in the sense that the actual movement of agents is driven by steering with locomotion (in this sense, the movement of the limbs and body) being estimated by animation.

animation industry, for understanding aggregate crowd behaviours, and for environment visualization such as when viewing architectural designs. Figure 1.2 shows a generalized overview of the crowd simulation solution. Two agents navigate an environment and must avoid colliding with the environment and each other on the path to their respective goals.

1.2 Locomotive Heterogeneity

The motivation of this dissertation stems from an apparent shortcoming in the literature regarding agent-based synthetic crowd models and human locomotion biomechanics.

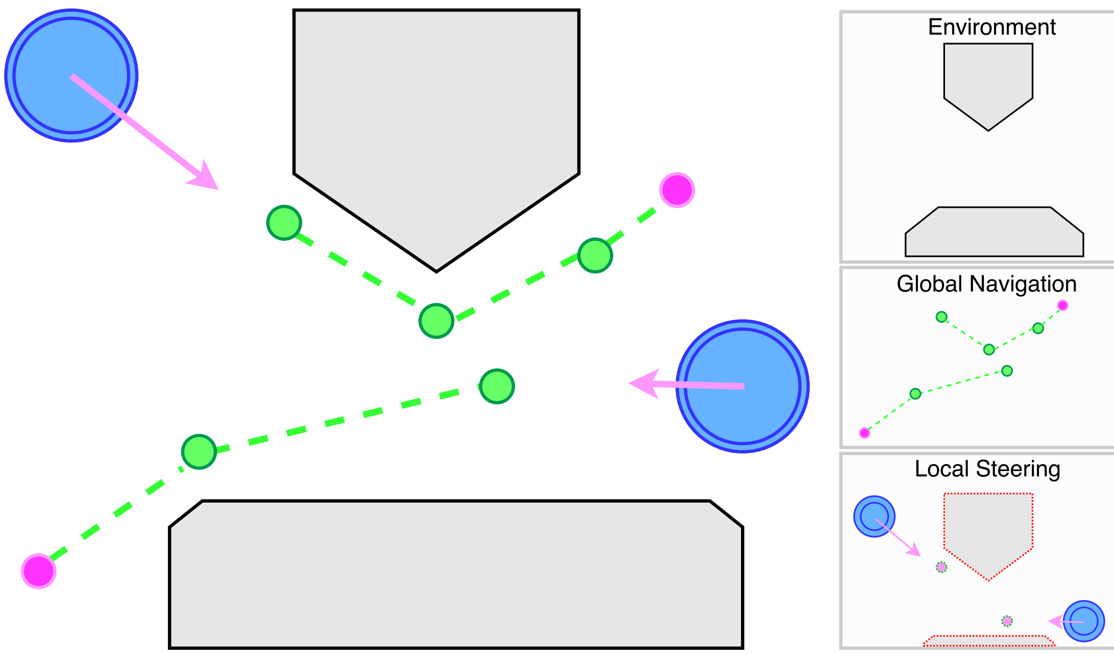


Figure 1.2. This figure represents an abstract overview of the steering level of crowd simulation. Two agents, the blue discs, navigate an environment, the grey shapes, and must avoid colliding with the environment and each other on the path to their respective goals. The green discs represent the short-term positional goals while the purple discs represent the final goal. The pink arrows represent the desired velocity of the agents.

Crowd simulators are historically derived from real data, physical analogies, behavioural assumptions, and/or computational geometry. In many cases, simulators rely heavily on these assumptions and generalizations, or simulators are developed under the guise of one problem space and then applied to human crowd modelling. For example, it is common practice to apply robotics steering methods to human crowd simulation.

Generally, heterogeneity is often possible, to varying degrees, at the steering level in most crowd models, but it is not explored or applied directly in evaluations or derivation of the model. In many cases, heterogeneity is inserted at the behavioural level in higher level synthetic crowds works with no regard to the diverse spectrum of human locomotion biomechanics. While homogeneous simulation is computationally beneficial, the impact of using homogeneous biomechanics in media, design, and policy goes beyond the processor and into the real world. Many models that lack representative outputs are applied in content creation tools from 3D modelling and animation suites to game engines. Similarly, crowd simulation tools which can and are used to guide built environment decisions and perform critical analysis often miss important human-centric considerations (Still, 2007). This discussion is taken further in both Chapters 3 & 5.

1.3 Crowds as an Open Loop System

In order to discuss crowd simulation at length, it is imperative to place the difficulty of creating and evaluating heterogeneous crowd simulators in a well understood framework. The literature on human movement simulation and animation in graphics is largely split between character control and crowd simulation. Historically, the approaches solve different problems. Character controllers resolve the motions of an individual character and their joints, while crowd simulators resolve the net movements of both individual and group motions.

These approaches can be framed as closed-loop and open-loop control systems respectively. The close-loop control system, here character control, has access to the output of the system as part of its input, in the form of a measured response. This affords complete control over the system via feedback-based control. On the other hand, the open-loop control system, here crowd simulators, are merely based on inputs. In crowd simulation, inputs are usually an individual agent's most immediate goal and their immediate perception of the environment and fellow agents. While agents receive feedback from the world in terms of modelled perception, the scope of emergent and chaotic effects are not captured by any measured response in the system. A useful analogy here is localized climate control via air conditioning systems versus attempted weather control in nature. An air conditioning system controls temperature and has a sensor which measures temperature, thus the input (a desired temperature) and the measured response (the current temperature) can be used to control local temperature directly, under the assumption that the system is closed. However, to measure only temperature in nature as a means to drive a weather control system remains intractable. The emergent and chaotic effects of natural weather systems can only be approximated and are not fully captured by any tractable measured response.

1.4 Dissertation Overview

This dissertation seeks to contribute the following: methods and metrics for comparative analysis of crowds over rich and in-depth benchmarks, models for investigating interesting gait patterns, and, perhaps most importantly, strong evidence for the inclusion of diversity in synthetic crowds and their applications. Towards these ends, the dissertation comprises the following chapters and their respective subjects:

Chapter 2 is the literature review and theoretical background pertaining to crowd simulation and synthetic crowd evaluation. This review supports the need, approach, and delivery of the works described in the dissertation.

Chapter 3 covers the methodology in terms of theoretical framework and an overview of the research arc in the dissertation. This chapter also covers a critical review or prior works in the applications of synthetic crowds in research and industry.

Chapter 4 outlines and presents the results from the first major study in the research arc–the importance of biomechanics in synthetic crowd simulations. This chapter is motivated by a critical review of crowd simulation evaluation approaches. A method for ground-truth free comparative analysis is proposed, including metrics and benchmarks. The chapter then covers the methods and results of five experiments which provide an in-depth look at steering model selections.

Chapter 5 outlines and presents the results from the second major study in the research arc–the importance of heterogeneity in locomotion biomechanics. Building from the previous chapter, the study is motivated by a critical review of the literature on the application synthetic crowds with a focus on safety-critical scenarios. The chapter then covers the methods and results of three experiments which provide an in-depth look at the impact of heterogeneity in locomotion biomechanics on the different level of steering.

Chapter 6 reviews the details of new methods used to generate the results in the dissertation. Additionally, the chapter covers speculation on new directions and critical analysis of the potential applications of the results.

Chapter 7 is the conclusion and aggregates and reviews the findings of all the studies in the dissertation. This is followed by the implications of these findings and their limitations. These sections motivate the discussion of future works in the areas the dissertation presents beyond what is discussed in Chapter 6.

Chapter 2

Background & Literature Review

This chapter is provided to explicate the literature and situate the problem space of the dissertation–heterogeneity of locomotion in biomechanical crowd simulation. To this end, this chapter focuses on the dissertation topics of the local steering and collision avoidance area of agent-based crowd simulation. Certain topics are covered from a high-level perspective and in-depth review and analysis is left to their respective chapters.

First, a review of Crowd Simulation models and their history from agent-based modelling to multi-agent systems is provided. Then the breadth of heterogeneity-capable agent-based crowd simulators is grouped by solution approaches and delineated. In particular, I focus on solutions which are primarily concerned with immediate locomotion decisions. Second, a deep review of synthetic crowds analysis methods lays the groundwork for understanding the difficulty of comparing and understanding the performance of a synthetic crowd model. This review also motivates the development of the comparative analysis method described in further detail in Chapter 4.

This review of literature is intended to provide the groundwork for later topics in the dissertation and is by no means a comprehensive review of all agent-based or multi-agent systems. The body of work in these areas are immense as they have proven useful in modelling large, complicated, and emergent systems beyond crowds. Nor is this review comprehensive of solutions in synthetic crowd simulations or physical character control. There are a multitude of exciting solutions in these spaces which are beyond the scope of this review.

2.1 Agent-based Crowd Simulation

This section provides an overview of agent-based particle models for local steering and collision avoidance in crowd simulation. Crowd simulation history necessarily starts with the history of early automata such as von Neumann's self-reproducing automata (Von Neumann, Burks, et al., 1966) and John Conway's Game of Life (Conway, 1970). These are examples of seminal cellular automata works that later spawned agent-based modelling. However, cellular automata are outside the scope of this report. The first documented use of agent-based modelling was on the dynamics of segregation (Schelling, 1971). From this point onward, agent-based modelling became a powerful way to model emergent behaviours in complex systems by encapsulating the functions (directives, reactions, rules) and knowledge (state, perception of surroundings) of an individual in a system as an agent in a simulation. The field and concept of agent-based modelling is a precursor for multi-agent systems, which are what agent-based crowd simulations are. In a multi-agent system, the interactions between intelligent agents and their environment are governed by simulation. A simulation encapsulates the representation of the world, the agents, and the passage of time. Generally speaking, the world is any environment and may be anything from graph paper grid to large scale digital 3D real-world like environments. The passage of time for simulations involves a steady update frequency or tracking of the change in time Δt. Depending on the simulation, the real-passage of time is most likely detached from simulation time. Simulations may require significantly more or less time to compute than what we perceive as real-time. The agents may be a cell within a world grid (cellular automata) or, closer to the scope of this paper, particles (or discs) whose physical traits (position and radius) are detached from the underlying world representation. The agents update their state in accordance with the simulation time.

The following models are all real-time capable agent-based particle simulations of local steering and collision avoidance. While all crowd simulators effectively handle the local steering and collision avoidance in some way, these models represent seminal works which focus on this particular portion of crowd simulation. Furthermore, all of these models can be parametrized, in sub-groups of the simulated crowd, to induce heterogeneity at the steering level.

2.1.1 Early Models

Early agent-based "Crowds" models were concerned with the animation of animal grouping in nature (herds, schools, flocks, etc). The first use of such procedural animation was in Eurythmy by Susan Amkraut and Michael Girard shown at the Film & Video Show of SIGGRAPH '85 (C. Reynolds, 1995). Soon after, this approach was formalized as a set of rules for combining desired velocities based on the intentions of the agents, known as Boids (C. W. Reynolds, 1987). These intentions included goal-seeking (to head towards a goal), separation (to avoid collision), alignment (to keep a heading with a group), and cohesion (to stay with a group). These simple rules work very well for replicating natural flocking and herding behaviour found in nature. The earliest application of this approach to humans defined inter-group relationships and emotional/sociological parameters (Musse & Thalmann, 1997). This model used line-line intersections for collision prediction, but it also included a multi-resolution approach to choosing whether collision avoidance was required at all. If the observer was far enough away from an agent, no collision avoidance technique was used. If the observer was close enough, a less expensive collision avoidance method was used that simply altered the speed of one of the agents. If the observer was close to the agent, the computationally more expensive technique of altering the angular velocity (turning) was utilized.

2.1.2 Velocity Models

The introduction of velocity obstacles (VO) for robotics VO (Fiorini & Shiller, 1998, 1993) provided a simple means of finding guaranteed collision free steering decisions. This method simply finds the abstract obstacle formed by the set of velocities for which two objects will collide at some point and then avoids choosing relative velocities within this "velocity obstacle" area as if it were an actual obstacle. In this way, VO methods are computational geometric methods that operate in velocity space. This is particularly import to robotics since local steering in robotics is primarily concerned with avoiding collisions in a guaranteed or provable way. Inter-robot collision can be costly in terms of equipment and repairs while robot-human collisions are considered unacceptable because of the potential for injury and danger to life.

This approach has since been extended several times to include reciprocal collision avoidance (RVO) (Van den Berg, Lin, & Manocha, 2008), optimal reciprocal avoidance for several agents (ORCA) (van den Berg, Guy, Lin, & Manocha, 2011), and generalized reciprocal collision avoidance (Wilkie, van den Berg, & Manocha, 2009). ORCA has become a common choice for collision avoidance in games and animation (Snape et al., 2012; Champandard, 2012). More recently, a biomechanically constrained hybrid motion synthesis model extends the RVO model such that full body character motions are synthesized with respect to steering and collision avoidance (Narang, Randhavane, Best, Shapiro, & Manocha, 2016).

2.1.3 Continuum Models

There is some analogous behaviour between crowd movement as particles and fluid dynamics. Several methods have been proposed to bring this analogy into practice with agent-based crowd simulation. This approach has the benefit of applying the same equations over all agents making the problem more tractable, that is this approach assumes the treatment of crowds as a continuum (Hughes, 2003). In particular, "crowd fluid" dynamics have been derived from gaskinetic equations (Helbing, 1992). Using these equations, macroscopic simulation can be performed if careful attention is paid to the differences in defining pressure and temperature to handle issues with density between actual fluids and crowd fluids. More recently, this has been applied to large scale crowds by converting crowds to potential fields and solving a density based system (Treuille, Cooper, & Popović, 2006).

Because of the treatment of the crowds as a whole, or continuum, heterogeneity is much more difficult to introduce. In fact, this approach often solved the problem of individual desired velocity by simplifying the equation into a continuum representation (like average velocity field) and leaning on plausibility of the simulation rather than an individual agent's ability to move independently. Additionally, it has been noted that while fluids can imitate many emergent crowd artefacts there are key differences which fluids do not account for well or at all, such as the aforementioned individuality of agents (like choice of exit or velocity), space filling phenomenon (crowds do not fill space uniformly) such as the corner bottleneck (crowds can bottleneck in open spaces). These methods are very good for large crowd simulations when the behaviour of an individual agent is not the focus of the simulation, such as in city rendering.

2.1.4 Force Models

In force-based crowd models the steering decision is based on a net of repulsion or attraction forces. This approach was first introduced to model the forces of social attraction and repulsion during locomotion in order to inform built environment designs (Helbing & Molnar, 1995). This model, Social Forces (SF), has been extended and applied to escape panic during evacuation in order to inform evacuation strategies (Helbing, Farkas, & Vicsek, 2000).

The models provide naturalistic behaviours and are computationally inexpensive, however they can suffer from severe oscillations, or unstable cyclic movement. While oscillations are natural in crowds, especially at bottlenecks, these can look unstable and unnatural due to particle sliding. Particle sliding is a larger general problem of particle-based crowd simulation. Since the particle model only represents position and radius, a particle-based steering decision may make unrealistic or biomechanically impossible movements. One approach that reduces oscillation is prediction, covered in more detail in Section 2.1.5. Models that layer the force-based model with line-disc intersection for prediction achieve excellent results (Karamouzas, Heil, Van Beek, & Overmars, 2009).

2.1.5 Predictive Models

Predictive models of local steering and collision avoidance predict imminent collisions and make early choices that effect the current steering decision. While collision avoidance is partially predicated on predictive-like decisions, this section focuses on models that explicitly predict future collisions and alter their current local steering decisions accordingly. This approach is particularly effective at reducing oscillations and step like or jagged movement.

The roots of predictive collision avoidance are very likely in naval steering and the need to avoid collisions with other boats in the near future in crowded ports. The roots of predictive collision for crowd simulation are in robotics. The works cited in Section 2.1.2 are all predictive. The basic VO approach avoids all potential future collisions with another agent by avoiding relative velocities within VO formed by their current velocities (Fiorini & Shiller, 1998, 1993). The extended forms of this method, in particular, RVO and ORCA both anticipate the reactive behaviour of other agents. In RVO, this is a achieved by assuming all agents make the same collision avoidance decisions then moving toward the end result of that set of decisions rather than stepping towards it (Van den Berg et al., 2008). This approach avoids the step-like or jagged patterns in the movement trajectories VOs may create. ORCA takes this assumption one step further and solves an optimization problem to find a collision free velocity of n-bodies in velocity space. The collision free velocity space is formed by the concatenation of half-planes induced by the velocities of the n interacting agents (van den Berg et al., 2011).

Perhaps the most straightforward predictive model is the linear one. In this case, the velocity of an agent is intersected with the disc of the combined radii of the two interacting agents using a simplified ray-disc intersection. The Social Forces model has been augmented with this method to greatly reduce oscillation artifacts (Karamouzas et al., 2009).

Methods modelling the space-time accurate representations of collisions are also possible. In space-time, where space is 2D, any given entity is a cone of all directions from the current position over time, and all other entities (from this point of view) are cylinders of the concatenated radii of both entities over time. This representation allows for a range of speeds and orientation changes to be represented as 3D cone-cylinder intersections. By segmenting these and optimally choosing the lowest cost set of choices for interacting agents, a steering decision can be made (Paris, Pettré, & Donikian, 2007).

Perception based prediction has also been utilized in the literature. In a synthetic-vision based steering algorithm, the optic flow of obstacles and interacting agents is used to predict future collisions (Ondřej, Pettré, Olivier, & Donikian, 2010). This paper relies on cognitive science which states that visual-stimuli extracted from optic flow can be used to control locomotion (Bruggeman, Zosh, & Warren, 2007).

Recent work in predictive modelling, WarpDriver, is a collision avoidance approach that uses probabilistic collision fields in its steering decision (Wolinski, Lin, & Pettré, 2016). Any given agent perceives other agents' current motion which is warped, or extended, in space-time to create a probabilistic field of potential future positions.

2.1.6 Planning Models

Directly related to predictive models, planning models produce a short-term path, or trajectory, to resolve imminent collisions. This approach, like predictive models is particularly well suited for avoiding oscillation and step-like or jagged movement. In addition to this, planning models also avoid particle sliding and unnatural movement. In particular, sliding disks can instantly change their forward direction, which is not natural for a biped. For people, complex interactions occur at doorways, including side-stepping for oncoming traffic, single-particle models may try to push through the doorway at the same time. This problem is partially eliminated with predictive models, but even these do not naturally handle side-stepping, which is inherently piecewise (composed of multiple steps).

One approach to planning is to model a discretized ego-centric perceptual field. This representation affords a space to plan on that is immediate to any given agent's perception, or local awareness planning. The ego-centric field is discretized such that nearer obstacles and agents are resolved at a finer granularity. By intersecting obstacles and performing linear prediction on this field, a short-term local plan can be formed (Kapadia, Singh, Hewlett, Reinman, & Faloutsos, 2012; Kapadia, Singh, Hewlett, & Faloutsos, 2009).

Multi-particle footstep-based models do not suffer from the same issues as single particle models. In addition to the aforementioned issues with single particle models, multi-particle models afford tighter packing, as the occupied space of the agent is represented by multiple smaller particles rather than one larger particle. As well, the foot particles can be used in space-time planning at the footstep level. Footstep-based planning is used by both the computer animation community and the robotics community, see Section 2.3. In crowd simulation, a space-time planning method for local steering and collision avoidance based on footstep placement affords a non-linear prediction of collisions, tightly packed crowds, piecewise stepping actions, and most importantly heterogeneity at the step level (Berseth, Kapadia, & Faloutsos, 2015; Singh, Kapadia, Reinman, & Faloutsos, 2011).

2.1.7 Data-Driven Models

Data driven models have the advantage of being empirically sound in recreating real-world scenarios. Most methods derive their data from video footage of aerial views of real crowds. The general approach is to extract the 2D trajectories of people interacting with other agents and obstacles and then use this data to inform local steering and collision avoidance.

Context-aware approaches use the local interaction features, such as environment features and the motion of the other agents, to resolve the closest matching action in the dataset (Lerner, Chrysanthou, & Lischinski, 2007). This approach can be extended to group behaviours as well (K. Lee, Choi, Hong, & Lee, 2007). Another more labour-intensive approach is to empirically understand these local interactions. Experiment-based modelling has been used to recreate single inter-agent interactions and large multi-agent interactions and then tune the model for real life data(Pettré, Ondrej, Olivier, Cretual, & Donikian, 2009). These context aware approaches may be automated in terms of clustering and selection using machine learning (Boatright, Kapadia, Shapira, & Badler, 2013).

Crowd simulation may also be seen as a two layer process, the group model that controls the formation of neighbours and the individualistic, or ego-centric, model that controls an individual's trajectory. By populating these models with learned data, morphing or interpolation can be used to extend them in time to form crowd simulations. This is accomplished, first, by generating an initial set of conditions, such as agent placements, and then iteratively advancing the trajectory model (Ju et al., 2010).

2.1.8 Hybrid Models

Hybrid models attempt to combine one or more of the aforementioned approaches to local steering and collision avoidance to produce a final steering decision. The Predictive Avoidance Model covered in Sections 2.1.4 & 2.1.5 is a good example of a basic hybrid model.

The High-Density Autonomous Crowds, or HiDAC system, combines psychological, and geometrical rules with social and physical forces. The model primarily handles collision avoidance in the same manner as social forces. A psychological layer provides information about context, such as being in a panicked stated, and affords context conditional rules such as stopping to avoid oscillations, waiting to form lines, and pushing in high density situations. The physical social force model also affords falling, which converts an agent to a static obstacle (Pelechano, Allbeck, & Badler, 2007). Many other features of HiDAC, such as bottleneck adaptive global path planning and information or state propagation, are outside the scope of this report.

Another approach to this method is to blend multiple steering decisions. A modular method for this approach is flexible such that blending can be achieved with a weighted sum, using heuristics to match the context of the agent, and fully computing all possible steering decisions then picking that which maximizes some fitness criteria (Singh, Kapadia, Hewlett, Reinman, & Faloutsos, 2011). This approach encapsulates the linear ray-disc prediction model from Section 2.1.5, a rule-based reaction phase, and a local short-term grid-based space-time planner. A state-machine keeps track of what steering circumstances the agent is in - predictive, space-time, or normal. However, this state-machine may be overridden by the necessity of reactive or crowd-based steering. This approach has, as well, several features that are outside of the scope of this report, such as perception and long or mid-term planning.

2.2 Synthetic Crowd Evaluation

The evaluation of crowd simulators is particularly difficult. An ideal scenario may be to directly compare a real life crowd or event with a simulation and measure the differences between the two. However, this sensible idea is fraught with issues. There are several scenarios important to the application of crowd simulation for which there are few real data recordings. This is especially the case in evacuation scenarios where footage or data may simply not exist and accurate re-enactment of the scenario is potentially dangerous and unethical. Furthermore, crowd simulators may have to perform under varying conditions in which the macroscopic, or aggregate, behaviour changes drastically. Imagine an evacuation scenario in which the initial conditions of the crowd involve leisurely or comfortable walking. An emergency evacuation may quickly evolve in to a high velocity and high density scenario. As agents evacuate, the behaviour may be to spread out in an open area away from the evacuated environment. The requirements of these three situations are very different, but they relate to one important application of crowd simulation. Thus, any singular approach to evaluation is not likely to capture the spectrum of macroscopic behaviours or the requirements for all applications.

Seminal work in this area is captured in *Chapter 6 - Validation of a computer model* of Dr. G. Keith Still's PhD thesis. This work separates validation of a crowd simulator into four categories: Component Testing, Functional Testing, Qualitative Validation, and Quantitative Validation (Still, 2000). The first two categories, Component and Functional Testing, are software engineering tasks and are outside the scope of this report. The latter two categories are concerned with Qualitative and Quantitative aspects of a crowd simulator. The following sections follow this direction and provide an overview of work in this area including Still's approaches. In this section, I focus on the breadth of the field pulling from key results and seminal methods to situate the content of the dissertation. The method presented Chapter 4 is directly relevant and a detailed overview of related methods can be found there.

2.2.1 Quantitative Evaluation of Simulations

Quantitative verification of crowd simulators is chiefly concerned with measuring some aspect of the crowd simulator. In the very best case scenario, this is a direct measurement of the difference between real and simulated data. However, since this is not always possible, several works have proposed other approaches to this problem.

Still's work proposed recreating several scenarios and techniques previously used in pedestrian traffic studies. In particular, this area focuses on recreating the fundamental diagram of specific scenarios. The fundamental diagram is a primary tool in transportation, traffic, and urban planning research. This diagram plots speed-density or flow-density relationships and highlights critical points. A critical point is the density at which speed or flow no longer increase and are very likely to rapidly decrease. Fundamental diagrams are useful in understanding when an environment no longer handles a crowd, or a crowd becomes dangerous or deadlocked. In Still's work, the fundamental diagrams between prior pedestrian experiments and his own, as well as others, and simulated experiments are directly compared. Still also recreates a previous real-world experiment of an evacuation scenario of a small room filled with people (Still, 2000). Similar work has comparatively evaluated crowd simulators by measuring the evacuation rate of rooms (Helbing et al., 2000). Follow-up work has shown that several guidelines and models differ in outcomes because of the method of measurement regarding fundamental diagrams (Seyfried et al., 2010). This work exemplifies the difficulty in normalizing evaluation, discussed later in Section 2.2.2. More recent work has used dense samplings of the fundamental diagram in bottleneck and group crossing evacuation scenarios to comparatively evaluate crowd simulators under the assumption that these scenarios capture fundamentally different behaviours (Haworth, Usman, Berseth, Kapadia, & Faloutsos, 2017, 2015).

Directly computable measures of crowd simulator performance can provide basic comparative feedback on crowd simulator performance. For example, crowd simulation is chiefly concerned with collision avoidance, thus the number of collisions can be measured directly (Shao & Terzopoulos, 2005). Similarly, several tasked-based measures such as path length, total kinetic energy, and total change in acceleration have been proposed (Kapadia, Singh, Allen, Reinman, & Faloutsos, 2009). However, these measures are scenario dependent. As well, the effort exerted by agents can be measured directly, and relates to the commonly found Principle of Least Effort in nature (Guy et al., 2010).

Comparatively evaluating crowd simulators is difficult but necessary to understand the relative performance and applicability of multiple simulators. Since many direct measurements may have unintuitive (to non-experts) or application specific meanings, approaches have been developed that combine benchmark scenarios of increasing difficulty and singular metrics derived from aggregates of independent measures (Singh, Kapadia, Faloutsos, & Reinman, 2009). This work defines a set of benchmarks that range from a single agent heading towards a goal, to two agents interacting on their paths to goals, to large high density scenarios. This work has been extended multiple times, and these works are covered in the following sections.

In scenarios where there exist a ground truth, it is to possible measure the performance of a crowd simulator with respect to the emulation of the real data. One approach is to try to recreate the initial conditions of the real data in the crowd simulator and measure the divergence between the two over time (Pettré et al., 2009). However, this method relies on accurate re-enactment and may not scale well to very large crowds. As well, the method does not take into account the non-deterministic nature of both real and simulated crowds. This is generally a problem with directly comparing real and simulated data. To solve this problem, one approach is to statistically model the error distribution between the two data sources (Guy et al., 2012). First, Bayesian inference is used to align the states of the simulation that correspond with the real data such that they can be directly compared. The comparison is made by measuring the entropy of the error distribution between the simulated and real data. This error distribution is estimated using a maximum likelihood estimator which is iteratively improved via expectation maximization. This approach scales well to larger crowds and handles the non-deterministic nature of crowd simulation well. Using functional Principal Component Analysis affords a multi-factored view of the sources of variability in synthetic pedestrian dynamics (Chraibi, Ensslen, Gottschalk, Saadi, & Seyfried, 2016). This method provides an efficient means of assessing the differences between experiment and simulation and thus can be used for model validation when ground truth is present.

2.2.2 Normalizing Quantitative Evaluation

The aforementioned evaluation methods all have minor shortcomings in terms of normalization. It is difficult, or even inaccurate, to directly compare two simulators on the same scenario without controlling for simulator parameters or other variables. Similarly, it is difficult, or inaccurate, to compare crowd simulators across scenarios without controlling for the differences in scenario.

One way of addressing cross scenario comparisons is to normalize a measure which is not entirely scenario dependent. The measurement of path lengths in SteerBench are scenario dependent and do not take into account the changes in velocity an agent experiences (Singh et al., 2009). To address this, a measure relating to the metabolic energy expenditure is utilized (Whittle, 2014). This measure directly accounts for changes in velocity by integrating velocity dependent energy expenditure over path of an agent during the simulation. By applying the Principle of Least Effort, the optimum effort for an agent in a scenario can be derived (Helbing et al., 2000). The actual effort of an agent in a scenario can then be normalized by the optimal effort. This new metric affords comparative evaluation of simulators across scenarios (Kapadia, Wang, Reinman, & Faloutsos, 2011). Similarly, the path measurements of SteerBench are rectified in ScenarioSpace by taking an egocentric reference agent approach normalizing with respect to optimal path length and time in a scenario (Kapadia, Wang, Singh, Reinman, & Faloutsos, 2011). However, these measurements are only valid in egocentric scenarios in which the reference agent is likely to interact with the scope of the environment.

Most importantly, comparative crowds analyses must consider the differing parameter space dimension between steering models. This is an issue both data-driven and synthetic comparative methods suffer from. While some of these parameters have intuitive direct effects, for example, the value of a *comfort zone* affects how close agents may come to each other, many are not intuitive yet may have large impact on the resultant behaviour or applicability of the model. Even when the meaning of the parameters is fairly intuitive, their aggregate effect, or their effect on the macroscopic behaviour of a large crowd, is not always easy to predict. Because of this, comparing crowd simulators directly and tuning their parameters is a difficult task. A crowd simulator, as is, may work better in particular conditions or be more suited for a particular task. Generally, all crowd simulators, when provided to the public, are given with default parameters. However, these default parameters may not be representative of the crowd simulators performance for all scenarios. Thus comparing any two simulators with default parameters raises the question of whether the comparison is fair or if the parameter settings themselves confound the results. To alleviate this, crowd simulator parameters may be optimized under the same conditions before comparison (Wolinski et al., 2014a, 2014b). This work finds the parameters of the crowd simulators to best fit the given data (real data or sketches) and then directly evaluates them.

2.2.3 Qualitative Verification of Simulations

Several important behaviours of real crowds appear at the macroscopic level and may occur in flux, or they may only exist in or during a small portion of the overall environment or simulation time. Behaviours such as vortices (vortex flow), lane forming (laminar flow), oscillations, deadlocks, and several other emergent behaviours ar difficult to quantify accurately or at all. While there are several rendered qualitative aspects of crowds such as upper body texture and accessories (McDonnell, Larkin, Hernández, Rudomin, & O'Sullivan, 2009; Maïm, Yersin, & Thalmann, 2009), this section focuses on qualitative aspects of emergent crowd behaviour and the tools to evaluate them. Still presented several qualitative methods in his thesis and relegated qualitative evaluation to informed expectations (Still, 2000). The expected macroscopic effects were: Edge Effects (edges of the crowd move faster than the centres); Finger Effects (lane forming or laminar flow in bidirectional crossing); Density Effects (crowd compression); Human Trail (movement regarding the principle of least effort). Edge Effects were evaluated by producing density and speed heatmaps over a hallway scenario. Finger Effects were evaluated by visually confirming the effect in a bi-directional scenario. Crowd Compression effects were validated by noting that the simulated crowd compresses when turning corners and in bottleneck scenarios. Finally, Human Trail effects were evaluated by observing crossing paths (shortest paths) for a goal square (goals in four corners), similar to human desire paths in parks.

Several approaches use application specific measures to quantify the performance of a crowd simulator. In virtual reality, the presence of the user is crucial to the success of the system. A comparative study was undertaken to measure the presence afforded to the user by different crowd simulators (Pelechano, Stocker, Allbeck, & Badler, 2008). In this study, the participants were given surveys immediately after completing tasks in a virtual cocktail party. The survey questions were designed to elicit the presence each participant felt in the company of the different crowd simulators (Social Forces, Rule-based, Cellular Automata, and HiDAC). Similarly, studies have utilized the perception of clones in crowd simulations to understand what contributes to heterogeneity, an important factor in the qualitative evaluation of heterogeneous crowds. This work used crowd agent clone identification tasks to understand what sort of behaviours or visual changes reduced the ability to spot clones (McDonnell, Larkin, Dobbyn, Collins, & O'Sullivan, 2008). Later work evaluated the spectators, or evaluators, ability to detect collisions in a simulation (Kulpa, Olivier, Ondřej, & Pettré, 2011).

Detection based qualitative evaluation moves the onus of viewing an expected behaviour from the evaluator to an algorithm. The SteerBug system is capable of detecting rule-based qualitative behaviours (Kapadia, Singh, Allen, et al., 2009). The strength of this system is the ability to detect micro to macroscopic behaviours and afford a simple ruled editing interface as well as a sketch interface for defining behaviours to detect. This makes the system extremely flexible. Similar works take a data driven approach, in that the system highlights simulated behaviours by matching them with real data (Lerner, Chrysanthou, Shamir, & Cohen-Or, 2010). The system can change contexts by utilizing different sets of real data to match against, effectively changing the evaluation.

Recent works have made use of the Level of Service (LoS) mapping to evaluate crowd simulators qualitatively by comparing the LoS flow-density and its qualitative understanding with the aggregate macroscopic crowd behaviours (Haworth, Usman, Berseth, Kapadia, & Faloutsos, 2017; Haworth et al., 2015). Fruin applied the concept of Levels of Service from transportation research to pedestrian traffic by mapping the quantitative values (flow rate or speed, and density) to a more qualitative one (levels or grades from A-F) (Fruin, 1971). Each successive level represents a decrease in service afforded by the scenario. Level A represents a low density free flowing environment without any path conflicts between people, and Level F represents a high density crowd with oscillations, pushing, tight packing, and deadlocked flow.

Tuning a crowd simulator for a particular behaviour or application, especially a high-dimensional model with unintuitive parameters, is also a difficult task. A specific crowd simulation method may be highly desirable for a particular application for several reasons but its behaviour may not match the desired outcomes. Given a pattern of behaviours as a performance criterion or a trade-off between performance objectives, parameter values of a steering algorithm may be selected that will produce a particular desired effect (Berseth, Kapadia, Haworth, & Faloutsos, 2014). This method also affords interactive blending for artistic control and simplified fine-tuning.

2.2.4 Scenario and Environment

The interactions of steering simulators and the environment are particularly useful. A steering simulator may feasibly be applied in any environment scenario. Thus, it is difficult to predictively evaluate simulators for scenarios not common in the literature or outside the scope of a particular application. To address this, work has been done to generate a representative scenario space that captures far-reaching features of different scenario designs (Kapadia, Wang, Singh, et al., 2011). This work samples the configuration space of scenarios, provides methods for statistically generating scenarios, and deriving subsets of the representative set ($10,000$ scenarios covering the spectrum of agent interactions and environment configurations). Steering simulators can be evaluated by computing coverage, the ratio of completed to total scenarios. The approximate converse of this work is computing the complexity of a scenario for crowd simulators (Berseth, Kapadia, & Faloutsos, 2013). This work affords the tuning and selective evaluation of crowd simulators, for example comparatively evaluating simulators on very difficult scenarios.

2.3 Character Control

This report is focused on biomechanical heterogeneity in crowd simulation, and, as such, it must touch on the rich area of character control, which has produced physically reactive, heterogeneous, biped locomotion at the single character level. This section provides an overview of seminal and recent approaches in digital character control. While there is an extensive and rich literature in robotics and biomechanics concerning biped locomotion, these works are outside the scope of this report.

There has been much success in physical and data driven models for heterogeneous biped locomotion of digital characters. In the following models it is possible to perturb, constrain, or parametrize joints, muscle, or neural model parameters to produce gait asymmetries, unique walking patterns, and falling. These solutions, while rich in their movement control and output, do not natively handle extrinsic factors found when planning in dynamic environments, agent-agent and agent-obstacle interactions, and high volume or high density simulations. In fact, the success of many of these models, especially earlier approaches, is highly dependent heavily constrained and controlled worlds.

The most successful physical character controllers have utilized at least one or a combination of optimization, neural networks, and data driven techniques. Early biped models recreated neural oscillators to produce walking patterns (Taga, Yamaguchi, & Shinizu, 1991). Neural oscillators are types or portions of Central Pattern Generators, biological neural networks on which isolated rhythmic patterns are predicated (Kuo, 2002). Later neural models focused on training neural networks by receiving joint or body sensor feedback as input and producing appropriate joint angles as output (Geng, Porr, & Wörgötter, 2006; Kun & Miller III, 1996; Miller III, 1994). It has been shown that this sort of walking behaviour can be evolved by using evolutionary optimization techniques on complex neural networks composed of simple neurons (Allen & Faloutsos, 2009a, 2009b).

Several models focused on first principles and simple controllers have been successful at producing life-like biped locomotion. A biped character's movement controller set can be composed using simple control strategies and feedback learning (Yin, Loken, & van de Panne, 2007; Faloutsos, Van de Panne, & Terzopoulos, 2001). However, generally, controllers must be carefully composed and tuned.

Of particular interest, is the integration of simplified biomechanical models such as the inverted pendulum model. This simple but effective model captures the movement of the centre of mass as a biped shifts weight on a single leg. This model provides a computationally low-cost means of solving for balance in physically controlled characters (Tsai, Lin, Cheng, Lee, & Lee, 2010; Kwon & Hodgins, 2010; Mordatch, De Lasa, & Hertzmann, 2010).

There have been several data driven models which derive motions from motion capture of real actors. This can be achieved in several ways: by applying joint torques derived from motion capture (Wrotek, Jenkins, & McGuire, 2006); optimizing captured motions for balance control and composing controllers (Sok, Kim, & Lee, 2007); modulating the reference motion capture (Y. Lee, Kim, & Lee, 2010); and adaptive controllers based on predictive models and motion capture data (Ge, Li, & Yang, 2012).

At a more microscopic level, direct simulation and optimization of muscle models has led to effective biped control. A flexible optimization based strategy affords the generation of locomotion for a variety of biped characters (Geijtenbeek, van de Panne, & van der Stappen, 2013). The value of this approach is the ability to model muscle or body dependent perturbations directly to generate heterogeneity or asymmetries in gait.

The most recent and advanced works in this area concern deep learning. In particular, the application of deep learning in layered, or hierarchical, controllers such that the biped may have some extrinsic characteristics found in crowd simulation, such as global planning in dynamic environments. Deep learning through hierarchical reinforcement learning avoids the need for prior knowledge about locomotion and higher level skills like path planning and environment interaction (Peng, Berseth, Yin, & van de Panne, 2017).

2.4 Conclusion

In this chapter, the groundwork is laid to begin an in depth discussion on analysis of and new approaches to biomechanical locomotion heterogeneity in agent-based synthetic crowd solutions and their applications. The history of the synthetic crowd led to the categorization and delineation of solutions in this particular area. Related approaches in the field of character control are also explored as the field is rich with heterogeneous physical (read biomechanical, or at least bio-inspired) solutions.

Chapter 3

Motivation & Methodology

This chapter attempts to concretely situate the *what?* and *why?* of this dissertation before providing an overview of the *how?* In doing so, I will attempt to place the work in frameworks that are both motivational to the work but also important in understanding its outcomes and impact.

This dissertation seeks to pursue and critically reflect on a specific issue in an area of research that is indicative of issues in applications of the work well beyond the scope of the dissertation. The research area is synthetic crowds and their applications in rendering and analytics. The specific issue is locomotion biomechanics at the steering level of crowd agent modelling.

Now that the *what?* has been covered here and in Chapters 1& 2, I will endeavour to situate the *why?*. First, to do so it is important to note a theoretical framework critical to motivating the problem space. This work is inspired and motivated by a feminist disabilities standpoint, and as such, this dissertation adheres strongly to the principles of the Social Model as opposed to the Medical Model. The Medical Model pathologizes illness, disability, and difference, while the Social Model centres the humanity and autonomy of the person or people of focus–placing the onus of such concepts as illness, disability, and difference on environment and the organization of society. In this way, these concepts are developed and reinforced by the institutionalization and regulation of bodies. Rosemary Garland-Thomson (Garland-Thomson, 2005) concisely and powerfully sums up this framing and its definition within these model frameworks:

> Feminism challenges the belief that femaleness is a natural form of physical and mental deficiency or constitutional unruliness. Feminist disability studies similarly questions our assumptions that disability is a flaw, lack, or excess. To do so, it defines disability broadly from a social rather than a medical perspective. Disability, it argues, is a cultural interpretation of human variation rather than an inherent inferiority, a pathology to cure, or an undesirable trait to eliminate. In other words, it finds disability's significance in interactions between bodies and their social and material environments.

Now that the motivation for the *what?* has been covered, the following sections attempt to impress the importance of the *why?* of this dissertation by examining the application space of synthetic crowd simulation, the focus on normative instead of inclusive frameworks and solutions, and how this affects outcomes in terms of their generality, usefulness, and real-world impact. The chapter ends by framing the approach taken to examine the problem space in this dissertation.

3.1 Applications of Synthetic Crowds

Crowd simulation has been applied in several areas of art, design, and evaluation. In particular, crowd simulation addresses the problem of scale in many fields. The costs required to produce large-scale scenes of crowds or evaluations of environments has been a primary driver of the application of crowd simulation in industry.

3.1.1 Computer Graphics and Animation

Early applications of crowd simulation in animation were cross-overs between disseminating technical achievements and works of art. The first use of flocking behaviours was in the animation Eurythmy (C. Reynolds, 1995). The Boids approach was the first formalized approach of flocking, schooling, and herding behaviours. The first animation using Boids was Stanley and Stella in: Breaking the Ice created by Craig Reynolds and Symbolics.

In film and animation, scenes of scale that are too costly to film with real actors or animate manually can be created with computer graphics using crowd simulations. Beyond the cost factor, using AI-driven crowd simulations allows the characters to react to their local situation and provides for emergent movement and behaviour. The first and perhaps most famous application of rich crowd simulations in film was the battle scenes in the Lord of the Rings series (MASSIVE, 2017). This software, MASSIVE, is also responsible for crowds in other films such as Ben-Hur, World War Z, John Carter, and Hugo. Another popular crowd simulation for use in the film industry is Golaem. This software is responsible for crowds in a diverse set of TV and films such as Chocolat, Florence Foster Jenkins, Race, Halo Wars 2, Pan, and Game of Thrones (Golaem, 2017). These tools all support animation and production pipelines and micro to macroscopic behaviour editing that includes context selective animations, AI, and heterogeneous movement details.

3.1.2 Analysis and Optimization

Crowd simulation reduces the need for human crowd studies in environment and event designs. Human-based (real participants) studies are very costly and require both an environment model and participants. Thus, there is a time and monetary cost motivation to using synthetic crowds as a source of dynamic analytics. In addition to these costs, several common crowded scenario cases for which environments must be tested, such as evacuations or escalated events, are unethical and extremely difficult to reproduce accurately or safely with real humans. In general, safety-critical scenarios, perhaps the most important to evaluate, are the most difficult to test because safety can not be guaranteed for participants, and often video evidence or actual events are not available, not released, or are extremely noisy due to the conditions of the event. In this space, real-world training, such as evacuation and safety response training, helps prepare people for such events. However, it is difficult to predict the real-world responses to such scenarios under a variety of conditions as they occur. Thus, crowd simulation allows us to safely sample this space and create predictive environments–environments which depend less on the training of their users and more on their design. In this way, the social model captures accurately the issues of inclusive environment design. Historically, design principles and methods place much of the onus on the environment's users, not on the environment or the policies which govern it. That is, we seek form over function, and when we do consider function, it is often from the perspective of intended use by ideal, *read: able-bodied*, users under ideal conditions, *read: average use*. This is then an issue for all environments from common-use to safety-critical scenario conditions. The space of tools and analyses for this work can be binned in to two areas, commercial and research. Seminal works from these areas are covered in the following sections.

3.1.2.1 Commercial

An industry has developed around the need to predictively evaluate architectural, urban, and event designs prior to investing in their construction or execution, as well as modifying, or fixing, already built environments. There are commercial tools which provide some crowd simulator-based analytics and visualization tools for understanding problems in designs. These range from large scale industry compliant crowd simulators for predictive scenario testing, such as Pedestrian Dynamics (INCONTROL Simulation Solutions, 2017), to fully integrated BIM compliant 3D environment editors such as MassMotion (Oasys Limited, 2017). This range includes: complex building simulators (Bentley, 2019; Oasys Limited, 2017; INCONTROL Simulation Solutions, 2017; SIMWALK, 2019); transportation focused simulation (PTV Group, 2019); evacuation focused simulation (Integrated Environment Solutions, 2019; TraffGo HT, 2019); retail focused simulation (CORE CROWD LLC, 2019); and film/TV focused simulation (MASSIVE, 2017; Golaem, 2017). These tools represent powerful engineering ventures and often use state-of-the-art simulators and analytics. However, they often rely on the use a single synthetic crowd steering model in their underlying animations and do not account for heterogeneous biomechanical locomotion in their evaluations. Chapter 4 explicates the issues with this panacea approach to choice in crowd model, especially in the analysis of buildings and scenarios.

3.1.2.2 Research

Historically, the application of crowd simulation in research has mainly focused on evacuation studies. This is mostly due the dangerous, dynamic, and very difficult to model behaviours that occur during panicked evacuation (Still, 2007). This section makes a brief overview of works that use crowd simulation to make environment design decisions. Work has been done to understand evacuation strategies such as herding, and grouping during evacuations (Helbing et al., 2000). Focus has been placed on evaluating environments and crowd dynamics using the placement of flow affecting obstacles to impact velocity or density in the crowd (Severiukhina, Voloshin, Lees, & Karbovskii, 2017; Helbing, Buzna, Johansson, & Werner, 2005). Pillar-like obstacles may be optimally placed in the path of bottlenecked crowds to reduce tangential momentum and increase crowd flow in evacuation scenarios (Berseth, Usman, Haworth, Kapadia, & Faloutsos, 2015; Jiang, Li, Shen, Yang, & Han, 2014). Recent work has shown that panel-like obstacles outperform pillar-like obstacles under varying initial crowd conditions and crowd densities (Zhao et al., 2017).

The concepts of Level of Service, introduced in Section 2.2.3, have been used to evaluate and optimize environments for varying crowd densities during evacuations (Haworth, Usman, Berseth, Kapadia, & Faloutsos, 2017; Haworth et al., 2015). This work in crowd density has been applied to speeding up the optimization of environments to facilitate crowd flow during evacuations by discretizing the search space and focusing on areas with high crowd flux/density (Haworth, Usman, Berseth, Khayatkhoei, et al., 2017; Haworth, 2016).

3.2 Normative Heterogeneity

This section explores the notion of homogeneity as a norm in crowd analytics and the use of proxies for heterogeneity. In Section 3.1.2.2, several papers that prescribe solutions to building features or concrete outcomes generated by synthetic crowds analytics are described. In these papers, there are echoes of a prevalent problem often overlooked. Susan Wendell (Wendell, 1996) captures this problem well:

> For instance, poor architectural planning creates physical obstacles for people who use wheelchairs, but also for people who cannot walk far or climb stairs, for people who cannot open doors, and for people who can do all of these things but only at the cost of pain or an expenditure of energy they can ill afford. Some of the same architectural flaws cause problems for pregnant women, parents with strollers, and young children.

In this way, misrepresentation, simplification, or lack of representation in synthetic crowd simulation which is then used in content creation, environment design, or policy development can be viewed as a form of erasure. While there may be many reasons for this approach, some of which are well-meaning, it is pertinent to take a critical approach when creating tools, methods, and procedures for informing the aforementioned applications. In the rest of this section, I explore these issues in my own work and identify a focus for the rest of the dissertation.

3.2.1 Prior Assumptions and Focus

Many of the synthetic crowd simulation based works, both in the commercial and research domains, make assumptions towards generating and probing tractable problem spaces. In this section, I explore what sort of assumptions have been made in works which I have co-authored and account for where they stem from. In many cases, by trying to control for confounds or reduce independent variables in analysis, works in this area simply overlook or overwrite representations of people.

My own work in this area began with these sorts of assumptions, so before analysing the work of others (in Chapter 5), I will examine my own. Specifically, I cover papers that make prescriptive findings about optimal environment layout for safety-critical scenarios, which rely heavily on the use of normative crowds.

In several papers I have co-authored, prescriptive environment designs and methods for deriving such designs have relied on mostly homogeneous and normative synthetic crowds (Haworth, Usman, Berseth, Khayatkhoei, et al., 2017; Haworth, Usman, Berseth, Kapadia, & Faloutsos, 2017; Haworth et al., 2016; Berseth, Usman, et al., 2015; Haworth et al., 2015). This approach has utility for reducing the independent variable set size, which can simplify the analysis and understanding of results. However, the assumption of the generality of the results by an external or non-expert reader could leave that reader overconfident and to overlook the use of these spaces by diverse users. While these papers are careful to note that these sorts of findings are not definitive and are limited in scope, it is possible for a reader, a non-expert, to apply outcomes to decision-making. In this way, preliminary, small scope, results may become part of a decision that affects those outside of the intended scope. In particular, the application of crowds can often be seen as overconfident. It has been noted that no single method could possible be panacea for the crowds problem (Still, 2000). As well, that often crowd simulators are presented as panacea with respect to modelling in that they "model the range of human behavior" (Still, 2007).

To go deeper in to this space it is important to focus and carefully define the heterogeneity in the scope of this dissertation. In particular, I wish to explore the lowest level of synthetic crowd simulation–the steering models. I want to understand the impact of heterogeneity in locomotion biomechanics. At the steering levels this translates in to how agent update decisions are made, in the real world it means how humans move at any given moment. For bipedal walkers without mobility aids, this means what motions the overall body makes in taking steps, stopping, avoiding collisions, and turning. This is the focus of this dissertation. However, this definition extends to those with varied mobilities including users of mobility aids, where steps are replaced with whatever means of locomotion those aids offer. For example, rather than planning discrete piecewise footstep plans, one may view, for example, a wheelchair as a differential drive controller. For this, you may consider continuous control signals in a differential drive controller model, creating piecewise control signal plans, or mapping the differential drive control to a simplified form, such as viewing the control plan as a piecewise parametrized curve. These latter models, those concerning mobility aids, are earmarked for future work as they are deeply important but outside the scope of this particular dissertation.

Finally, I restate that the focus of the dissertation is entirely within the realm of synthetic crowd simulation. Often, disability and health intersect with race, class, sex, and gender with respect to the forms of oppression people experience. These are important dimensions of the problem space the dissertation addresses–particularly in the application of simulations, and the inclusivity of teams making simulation-based decisions or creating simulation-based media. However, these dimensions are outside the scope of the dissertation. That is, addressing the forms the application of crowd simulation takes, requires several very different studies outside the scope of this area of study. This work, particularly critical collaborative interdisciplinary approaches, are earmarked as deeply important future works. However, the impacts of the various components in this problem space are actively studied in the many areas or disability studies, in particular the role of environment design in the disablement process and its impact on mobility, health, and socio-economic factors of community members with disabilities (Clarke, Ailshire, Bader, Morenoff, & House, 2008; Clarke & George, 2005; Renalds, Smith, & Hale, 2010).

The remainder of this discussion, a deeper dive into the literature around this space, is left for Chapter 5 Section 5.1. I explore the background of this issue in the literature spanning from normative assumptions in modelling to more particular misrepresentations of group categories in experiments.

3.3 Approach

This dissertation, serves the dual purpose of 1) placing an emphasis on a critical approach in decision-making with respect to the application of simulation as an apparatus for dynamic analysis and content creation, and 2) revealing, in the manner required by the academic and scientific communities, that such an issue is prudent and worthy of consideration in the defined area of research.

To achieve this goal, I reflect on the literature and build a critical view of the application of synthetic crowds, or crowd simulation. This critical reflection occurs here in this chapter but also at the outset of each following chapter to frame the importance of a critical approach in this area of research. I then address the questions posed in Chapter 1 by piecewise rigorous exploration of hypotheses generated by these questions in a bottom-up hierarchical manner. I separate the notion of locomotion biomechanics from heterogeneity to carefully examine both individually. The first study, regarding biomechanics in steering, serves to highlight not only that there are fundamental differences between steering model outcomes, but also that biomechanical steering provides a level of fidelity which affords a different view of analysis. The second study, regarding heterogeneity in biomechanics, serves as a set of controlled counter examples to the assumption of normativity in crowd outcomes. Taken together these premises support the conclusion laid out in this chapter–biomechanical heterogeneity and the representation it affords is important in synthetic crowds.

To understand the relative performance of locomotion biomechanics in synthetic crowds from multiple perspectives, I devise a new method for comparative study of steering models in synthetic crowds which is ground-truth free and comprehensive. This method attempts to normalize the difference between models in a novel manner and then test the models over a dense and broad sampling of scenario types of increasing complexity. All sources of heterogeneity not associated with model selection are also removed–parameters such as desired speed are equal amongst all models and agents. Rather than devise difficult to interpret measures or scores, the outcomes of this study are crowd-wide measures with direct crowd movement/interaction interpretations. Along this vein, the measures are taken statistically and comparatively evaluated such that outcomes have a theoretical backing that can be checked against assumptions. This is in contrast to several methods of comparative crowd evaluation which invent new methods that may be seen as esoteric or involve numerical manipulations or definitions that abstract away from a rigours understanding of the impact of model selection. This study also combines rigorous qualitative analysis to ensure that emergent behaviours and failures are captured–something quantitative-only approaches may overlook. Additionally, placing a state-of-the-art biomechanically based steering model alongside the industry and research standard models in this large scale study we can better understand the importance of the biomechanical approach at the steering level.

To understand the importance of heterogeneity in locomotion biomechanics, I take the state-of-the-art biomechanical model and produce a well-studied set of conditions, specifically Temporal Gait Asymmetry in patients post-stroke, to induce heterogeneity at the locomotion level. I also produce models to comparatively control for and understand whether the proxies mentioned in this chapter are sufficient for the types of outcomes needed in safety-critical applications. These models are then comparatively evaluated across a hierarchy of "impact" in a bottom-up approach: the footstep action space; the agent collision space; and crowds space. That is, I carefully explore the impact of locomotion heterogeneity from the way steps are produced, to the absolute motion of the agent, to the emergent behaviours and crowd-wide scenario outcomes.

Finally, I review the issues brought forward by this dissertation and use the findings from the above explorations to develop some preliminary approaches to addressing them.

3.4 Summary

This chapter covers the theoretical background for understanding the problem space and reasoning behind the dissertation, namely that the feminist disability approach requires a more critical intersection of crowd simulation work and disability work. The chapter then outlines examples for exactly why this is so important, particularly now, for the crowds simulation and analytics research. Finally, the approach taken throughout the dissertation to satisfy these ends is delineated. In particular, the dissertation takes a controlled hierarchical approach, first separating the relevant concepts iteratively exploring them to build a case for improvement. At the top of this bottom up hierarchy, I provide some preliminary solutions towards addressing the issues presented in the dissertation.

Chapter 4

Importance of Biomechanics

This chapter explores quantitative synthetic crowd evaluation as a difficult but necessary task both generally and towards understanding the importance of biomechanical steering models in crowd simulation. The literature on quantitative crowd evaluation is largely split into two approaches, data-driven and synthetic comparative crowd evaluation. That is, methods are either data driven, from samples of real life data, or synthetic, from measures based on the simulated crowd under varying conditions. This chapter covers the body of quantitative comparative crowd evaluation with a focus on synthetic methods as a necessary but problematic approach to understanding the performance of synthetic crowds outside the space of real data.

Both approaches to comparative crowd evaluation have methodological drawbacks. Data-driven analyses require data on which to operate. Typically, these methods can be split into two categories, those predicated on learning from real data and those predicated on the simulation re-enactment of real scenarios. The re-enactment methods are highly dependent on the accuracy of the re-enactment itself, and it is clear that generally any given crowd will not recreate the same behaviours on two instances of the same scenario (crowds are generally non-deterministic). The learning methods are difficult to generalize outside of the training data conditions. Typically, the problem exists that, while a data driven method may train and test well, it is near impossible to test on conditions not in the training or testing set. This is partially because a complete and accurate real life crowds data set is intractable across the space of possible situations, contexts, and cultures. In particular, those situations which are deadly are of the highest importance but lack data. Synthetic comparative crowds analyses suffer from the opposite issue, in that there is no ground truth for which to compare results to. Synthetic crowds can be compared in any environment handled by the crowd simulator and do not rely on real world data. This is problematic because it is difficult to say if the model is reproducing desired behaviours, and if the behaviours are sufficiently "accurate."

A key issue across all comparative crowds analysis methods is the differing model parameter dimensions across the compared models. For example, the dimensions of a force-based model will likely differ from a space-time planning model. Additionally, the parameters will control properties of the models which produce different emergent behaviours in the resultant crowd simulations. This is effectively solved by optimizing a model's parameters with respect to real data prior to comparative evaluation (Wolinski et al., 2014a, 2014b). While this works very well when real data is present, and the comparison is contained to that particular scenario, this method is not directly applicable to comprehensive and exhaustive testing in synthetic scenarios. Furthermore, in (Berseth et al., 2014) the problem of optimizing a model for a particular metric is effectively solved. This method allows a crowds designer to decide what is important about the crowd behaviour they wish to achieve, in terms of quantitative metrics, and find the best parameters set for that model with respect to those metrics. In this chapter, I propose the effective amalgamation of these two methods for comprehensive comparative analysis of crowd simulation models in synthetic scenarios.

This chapter represents, to the best of my knowledge, the largest scale comparative crowds evaluation study presented thus far. Crowds models are typically applied across a range of applications in industry, from art, via animation and gaming, to policy and design decisions, via environment scenario simulation. The issue of model applicability is rarely addressed in the most important of these scenarios, where decisions may impact real world outcomes (Still, 2007). This study explores where and why models succeed and fail across a variety of scenarios from both quantitative and qualitative perspectives. Specifically, this study tries to understand the applicability of models and provides a controlled study to assist end user's in applying models for their own use.

4.1 Background

This section provides a more in-depth review of all relevant literature in quantitative comparative crowds analysis and is intended to extend Chapter 2 Section 2.2.1. Continuing the conversation from there, there are drawbacks in terms of applying models to different applications and our confidence in their results (Still, 2007). Several methods have since been developed to carry out the important and necessary task of comparative crowds analysis. Comparative crowds analysis can be broken into quantitative methods and qualitative methods. The quantitative methods are of particular interest for model design, validation, and application because they provide meaningful numbers to base decisions on. Qualitative analysis is difficult to encompass because of the expectations on may have in terms of outcomes. Context, culture, emotion, and several other factors affect the outcomes of real world crowds and in terms of models we can only hope to recreate expected behaviours under very specific conditions (Still, 2007). This review of works focuses in the quantitative area of the literature.

The first works to blend traffic and comparative pedestrian analysis in the scope of environments adapted dynamic highway traffic analysis to the space of pedestrian environment (Fruin, 1971). This work examined the impact of environment design on the quality of pedestrian crowd flow. An early and particularly important work in this area for synthetic crowds was the Social Forces model (Helbing et al., 2005). This crowd model is discussed in detail in Section 4.2.1.1 and relies on forces to model a net steering force per time-step. Most importantly, this work examined the impact of different scenarios on the crowd simulator. Follow-up work in this area studied the impact of panic comparatively across level of panic and velocities (Helbing et al., 2000). These studies provide insights to within simulator performance under different conditions as well as across parameters, such as desired velocity and "panic". More recently this work looked at the transition phases between crowd behaviours, such as laminar to stop-and-go, with a comparative analysis of previously computed fundamental diagrams (Helbing, Johansson, & Al-Abideen, 2007).

Direct comparison of synthetic crowds via data-driven, i.e. predicated on real data, methods has a rich history in the literature (Junior, Musse, & Jung, 2010; Zhan, Monekosso, Remagnino, Velastin, & Xu, 2008). A sampling of these methods which relate to comparative crowds analysis is covered here. An important paper in this area focuses on comparing real sub-crowds within the same event (Johansson, Helbing, Al-Abideen, & Al-Bosta, 2008). In particular, this paper focuses on issues found in all comparative crowd analysis: when looking at density there are multiple local densities where the maximum occurring density is what relates to safety; the difference in impact across cultures and contexts on desired speeds; the difference of measurements used in the literature; the difficulty of using fundamental diagrams prescriptively. For data-driven comparative analysis of synthetic crowds, methods typically set up the problem as one of a similarity measure to the real data. By sampling microscopic state-action pairs from real data, synthetic crowd can be directly compared per simulation time-step, i.e. their short-term decisions, by measuring the difference as a function of density or proximity (Lerner et al., 2010; Lerner, Chrysanthou, Shamir, & Cohen-Or, 2009). Tangentially, under the assumption of small grouping in the majority of crowds, metrics such as dispersion, distortion, and group outlier percentage can be computed on said state-action pairs (Karamouzas & Overmars, 2012). Similarly, by reconstructing a scene, metrics which measure the difference, or distance, between real and synthetic can be used. For example the distance between distributions of agents between real and reconstructed synthetic scenarios across measurement areas (Dupre & Argyriou, 2017; Banerjee & Kraemer, 2011), or using several similarity metrics such as (Jablonski, Argyriou, Greenhill, & Velastin, 2015). However, complete and accurate scene reconstruction implies a modelling pre-step cost, and the accuracy of the scene reconstruction may greatly impact results. This cost and potential source of error is directly related to scenario complexity and may become intractable. The concept of measuring distribution distances is significantly extended in the formulation of 4D crowd phase space histograms (Musse, Jung, & Cassol, 2016; Musse, Cassol, & Jung, 2012). These histograms capture macroscopic flow details. A clustering method predicated on iteratively refining cluster members based on distribution distance affords the analysis of more micro "sub-flows". This method allows those choosing synthetic crowd models to pick the model for particular applications based on their performance in the same environments limited to position and velocity space measures in simple scenarios. Similarly, viewing crowd behaviours as a distribution of location-orientation pairs one can formulate trending paths (H. Wang, Ondřej, & O'Sullivan, 2017). This provides a useful qualitative abstraction of crowd behaviours in the form of abstract patterns. Similarly, distance measures, such as KL-divergence or per-state predictive likelihood from pairs of datasets. Beyond measuring distribution distances, other methods provide meaningful analysis of deviation from real data. Using functional Principal Component Analysis affords a multi-factored view of the sources of variability in synthetic pedestrian dynamics (Chraibi et al., 2016). Additionally, the method removes the effect of lateral swaying from real data to normalize with respect to non-biomechanical crowds modelling. Human bipedal locomotion induces swaying and is in part due to energy saving distribution of mass to maintain balanced forward motion. Removing this unfairly gives credence to models which ignore this motion. Evidence shows that shoulder movement, when modelled at the collider levels, induces changes in dense crowd manoeuvring and the perception of crowds (Hoyet, Olivier, Kulpa, & Pettré, 2016; Stüvel, de Goeij, van der Stappen, & Egges, 2015). These data-driven methods highlight the issues with attempting to reconstruct real data scenes for model validation in data-driven comparative analysis, as the studied models do not perform well at scenario reproduction. Authors across this literature are clear to note that it is difficult to conclusively make model choices based on these types of analyses.

Another approach to comparative analysis involves quantitative metrics computed on the synthetic crowds themselves without the underlying assumption of a data-driven ground truth to make meaningful comparisons. An early a method in this area used numerous per agent and aggregate metrics and a body of scenarios composing a benchmark set designed stress test crowd simulators (Singh et al., 2009). To make such a large set of diverse metrics broadly usable, the metrics, both per agent and per crowd, may be combined in a weighted sum producing a single score per scenario. An important follow-up to this work reformulates the quantitative metrics as quality metrics based on their optimal values for a given scenario (Kapadia, Wang, Reinman, & Faloutsos, 2011). This functionally allows cross scenario comparison in addition cross simulator comparison. This work also identifies a failure set as a function of a simulator's ability to meet a minimum standard defined by the user. Building on this approach, large scale cross scenario comparison affords even more insights into performance. By sampling a well constrained space of empirically important scenarios, which enforce interaction with an egocentric agent, normalized metrics can be computed over a battery of the possible interactions an agent may face, both agent-agent and agent-environment (Kapadia, Wang, Singh, et al., 2011). This affords a deep stress testing of a model's abilities under a procedural set of scenarios which is extensible by reconfiguring the parameters for generation. This approach focuses on the agent-centric, i.e. egocentric, computation of parameters and forms a strong basis for model analysis but may misinterpret crowd level interactions and performance. Another approach, focused on comparison under evacuation conditions, measures the distributions of flow metrics between crowd models (Viswanathan, Lee, Lees, Cheong, & Sloot, 2014). This work provides important insight in differences between emergent behaviours in crowd models under dense conditions.

Inverting this problem by viewing the crowd simulation from the point of view of the environment, and the impact it has on the crowd, is an important perspective. Understanding, comparatively, the difficulty of an environment for a group of simulators helps us understand the applicability of steering models for particular environment designs (Berseth et al., 2013). This provides insight into the type of difficulties particular simulators face in environments designs. A follow-up to this work focuses on the interactions between parametrizable environment elements and crowds during the optimization processes (Berseth, Usman, et al., 2015). This work reveals how the different behaviours a model reproduces affect the optimization process of an environment and vice versa, the environment layout produces interesting and difficult to predict effects on the simulated crowd behaviours. This work found that optimal configurations often produced emergent cooperative behaviours like laminar flow and vortices. Comparative analysis of synthetic crowds and the effects on flow and density similarly provides insight into the cross-density behaviours which emerge and which may be simulator dependent during this optimization process (Haworth, Usman, Berseth, Kapadia, & Faloutsos, 2017). In particular, patterns of pillars for counter-flows (bi-directional hallway traffic) and for bottleneck flows (uni-directional egress traffic) differed for both the crowd model and the density. It was found that in some cases, a scenario optimized under high density crowd conditions produced designs that improved flow across density conditions. It was also found that pillars near egress points caused the formation of lanes and funnels but that these arrangements differed for each crowd simulator. Similarly, in counter-flow optimizations, optimal pillar placements either produced laminar flow or vortices depending on the crowd model. These works conclusively demonstrate the impact that model choice has on decision-making during crowd-aware design processes.

As previously mentioned, qualitative results in crowd simulation are highly subjective, i.e. dependent on expectations. However, these expectations may not generalize across contexts or cultures. Early work in this area sought to reinforce subjective expectations (Still, 2000). Similarly, making use of Virtual Reality to perform direct perceptual validation of models provides an in depth, in-person qualitative review but remains subjective (Pelechano et al., 2008). Defining and identifying the occurrence or emergence of qualitative phenomenon is very difficult, especially as crowd size grows. To address this problem, the Steerbug platform provides a rule- and sketch-based interface in a large scale behaviour detection framework for synthetic crowds (Kapadia, Singh, Allen, et al., 2009). This work affords the in-depth and user guided qualitative analysis of crowds at any scale. In a similar vein of work are anomaly detection methods for crowds (Charalambous, Karamouzas, Guy, & Chrysanthou, 2014; Mehran, Oyama, & Shah, 2009). These may be directly repurposed for the qualitative analysis of synthetic crowds.

A persistent and difficult problem across the space of comparative crowds analysis is the difference in dimensionality between models being compared, and furthermore, the underlying approach to the model itself. This is an important confounding factor in comparative analysis of synthetic crowd because the space of scenarios is infinite and the models in question may be biased towards particular scenarios. To address this issue, it is possible to first calibrate all the models to be compared with respect to real data (Wolinski et al., 2014a). This is a data-driven model optimization process which affords a fairer direct comparison between steering models. Another approach, not reliant on ground truth data, is the Steerfit process which optimizes models parameters by minimizing or maximizing some criteria (Berseth et al., 2014). This affords cross models comparison, by optimizing models for the same criteria (crowd metric) under the same conditions (scenarios), as well as behavioural tuning for content generation, by providing a straightforward means of exploring the model parameter space for non-experts.

This work is inspired in part by model validation procedures and recommendations in simulation modelling (Sargent, 2009). Following this approach, this dissertation presents data analysis of industry and research standard models to help model consumers choose the model-type for their application. Additionally, this paper provides a method, measures, and comprehensive testcases set which helps model producers follow these recommendations when devising new crowd simulation models. The majority of the presented testcases are pulled from the long history of comparative crowds analysis and represent the space of scenario crowd simulators are utilized in. In addition to these testcases, some cases from real world designs are included to represent certain complexities found in built environments which synthetic crowd struggle to solve.

4.2 Overview

This study explores the quantitative and qualitative performance of steering models under several conditions. This method, including the model normalization, benchmarks, and metrics, are designed to address limitations of comparative evaluation of synthetic steering models. Prior comparative analysis methods for steering models, as well as, literature delineating new steering models has several notable limitations. While the specifics and limitations of particular comparative methods were outlined in Section 4.1 the following is a compiled list of limitations in the crowd simulation literature.

- **"Ecologically valid" evaluation tasks.** Most crowd simulators are evaluated by using a small subset of mostly ecologically invalid scenarios, such as diametric circles, four-way crossing, or very specific use cases like plane evacuation scenarios. These are not ecologically valid across applications, and not necessarily ecologically valid within crowd simulations. A particular shortcoming of interest is evaluation without static obstacles or where the static obstacles are placed in ways as to not actually impact evaluation.

- **Ranked outcomes without normalization.** Several approaches rank evaluation outcomes, such as one better than the other, without much statistical validation, usually under the assumption that the models can be directly compared without any normalization and that underlying outcome measure distributions are the same.

- **Missing ground truth or reliance on ground truth.** Relying on a ground truth makes an assumption of perfect predictability in real crowds, and is intractable under the space of conditions due to: 1) lack of data; 2) the cost of reproducing the initial conditions of a single scenario; and 3) the chaotic-like nature of human crowds (mainly that any small change in initial conditions create very large changes in outcomes especially for any given single participant in a crowd). Because of these limitations, it is common to generate ground truths from noisy and specific data (for example, limited catastrophe footage), or construct non-ecologically valid scenarios to record real data with experimental participants. However, these participants are not performing in the context of their regular environment.

I propose a new method, the amalgamation of two previous approaches, to alleviate the reliance on ground truths while normalizing high dimensional models for direct comparison. Each crowd simulation model (specifically steering model in this dissertation) is first optimized for the same metric using the parameter optimization method in (Berseth et al., 2014) and then compared across micro and macro metrics in the spirit of (Wolinski et al., 2014a). This affords a within scenario across simulator comparison method, that is, an analysis such that each model's approach can be directly compared under each set of conditions. Specifically, in the first step, the models must be normalized on some metric derived from the simulated agents performance. Humans naturally minimize effort during locomotion tasks (Zarrugh, Todd, & Ralston, 1974; Minetti & Alexander, 1997; Kuo, 2001; Bertram & Ruina, 2001), thus, the parameter sets are optimized with respect to metabolic energy expenditure, referred to as 'effort' throughout the dissertation. A crowd-wide definition of this metric can be found in Section 4.2.2.2. The final parameter sets can be seen in Appendix A.

All models are then tested under a battery of scenarios that span applications from research to industry. Some of these scenarios are pulled from prior literature (Sections 4.3, 4.4,& 4.5) and some are difficult combinations of scenario elements found in built environment designs (Section 4.6,& 4.7). This approach provides insight into the implications of prior comparative methods as well. In particular, the benchmark set focuses on scenario types with global and local features of increasing difficulty for models to solve and are designed to test the limits of the models. That is, we expect models to fail under several conditions, and by eliciting these conditions we can study their flaws, performance, and provide empirical evidence for their applicability in different domains. By combining both new and previously researched benchmarks together in a hierarchical bottom-up fashion from the most straightforward solution expectations to the most complicated, this method presents a novel means of rigorously benchmarking steering algorithms. First, the models are tasked with solving simple interaction scenarios designed to elicit whether models have any core faults, such as resolving common sub-scenarios like moving towards a goal or interacting with obstacles. Second, the models are run in a battery of scenarios that force agent-agent and agent-obstacle interactions with at least one agent. Third, models are placed in conditions of increasing density in scenarios that mimic sub-scenarios of a larger scale environment movement (such as a bottleneck egress point, or a group crossing point). Fourth, the models are tasked with solving a larger scale egress with several concavities.

Each experimental condition is a scenario space $\mathbb{S} = \langle O, A \rangle$ defined by the distribution of obstacles O and agents A in an environment. An obstacle $o \in O$ is defined by its position, rotation, and shape, while an agent $a \in A$ is defined by their model, initial conditions (starting position and rotation), and goals (a series of world positions). Both O and A may be generative or explicit in that they are defined by the bounds of their parameters in which they may be generated or explicitly defined by given parameters respectively. This allows for a scenario $\mathbf{s} \in \mathbb{S}$ to be fully or partially generated from the scenario definition $\mathbb{S}$. Thus, each $\mathbf{s}$ is a sample of the experiment defined by the parameters of $\mathbb{S}$. Each of these $\mathbb{S}$ is discussed in detail in their respective sections.

Several parameters of this study are constant across all the experiments conducted. Those aspects of the study are described here.

4.2.1 Independent Variables

This study focuses on four different model-types that span the spectrum in terms of implementation and common use in industry and research. The theoretical and historical bases of each model is discussed here as a backbone for discussions later in the section. There is a large body of evidence in the literature that outcomes of crowd simulation are deeply dependent on model choice, see Section 4.1. While the proposed method attempts to normalize differences between models for comparison, the following choice of models is based on: 1) their ubiquity in research and industry; and 2) demonstrated deviations from each other both quantitatively and qualitatively across the body of literature and reported outcomes.

4.2.1.1 Net Force Models

In force-based crowd models the steering decision is based on a net of repulsion and attraction forces. This approach was first introduced to model the forces of social attraction and repulsion during locomotion in order to inform built environment designs (Helbing & Molnar, 1995). This model has been extended and applied to escape panic during evacuation in order to inform evacuation strategies (Helbing et al., 2000). This study utilizes Social Forces, which we will refer to as $sfAI$ throughout this chapter.

4.2.1.2 Predictive Models

Predictive models of local steering and collision avoidance predict imminent collisions and make early choices that effect the current steering decision. While collision avoidance is partially predicated on predictive-like decisions, these models explicitly predict future collisions and alter their current local steering decisions accordingly. This approach is particularly effective at reducing oscillations and step like or jagged movement. Several models utilize predictive methods, including all models in this study except for *sfAI*. This study utilizes the predictive avoidance model, which will be referred to as *pamAI* throughout this chapter, an extension of the Social Forces net force model (Karamouzas et al., 2009).

4.2.1.3 Velocity Models

The introduction of velocity obstacles (VO) for robotics VO (Fiorini & Shiller, 1998, 1993) provided a simple means of finding guaranteed collision free steering decisions. This method finds the abstract obstacle formed by the set of velocities for which two objects will collide at some point and then avoids choosing relative velocities within this "velocity obstacle" area of velocity space as if it were an actual obstacle.

Extended forms of this method, in particular, RVO and RVO2 both anticipate the reactive behaviour of other agents. In RVO, this is a achieved by assuming all agents make the same collision avoidance decisions and moving toward the end results of that set of decisions rather than stepping towards it (Van den Berg et al., 2008). This approach avoids the step-like or jagged patterns VOs may create. RVO2 takes this assumption one step further and solves an optimization problem to find a collision free velocity in velocity space. The collision free velocity space is formed by the concatenation of half-planes induced by the velocities of the N interacting agents (van den Berg et al., 2011). This study utilizes RVO2, which we will refer to as $rvo2dAI$ throughout this chapter, a predictive and reciprocal model.

4.2.1.4 Space-time Planning Models

Planning models produce a short-term path, or trajectory, to resolve imminent collisions. One approach to planning is to model a discretized ego-centric perceptual field. By finding the overlap, or intersection, of obstacles and agents and performing linear prediction on this field, a short-term local plan can be formed (Kapadia et al., 2012; Kapadia, Singh, Hewlett, & Faloutsos, 2009).

Footstep-based planning is used by both the computer animation community and the robotics community. In crowd simulation, a space-time planning method for local steering and collision avoidance based on footstep placement affords a non-linear prediction of collisions, tightly packed crowds, piecewise stepping actions, and most importantly heterogeneity at the step level (Berseth, Kapadia, & Faloutsos, 2015; Singh, Kapadia, Reinman, & Faloutsos, 2011). This study utilizes the footsteps model, which is referred to as $footstepAI$ throughout this chapter.

4.2.2 Dependent Variables

This research focuses on the crowd wide performance of agents with respect to scenario and steering model. To capture and understand what we consider to be the most important factors in understanding quantitative crowd performance we consider contacts, effort, flow rate, and path length as defined here.

4.2.2.1 Unique Contacts

Crowd simulation can be thought of as a hierarchical set of considerations for entities to achieve while traversing an environment towards some goal. In agent-based, and more specifically particle-based, crowd simulation at the local steering level of the hierarchy, a primary consideration is collision avoidance. The term collision may have a subjective meaning however. Qualitatively, several models, especially force-based models, may have what can be thought of as Newtonian collisions in which the particles never actually touch but their repulsive forces create motion which appears, at a distance, to be the result of a collision. In this research, we are specifically trying to understand where models fail or succeed in solving scenarios. Considering this, collisions are defined in terms of interpenetration contact, which can be considered an invalid state for a particle simulation to be in. To understand how a model performs with respect to contact, the total number of unique interpenetration contacts is counted over the course of a scenario simulation for each agent's particles with all other agent's particles as follows:

$$C(A) = \sum_{t \in T} \sum_{(a,b) \in \binom{A}{2}} \sum_{\substack{(p_1,p_2) \\ \in (a_p) \times (b_p)}} o(t, p_1, p_2) \tag{4.1}$$

where T is the set of all time samples, t is the index of each sample, A is the set of all agents, a_p and b_p are agent a's and agent b's particles respectively[1], and $o(t, p_1, p_2)$ is an interpenetration contact indicator at time t defined as:

$$o(t, p_1, p_2) = \begin{cases} 1 & \text{if } {}^{||p_2^c - p_1^c|| < p_2^r + p_1^r and}_{o(t-1, p_1, p_2) = 0} \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where p_1 and p_2 are particles defined by their centres p_1^c and p_2^c with radii denoted by p_1^r and p_2^r respectively.

4.2.2.2 Effort

Measuring effort helps provide insight into how well a model mimics human behaviours as humans naturally optimize effort expenditure when walking. Effort spent by a person walking can be effectively expressed as total metabolic expenditure over a path (Guy et al., 2010). In this study, the total effort expenditure over the crowd is defined as the sum of all agents' effort over their path integral:

$$E(A) = \sum_{a \in A} m_a \int (e_s + e_w |\mathbf{v}_a|^2) dt \tag{4.3}$$

where m is the mass of each agent (in this study, m is homogeneous and set to 1 Kg to remove size/mass heterogeneity which is outside the scope of this dissertation), e_s and

[1]For most algorithms the agent is defined by a single particle (i.e. $|(a_p)| = |(a_b)| = 1$), but this is not always the case (e.g. the footsteps model).

e_w are biological constants set to empirical averages for human walking, 2.23 J/Kg·s and 1.26 Js/Kg·m^2 respectively, and v_a is the velocity of agent a.

4.2.2.3 Flow Rate

Pedestrian dynamics have been studied in the context of environment traffic, analogous to vehicular traffic (Fruin, 1971). The application of qualitative labels to density and predictive flow qualities in pedestrians traffic flows facilitates the discussion and measurement of pedestrian flow dynamics. In this study, we define flow rate as the rate at which agent's complete their final goal over the entire simulation:

$$F(A) = \frac{|A|}{T_f} \tag{4.4}$$

where T_f is the final simulation time. While there are several approaches to defining flow rate, this formulation is a simple and tested method that can be thought of as goal completion rate (the area or point at which the flow is measured is the goal area) (Haworth, Usman, Berseth, Kapadia, & Faloutsos, 2017).

4.2.2.4 Path Length

Measuring total path length is a straightforward way to understand the impact of emergent behaviours in crowd simulators. Ideally, humans minimize the length of the path they travel on towards their goals. In crowds, this produces interesting emergent behaviours as the shortest linear path is rarely possible. In this study, the path length is approximated over a poly-line trajectory of each agent:

$$P(A) = \sum_{a \in A} \sum_{s \in S_a} ||s|| \tag{4.5}$$

where S_a is the poly line path traversed by agent a defined by the linear segments s whose endpoints are captured at each sampling.

4.2.3 Additional Qualitative Discussion and Failure Ratio

It is important to explore the emergent behaviours of each of the models as well as how each of the models fail under particular conditions. In each scenario, success is defined as, "all agents have complete their goals." In addition to the aforementioned crowd measures, we also measure failure rate as the ratio of successfully completed scenarios to the entire study testcase set. For failed and outlier scenarios, the reasoning behind the failure is explored and discussed. Additionally, the emergent behaviours of successful scenarios are explored and discussed.

4.2.4 Statistical Analysis

In the comparison of steering models, care must be taken in the assumption made about outcomes and their significance. The method of comparison described here assumes nothing about the compared algorithm set, and is intended to be applied broadly (to new and different algorithms). Instead of presenting a new metric or method which moves away from classic and well-known statical tests, this method aims for a rigorous application of well-known techniques which can be understood from the literature regarding their construction and use. These methods also assume a generalizability to application. This means a statistical method is needed which is generally applicable for comparison across multiple groups without assumptions of underlying distributions of the groups and handles the relative ranking of outcomes. In statistical hypothesis testing, there is ultimately a tension between broad applicability and agreeable panacea in such a case. On one hand, common parametric tests like ANOVA assume normality of underlying distributions (but can be robust to large sample sizes), while on the other hand non-parametric tests do not assume normality, but they may assume homoscedasticity for particular interpretations.

The data presented in this chapter has undergone careful analysis to understand what methods of statistical significance may be applicable. Given the above constraints, the fact that heteroscedasticity is sometimes present (verified via Levene's test), outliers are highly dependent on scenario type, and the number of samples is relatively large (always $N > 20$), I propose using the Kruskal Wallis test (Kruskal & Wallis, 1952). Kruskal Wallis has interpretations based on assumptions about the underlying distributions, these are: (1) when the distributions of the data are homoscedastic, i.e. they are identically distributed, the test can be interpreted as a test between medians; (2) if they are additionally distributed symmetrically the test can be interpreted as a test between means; and (3) if no assumption is made about underlying distribution the test can be interpreted as dominance of one distribution over another. In this work, I make use of (3), to provide a general approach with no assumptions regarding underlying distributions. The use of boxplots helps identify where this dominance occurs by characterizing the distribution graphically. Additionally, should there be a significant result, this test is then follow with a battery of post-hoc tests to find the location of these pairwise significant differences and help control for the errors that may stem from using a panacea method. This battery includes the pairwise multiple comparison post-hoc tests Conover (Conover & Iman, 1979), Dunn (Dunn, 1964), and Nemenyi (Nemenyi, 1963). In the case of rank ties the FWER (Hochberg & Tamhane, 1987) & FDR (Benjamini & Hochberg, 1995) correction methods for Conover and Dunn are used, and the Chi-square null distribution of the test statistic for Nemenyi is used. In this proposed method, results are only reported as significant if *all* pairwise post-hoc tests are $p < 0.01$. Taken together, this strategy should provide for broad applicability while retaining a high degree of confidence in results.

4.3 Simple Interactions

This experiment comparatively explores the performance of each model in a set of thirty scenarios commonly used in the analysis of models in the literature. The majority of the set was previously defined as a spectrum of difficulties for which to stress test specific areas of performance in steering models. Each scenario is explicitly defined with agent initial conditions, goals, and obstacles where applicable.

The scenarios range from a single agent in an empty environment navigating toward a nearby goal to several agents arranged in concentric circles with diametric goals. Of particular interest are those scenarios commonly used in the literature, such as the diametric agent scenario, *circle 20*.

4.3.1 Material and Methods

In addition to the materials and methods outlined in Section 4.2, this experiment is performed on the following scenario set.

4.3.1.1 Scenario Set

All scenarios in this experiment have explicitly defined initial conditions and thus were simulated one time, (experiments showed repeated simulation does not provide additional information). This set comprises 28 scenarios common in the crowd simulation literature which build from straightforward to complicated and span expected interaction cases. These scenarios are outlined in Table 4.1.

4.3.2 Results

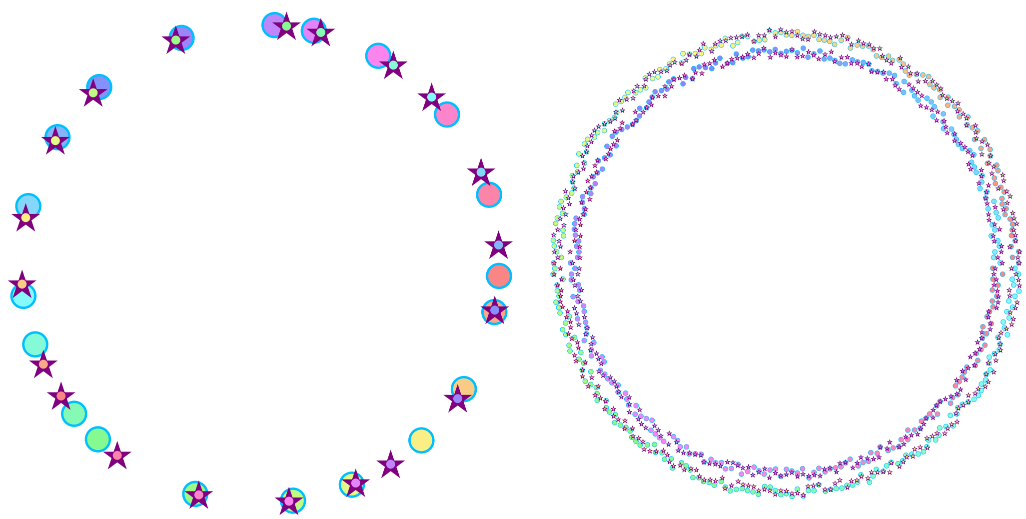
The results for the study are separated in tables by metric: total unique contacts in Table 4.2; total effort in Table 4.3; flow rate in Table 4.4; and total path length in Table 4.5. Each table is coloured as a per row heatmap (where green is the better score for that metric) to highlight the success of each steering modelling approach.

4.3.3 Discussion

In this study, across all benchmarks all models are configured to a desired and maximum speed of approximately 1.3m/s. This is to control for the heterogeneity of parameter settings, since the rest of this dissertation addresses that separately. While higher flow rate

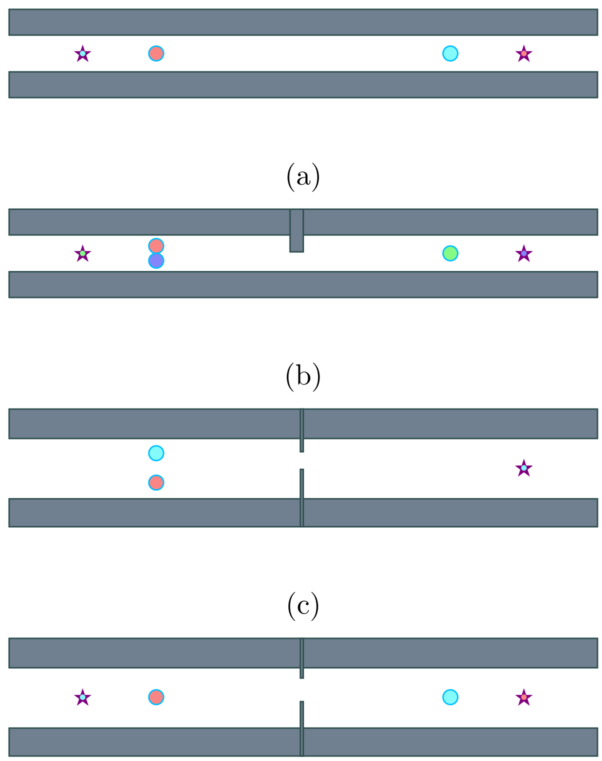
Table 4.1: All simple interactions scenarios with descriptions.

Scenario	Description	Focus
simple 1	One agent with obstacle immediately behind.	**Model Validation**
simple obstacle 1	One agent moving towards goal behind obstacle.	
simple wall	One agent moving towards goal behind wall.	
curves	One agent steering through a piecewise quantitzed S-curve.	
oncoming 1	Two agents moving towards a head on collision.	**One-on-One Interactions**
crossing 1	Two agents crossing paths at a right angle.	
oncoming trick	Two agents meeting their goals near but not crossing each other.	
crossing trick	Two agents meeting their goals near but not crossing each other at right angles.	
similar direction	Two agents with marginally different goals moving in the same direction.	
oncoming obstacle	Two agents crossing paths with an obstacle at the crossing point.	**Agent-Agent Interactions**
crossing obstacles	Two agents crossing paths at a right angle with an obstacle at the crossing point.	
surprise 1	Two agents crossing at the corner of a large obstacle blocking line of sight.	
squeeze	Two agents crossing in a narrow corridor with enough space for both.	
doorway one way	Two agents passing through a doorway from one side.	
doorway two way	Two agents passing through a doorway at their crossing point from both sides.	
fan in	A group of agents with the same goal.	**Group Interactions**
fan out	A group of agents with spread out goals.	
cut across 1	One agents crossing the path of a group of agents.	
3 way confusion 1	Three agents crossing paths.	
4 way confusion	Four agents crossing paths at a single crossing point.	
4 way confusion obstacle	Four agents crossing paths with an obstacle at the crossing point.	
frogger	One agent crossing the paths of three agents moving perpendicular to the agent.	
oncoming groups	Two groups of agents crossing paths.	
3 squeeze	Two agents crossing paths with one agent in a narrow corridor.	
double squeeze	Two agents crossing paths with two agents in a narrow corridor.	
wall squeeze	Two agents and one agent passing through a doorway at their crossing point in a narrow corridor.	
circle 20	Twenty agents on the circumference of a circle with diametric goals.	**Large Scale**
concentric circles	Five hundred agents on the circumference of two concentric circles with noisy diametric goals.	



(a) (b)

Figure 4.1. The two diametric goals scenarios from the common set where each agent is a coloured circle with a matching goal star. (a) the *circle 20* scenario with 20 agents, and (b) the *concentric circles* scenario, a related but significantly more difficult case with 500 agents in two noisy concentric circle arrangements. Note that (b) is multiple times larger than (a) and is scaled for visualization here–the diameter of (a) is approximately 20m and of (b) is approximately 100m. Several variants of the diametric goals scenario have been used throughout the crowd simulation literature.



(a)

(b)

(c)

(d)

Figure 4.2. The subset of common scenarios that involve difficult hallway configurations (the intersection of people, space, and activities in the simple interactions): (a) *squeeze* (b) *wall squeeze* (c) *doorway one way* (d) *doorway two way*.

Table 4.2: Total unique collisions for all common scenarios.

Scenario	*footstepAI*	*pamAI*	*rvo2dAI*	*sfAI*
simple 1	0	0	0	0
simple obstacle 1	0	0	0	0
simple wall	0	4	4	0
curves	0	8	7	0
oncoming 1	0	2	2	0
crossing 1	0	0	0	0
oncoming trick	0	0	0	0
crossing trick	0	0	0	0
similar direction	0	0	0	0
oncoming obstacle	0	4	3	0
crossing obstacle	0	4	2	N/a
surprise 1	2	5	3	N/a
squeeze	0	2	2	N/a
doorway one way	0	4	4	0
doorway two way	0	4	3	0
fan in	0	0	0	0
fan out	0	0	0	0
cut across 1	0	0	0	0
3 way confusion 1	0	0	0	0
4 way confusion	0	4	2	0
4 way confusion obstacle	0	8	7	0
frogger	0	2	2	0
oncoming groups	0	0	14	0
3 squeeze	0	2	2	0
double squeeze	0	0	0	0
wall squeeze	0	5	8	0
circle 20	2	4	0	0
concentric circles	7108	44078	25668	819256

Table 4.3: Total effort for all common scenarios. All units are kgm^2/s^2

Scenario	*footstepAI*	*pamAI*	*rvo2dAI*	*sfAI*
simple 1	0.89	0.11	0.26	0.22
simple obstacle 1	34.65	30.24	30.33	29.86
simple wall	232.02	208.36	207.44	209.21
curves	99.04	89.46	89.46	95
oncoming 1	140.56	324.31	329.31	378.94
crossing 1	136.12	122.92	123.12	122.58
oncoming trick	68.88	60.05	64.4	58.42
crossing trick	60.6	52.37	51.64	51
similar direction	352.36	324.45	336.92	323.78
oncoming obstacle	144.19	128.32	127.66	139.58
crossing obstacle	134.12	120.76	123.43	N/a
surprise 1	201.2	209.55	194.59	N/a
squeeze	177.35	161.21	377.11	N/a
doorway one way	180.35	160.68	162.78	180.27
doorway two way	177.63	161.22	161.14	169.79
fan in	424.9	529.19	532.36	527.64
fan out	576.82	229.85	231.11	236.93
cut across 1	875.81	788.67	838.52	785.81
3 way confusion 1	232.53	211.11	234.26	213.2
4 way confusion obstacle	284.51	258.87	259.91	266.18
4 way confusion	282.32	256.97	256.79	307.21
frogger	255.65	127.66	126.91	138.19
oncoming groups	2036.87	1886.42	1885.43	1909.49
3 squeeze	266.59	241.36	242.44	246.13
double squeeze	355.32	380.98	384.67	391.33
wall squeeze	269.7	242.69	250.18	253.3
circle 20	1371.3	1205.53	1233.65	1195.67
concentric circles	185450.2	175112.73	213682.3	675853.83

Table 4.4: Flow rate for all common scenarios. All units are *agents/s*

Scenario	*footstepAI*	*pamAI*	*rvo2dAI*	*sfAI*
simple 1	2.5	20	20	20
simple obstacle 1	0.1	0.14	0.17	0.15
simple wall	0.05	0.08	0.09	0.08
curves	0.03	0.05	0.06	0.04
oncoming 1	0.1	0.2	0.23	0.37
crossing 1	0.1	0.14	0.16	0.14
oncoming trick	0.19	0.26	0.25	0.3
crossing trick	0.22	0.31	0.39	0.34
similar direction	0.04	0.05	0.06	0.05
oncoming obstacle	0.09	0.13	0.16	0.11
crossing obstacle	0.1	0.13	0.16	N/a
surprise 1	0.07	0.07	0.1	N/a
squeeze	0.08	0.1	0.03	N/a
doorway one way	0.07	0.1	0.12	0.08
doorway two way	0.08	0.1	0.12	0.09
fan in	0.27	0.13	0.15	0.13
fan out	0.09	0.26	0.31	0.24
cut across 1	0.13	0.18	0.17	0.19
3 way confusion 1	0.12	0.17	0.15	0.16
4 way confusion	0.19	0.25	0.3	0.19
4 way confusion obstacle	0.19	0.23	0.3	0.24
frogger	0.18	0.13	0.16	0.12
oncoming groups	0.2	0.27	0.33	0.27
3 squeeze	0.11	0.15	0.19	0.15
double squeeze	0.15	0.36	0.45	0.15
wall squeeze	0.11	0.15	0.17	0.14
circle 20	0.91	1.34	1.54	1.42
concentric circles	3.54	3.13	3.52	2.14

Table 4.5: Total path length for all common scenarios. All units are m

Scenario	*footstepAI*	*pamAI*	*rvo2dAI*	*sfAI*
simple 1	0.01	0.01	0.08	0.07
simple obstacle 1	9.38	8.9	8.97	8.9
simple wall	64.98	61.85	61.33	62.15
curves	27.78	26.54	26.45	28.04
oncoming 1	38.72	37.45	37.47	38.3
crossing 1	37.63	36.43	36.4	36.5
oncoming trick	18.35	17.44	17.48	17.42
crossing trick	16.27	15.34	15.29	15.21
similar direction	98.68	96.52	99.61	96.55
oncoming obstacle	39.42	37.64	37.72	38.96
crossing obstacle	36.9	35.74	36.5	N/a
surprise 1	55.78	61.18	57.52	N/a
squeeze	48.93	47.45	75.83	N/a
doorway one way	49.21	47.63	48.14	49
doorway two way	49.06	47.45	47.61	49.06
fan in	116.6	157.35	157.38	157.32
fan out	161.02	68.03	68.36	69.82
cut across 1	243.11	234.33	247.85	234.24
3 way confusion 1	64.32	62.57	69.21	63.28
4 way confusion	77.66	75.64	75.85	78.45
4 way confusion obstacle	78.42	75.41	76.86	77.82
frogger	70.25	37.45	37.55	38.3
oncoming groups	569.54	559.09	557.46	564.76
3 squeeze	73.5	71.3	71.72	72.39
double squeeze	97.62	112.69	113.62	98.9
wall squeeze	74.1	71.66	73.07	73.84
circle 20	375.79	355.42	364.63	355.86
concentric circles	47364.99	46208.43	62603.5	172755.47

is generally better, adhering to the desired and maximum speeds is critical to achieving expected results for whatever scenario a designer may be modelling. In Table 4.4 the first sign of a common thread which holds throughout the larger study, appears wherein footsteps has a slightly lower flow rate than the other models.

The *footstepAI* model has the best performance in the unique contacts except for the *circle 20* scenario. The *footstepAI* model also has the worst performance in terms of flow rate, path length, and effort. In the effort metric results, there are four outlier scenarios: *oncoming 1*, *fan in*, *double squeeze*, and *concentric circles*. Despite the parabolic paths and the longer time taken, in these scenarios the *footstepAI* solution is more energy efficient. This is somewhat well reflected in the flow rate and path length metrics as well.

Broadly speaking *pamAI* has the best performance in terms of total effort and path length. However, *rvo2dAI* has the best performance in terms of flow rate. The *sfAI* model did not complete the *crossing obstacle*, *surprise 1*, and *squeeze scenarios*. These differences in related measures highlight the importance of evaluating the qualitative results.

4.3.4 Qualitative Discussion and Failure Set

The overall low flow rate of the *footstepAI* model is primarily due to its ability to utilize nearby free space and make non-linear motions, like sidesteps and turning steps, to resolve collisions and path direction. Additionally, the *footstepAI* model is non-holonomic. This means the model, unlike the other models in this study, can not simply move along an available degree of freedom to resolve steering issues, it *must* plan a series of footstep actions to do so. In this set of cases, we begin to see the importance of biomechanics and how it may be overlooked at the micro scale.

A particularly extreme scenario, highlighting an artefact of disregarding biomechanics and/or all aspects of physical simulation, is the *simple 1* scenario. The *footstepAI* model completes a multi-point turn at a speed of approximately one full turn, or 360°, in one second. The *rvo2dAI* and *sfAI* models complete near instantaneous rotations corresponding to an angular velocity of 3600° in one second. Interestingly, *pamAI* ignores rotation in completing the scenario. The insight here is that often open source steering model implementations clamp linear velocities to fall between their minimum and maximum (if they are explicit parameters). However, they may not control for torques and induced angular velocities. Therefore, at agent initialization, when the velocity of the agent is a zero vector and the agent is not in motion, the agent may rotate as fast as necessary. Additionally, often the description of steering models does not include torque control. That being said, the impact of this artefact is part of a rare edge case and is dependent on the desired use of a model. In rendering, like animation or games, this artefact would have negligible impact. The agent would within the first frame (or at most very few frames), outside of human perception, simply appear to be facing the correct direction. However, in analysis there is an interplay between the scenario, metrics, and the model. It is clear that the artefact drastically inflates flow rate because the completion time is the minimum given the update frequency (1agent/0.05s) and decreases the effort because the effort measure does not capture angular velocity. Finally, the scenario is extremely trivial in design and has no practical value in application, but does serve to explicate issues such as this in evaluation. In practice, and in other scenarios in the remainder of this dissertation, this artefact has no impact.

There are two low performing effort outliers for the *footstepAI* model, *fan out* and *frogger*, are due to an interesting parametrization artefact. The balance between energy coefficients, the parametrization used to blend energy costs in the model's optimization scheme, have interesting impacts on qualitative results. In cluttered forced interaction type scenarios (obstacles and agents), the model parametrization used in these experiments produces a minimization of effort–an effect reflected by the well performing outliers with respect to effort mentioned in the quantitative results. However, in very open scenarios, especially those with no obstacles and which require (or would be well solved by) straight paths, the *footstepAI* model produced an angular path of two lines such that the two lines and the shortest path form an obtuse triangle. This is primarily because very straight footsteps of average step length have a lower cost overall, with respect to the internal energy functions used by the model, than making shorter or perturbed footsteps. That is, there is a mismatch in the distance to the goal and how many low energy steps fit in that distance. This can be fixed by increasing the time cost weight, which affects the overall shape of the agent trajectory. However, somewhat unintuitive, this is not an effort optimal strategy in general–something reflected on later in Chapter 6 Section 6.2.2.

The diametric goals scenarios, *circle 20* and *concentric circles*, are good scenarios for understanding the impact scale may have on revealing issues in models. In these scenarios, the outcomes are mostly flipped. The *circle 20* scenario is relatively easy to resolve, and for the most part all algorithms produce very smooth vortices. However, minor difference in how they achieve this lead to different results. The *footstepAI* model makes a somewhat piecewise vortex in that three main groups of agents stick together and three outlier agents which do not quite fit along the ring of these groups move along the outside of them. The vortex like movement happens at the last possible point where the group begins to compress sin the centre of the scenario. The *pamAI* and *rvo2dAI* models produce very similar patterns of near perfectly balanced vortex that moves in a smooth circular formation almost from the very beginning of the scenario. In the *concentric circles* scenario, the predictive nature of the *footstepAI* and *pamAI* models produces nice "optimal" vortices for such a noisy and large case. The *sfAI* model, however, struggles in an interesting manner. The agents that reach the compression point first are trapped by the agents on the outer edges of the group which forms near the centre. The group near the centre oscillates at a high frequency, making no progress toward their goals. The outer layer forms a vortex which remains on the outer layer until linear paths to the goals can be made. This continues, with each "shell" of agents leaving the centre until the entire agent set completes.

The *sfAI* model was the only model to fail under these conditions. These relatively straightforward conditions, where balanced oncoming agents collide, highlight a general problem for force-based models. These models without noise or predictive forces, particularly under heterogeneous conditions or similarly parametrized agents, may reach equilibrium in the net force between two or more interacting agents. This may continue in a functional stalemate with none of the interacting agents completing the scenario. In some cases, numerical error may eventually resolve the conflict by causing gradual sliding between the particle until the agents' net forces overcome their equilibrium.

The results in this experiment highlight an interesting trend in the biomechanical model, *footstepAI*. The model appears to do worse, with respect to the majority of metrics. However, careful qualitative review highlights what is actually happening. In any given path, the other three approaches will produce linear paths–in the completely unobstructed scenario a path between any two points is a perfect line. However, *footstepAI* is predicated on a piecewise trajectory of parabolic curves. These curves, at any point, will be longer than their linear approximation counterparts. Additionally, the COM follows these curves and thus the velocity modulates over them. So what appears to be an under-performing approach to synthetic crowds is, in most cases, actually a significant increase in fidelity.

4.4 Forced Environment and Agent Interactions

This experiment comparatively explores the performance of each model in a comprehensive set of scenarios, the Representative Set. These scenarios were originally intended to force interactions with a reference agent such that normalized metrics (with respect to optimal values) could be computed to make direct steering model performance comparisons across scenarios (Kapadia, Wang, Reinman, & Faloutsos, 2011). Here, the crowd-wise metrics, as described in Section 4.2.2, are used instead of reference agent normalized metrics, as this experiment seeks to understand the crowd level performance of these modelling approaches.

4.4.1 Material and Methods

In addition to the materials and methods outlined in Section 4.2, this experiment is performed on the following scenario set.

4.4.1.1 Scenario Set

This experiment's scenario set is defined by a procedural scenario space designed to force interactions with a reference agent. The Representative Set $\mathbb{S}_r$ is a subset, of cardinality $10,000$, of this space such that the "*coverage*" of models converges (Kapadia, Wang, Singh, et al., 2011). In $\mathbb{S}_r$, each scenario is explicitly defined with agent initial conditions, goals, and obstacles. An example scenario from the representative set, *ID: 1999*, can be seen in Figure 4.3.

4.4.2 Analysis

The summary statistics (median, first quartile, third quartile, IQR, max, min, and outliers) of each of the four metrics are computed over the intersection of all completed scenarios. All measures are tested using the methods outlined in Section 4.2.4 on the

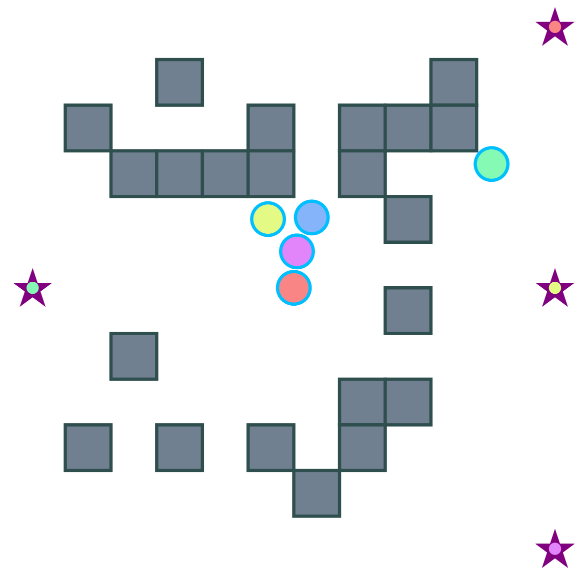


Figure 4.3. An example scenario from the representative set $\mathbb{S}_r$. In this particular scenario, there are five agents and four goals (one agent shares a goal with another). There is a hallway like feature and a high potential for agents crossing paths while navigating the obstacles.

intersection of all completed scenarios.

4.4.3 Results

Boxplots of unique contacts, effort, flow rate, and path length statistics in Figure 4.4. The completion rates for each model are reported in Table 4.10. There is a significant difference present amongst the models for each measure ($N = 7371$, $p < 0.01$). The battery of post-hoc tests for the unique contacts measure reveal there is no significant difference between *pamAI* and *rvo2dAI*, but all other combinations of models are significant ($p < 0.01$), see Table 4.6. All pairwise comparisons of steering models for effort and flow rate are significantly different ($p < 0.01$), see Tables 4.7 & 4.8. For the path length measure,

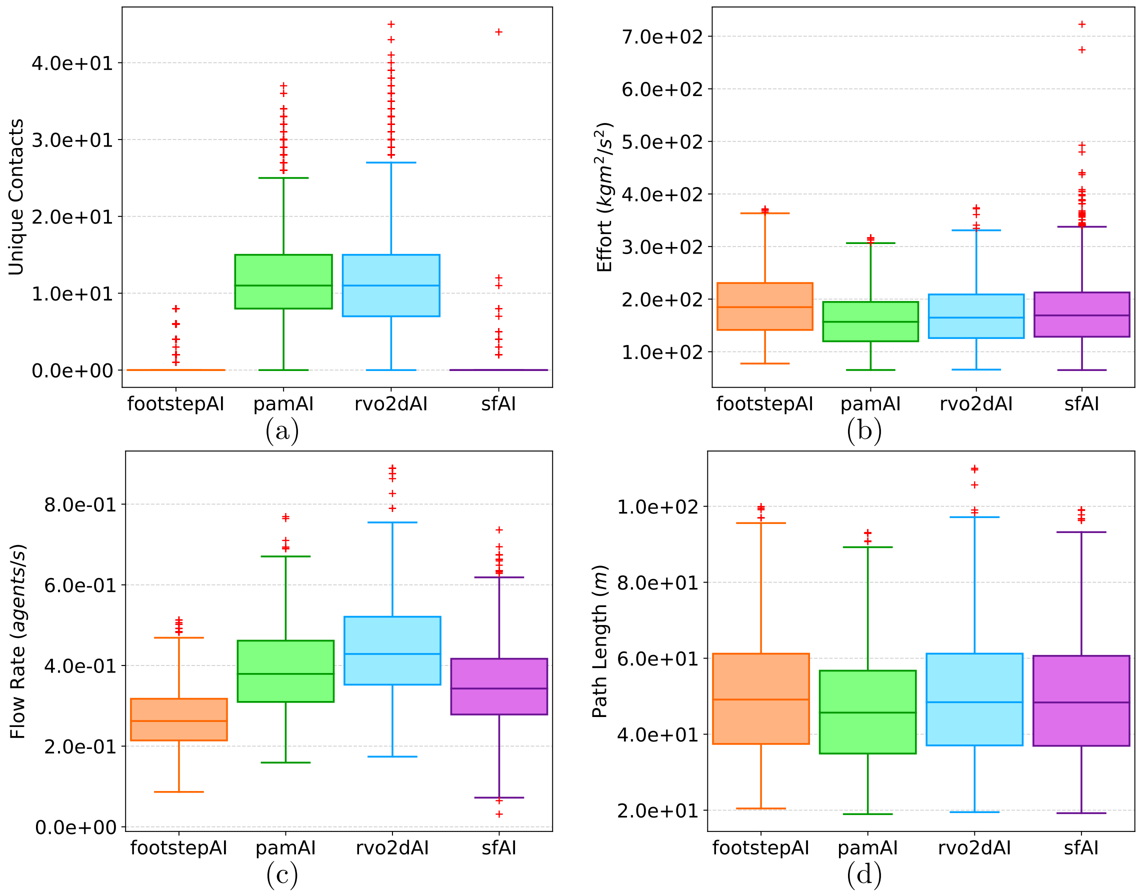


Figure 4.4. Boxplots of (a) unique contacts, (b) effort, (c) flow rate, and (d) path length statistics for the *footstepAI*, *pamAI*, *rvo2dAI*, and *sfAI* models in the forced interactions scenarios.

pamAI is significantly different from all other models ($p < 0.01$), all other combinations are not significant, see Table 4.9.

4.4.4 Discussion

In this study, I want to understand the overall performance of the agents when they are forced to interact with themselves and the environment under numerous configurations.

Table 4.6: Matrix of statistically significant results for the unique contacts metric in the Forced Interaction Scenarios. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	x	-
sfAI	✓	✓	✓

Table 4.7: Matrix of statistically significant results for the effort metric in the Forced Interaction Scenarios. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	✓	-
sfAI	✓	✓	✓

Table 4.8: Matrix of statistically significant results for the flow rate metric in the Forced Interaction Scenarios. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	✓	-
sfAI	✓	✓	✓

Table 4.9: Matrix of statistically significant results for the path length metric in the forced interaction scenarios. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	x	✓	-
sfAI	x	✓	x

Table 4.10: Total number of completed scenarios for each model in the forced interactions scenario set ($|S| = 10000$).

AI	**Completed Scenarios**	**Success Ratio**
footstepAI	9825	98.25%
pamAI	9921	99.21%
rvo2dAI	7921	79.21%
sfAI	9440	94.40%

These scenarios are relatively unchallenging in terms of crowd-level challenges (large numbers of interacting agents), but are individually challenging for the agents as they force micro interactions between agents and obstacles in a small space. Ultimately, the scenario set in this experiment is a good sampling of the state space of micro agent interactions. The results for this study tell two stories. The first lies in the source of this benchmark and the difference between its original intended use and the use made of it here. This is primarily because of two factors 1) the inclusion criteria for these studies is predicated entirely on the completion of all goals of all agents in the scenario, which differs from the original benchmark criteria and 2) $\mathbb{S}_r$ is specifically designed for egocentric metric evaluation. In particular, the path length outcomes reveal the purposefully limited definition of the $\mathbb{S}_r$ which focuses on scenarios with individual interactions instead of crowd wide dynamics. Since all agents are distributed over the small and constrained definition of t$\mathbb{S}_r$ and all models use the same high-level path planner, there is little room for deviations in paths and thus path lengths.

The *pamAI* and *rvo2dAI* models stand out for having the highest unique contacts counts with numerous outliers, while also having the highest flow rates.

Interestingly, though there are significant differences, the size of these differences between models with respect to the effort and path length measures is negligible. This is most likely an artefact of the scenario space. The scenario is rather small and is designed primarily for forcing interactions in the building blocks of larger environments, not stress testing the more complicated environments which may be made up of these building blocks. Thus, the path lengths in general are within a small range. Similarly, the same occurs within the effort measure, however *footstepAI* is offset primarily by its parabolic paths.

4.4.5 Qualitative Discussion and Failure Set

In this study, we see the beginning of a trend that holds in other parts of the larger study, *pamAI* and *rvo2dAI*, struggle with static obstacles, particularly in tight corridors with competing agents. This causes a number of artefacts such as ghosting through obstacles and agents.

The outliers for *rvo2dAI* and *pamAI* in the unique contacts and flow rate metrics are interesting. These outliers are almost exclusively caused by pushing which, in tight environments (i.e. nearby obstacles), produces model specific problems. A couple of *rvo2dAI* agents can push another agent (ie the only free velocity on velocity space is in the opposite direction) out of their way until a forward facing free velocity is found for the pushed agent. For *pamAI*, the similar issue is primarily with forced oscillations under tight environment passing conditions. As well, while *rvo2dAI* and *pamAI* both produce excellent reciprocal avoidance amongst agents in open environments (as seen in Section 4.3), they struggle with stationary obstacles. For *rvo2dAI*, this appears to be because, occasionally under competing half planes in velocity space, an *rvo2dAI* agent will find free velocities inside of the obstacle who's half plane is not well-defined. For *pamAI*, this appears to be because of the importance of balance in the definition of forces. In *pamAI*, there are agent avoidance, obstacles avoidance, goal, and predictive agent avoidance forces, however there is not predictive obstacle force. This appears to cause, under the right conditions, a net force that overwhelms the obstacle avoidance force.

The failure sets of the models in the forces interactions scenarios reveal interesting shortcomings of all the models in dense environment conditions. The *footstepAI* model does not complete scenarios if it spawns too close to an obstacle and is facing it at close to orthogonal small number or total - a situation that does not occur often. This was partially fixed by random sampling in the footstep action space and the definition of in-place turning actions (Berseth, Kapadia, & Faloutsos, 2015). However, under the right obstacle orthogonality conditions the planner never finds a solution to place the first step. The *pamAI* model, under very rare conditions, suffers from equilibrium of opposing agents in tight corridors, while *sfAI* suffers from this same problem much more often because of additional corner cases (where agents attempt to round a corner in opposite directions). These problems stem from the definition of the model. Since the models are based on the resolution of a net force, it is possible to reach a physical equilibrium between agents which never resolves. The *rvo2dAI* model violates walls (static obstacles) occasionally and this scenario has relatively large block shaped obstacles which once inside cannot be planned out of. This is because in velocity space static obstacles are not well-defined. While searching for the free velocity space, an *rvo2dAI* agent may see an obstacle in velocity space, especially at the seams between obstacle faces, as free (or open) and move into the obstacle. Once inside, the agent has no way of exiting because the agent is now within the half planes formed by the obstacle faces.

4.5 Intensifying Density

This experiment comparatively explores the performance of each model in two common sub-sections of large building designs, the bottleneck egress and bi-directional crossing groups scenario. In particular, the main independent variable of this experiment is density, over which all the study measures are captured. These two subsection designs are common in well constrained pedestrian-obstacles interaction and optimization analysis studies(Haworth, Usman, Berseth, Khayatkhoei, et al., 2017; Zhao et al., 2017; Severiukhina et al., 2017; Feng, Yu, Yeung, Yin, & Zhou, 2016; Berseth, Usman, et al., 2015; Jiang et al., 2014; Johansson & Helbing, 2007). Here, the density is intensified over small steps in the fashion of a sensitivity analysis–the outcome being readable fundamental diagrams for all the measures. Furthermore, pedestrian density can be easily mapped to Levels of Service for expected qualitative outcomes of pedestrian behaviours (Fruin, 1971).

It is important to point out that these sub-section scenarios avoid the myriad of other considerations generally required in larger evacuation simulations. These include panicked behaviours models, signage, evacuation plans, and evacuation assistants (group leaders) in a larger environment. By focusing on a well constrained sub-section of the environment, the study can focus on the impact of density on the crowd performance at egress and crossing points.

4.5.1 Material and Methods

In addition to the materials and methods outlined in Section 4.2, this experiment is performed on the following scenario set.

4.5.1.1 Scenario Set

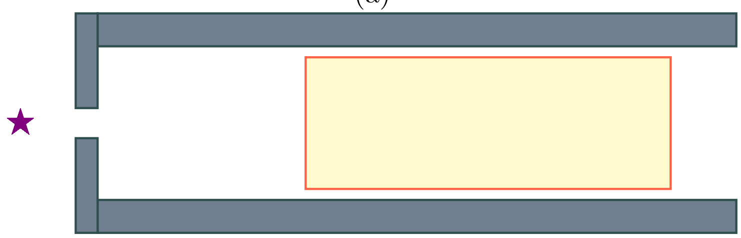
This experiment involves two sub-enviroments commonly found in built environments. The bottleneck egress scenario is a type of common uni-directional egress scenario where the egress point is smaller than the hallway leading to it. This design and its key dimensions can be seen in Figure 4.5a. The bi-directional crossing groups scenario is a scenario where two groups are passing in a hallway. This design and its key dimensions can be seen in Figure 4.6a.

Density This experiment consists of 37 samples of crowd densities from 0.2 to 2.0 agents/m^2 at steps of 0.05 agents/m^2 corresponding to 37 scenario spaces for each environment - the bi-directional crossing groups scenario and bottleneck egress scenario.

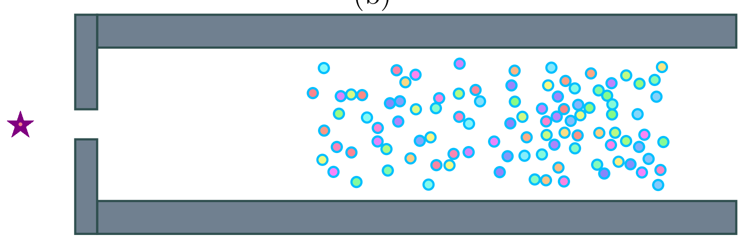
4.5.2 Analysis

Fundamental diagrams have a long history of use in vehicular traffic analysis and have been applied to pedestrian traffic in built environments (Fruin, 1971). This area of analysis has since seen a large amount of attention in the literature (Vanumu, Rao,

(a)

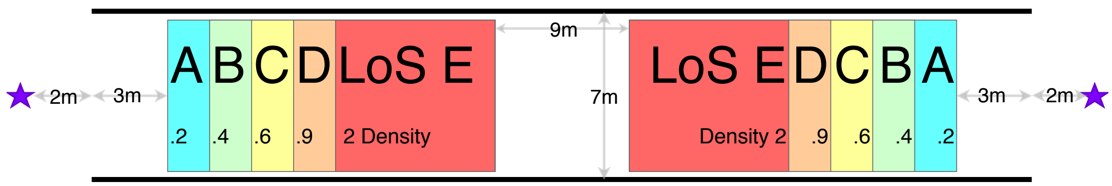


(b)

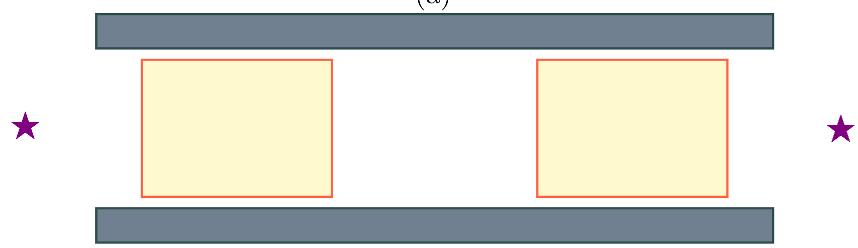


(c)

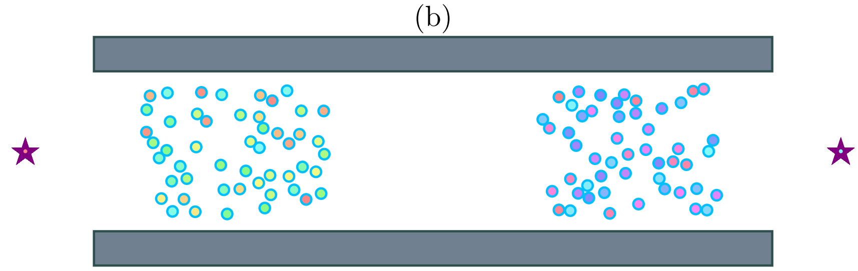
Figure 4.5. The bottleneck egress experiment from the attenuating level of service study spawns agents inside the hallway with goals just outside if a standard sized door. (a) An experiment outline showing common densities and their associated Level of Service label, as well as the dimensions of the scenario space's important features; (b) the density level 1 agents/m^2 scenario space; (c) a single corresponding scenario generated from (b). Note that (a) is a granular representative figure, this experiment consists of 37 samples of the density from 0.2 to 2.0 agents/m^2 at steps of 0.05 agents/m^2



(a)



(b)



(c)

Figure 4.6. The bi-directional crossing groups scenario from the intensifying density study spawns agents inside the hallway with opposing goals just outside. (a) An experiment outline showing common densities and their associated Level of Service label, as well as the dimensions of the scenario space's important features; (b) the density level 1 agents/m^2 scenario space; (c) a single corresponding scenario generated from (b). Note that (a) is a granular representative figure, this experiment consists of 37 samples of the density from 0.2 to 2.0 agents/m^2 at steps of 0.05 agents/m^2

& Tiwari, 2017). The proposed method extends the fundamental diagram concept by computing the study measures across densities. It is clear from the literature that exact fundamental diagram shapes are highly dependent on experimental design, measurement definition, and sociocultural factors of the pedestrians involved. This analysis therefore seeks to see how and when fundamental diagrams converge or diverge across steering models.

4.5.3 Results

The fundamental diagrams, with shaded error regions, for all four measures of the bottleneck egress scenario can be seen in Figure 4.7. The fundamental diagrams, with shaded error regions, for all four measures of the bi-directional crossing groups scenario can be seen in Figure 4.8.

4.5.4 Discussion

Bottleneck Egress Scenario. The most prominent feature of these results is the critical point, the point at which density increases produce diminishing or negative returns. This is particularly so of the *sfAI* model. This is entirely due to the models under damped net force nature which causes highly oscillatory motions (Wolinski et al., 2014b). This is most clear in the path length and unique contacts. Under this scenario's conditions all models should have approximately the same path length–there is little room for deviations. However, because of oscillations, even this metric is inflated after the critical

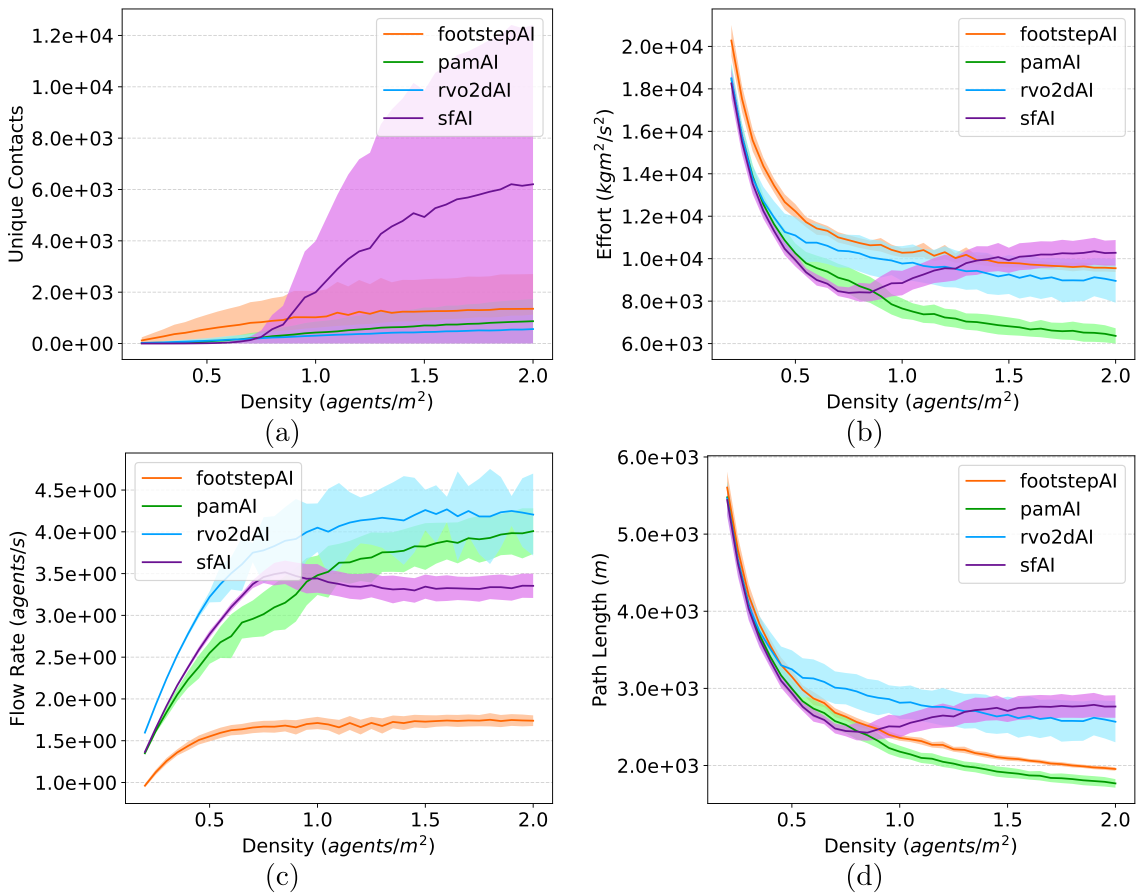


Figure 4.7. Fundamental Diagrams of (a) unique contacts, (b) effort, (c) flow rate, and (d) path length over 37 samples of increasing density [0.2,2.0] for the *footstepAI*, *pamAI*, *rvo2dAI*, and *sfAI* models in the bottleneck egress scenario. The lines and shaded areas represent the means (μ) and standard deviations (σ) respectively.

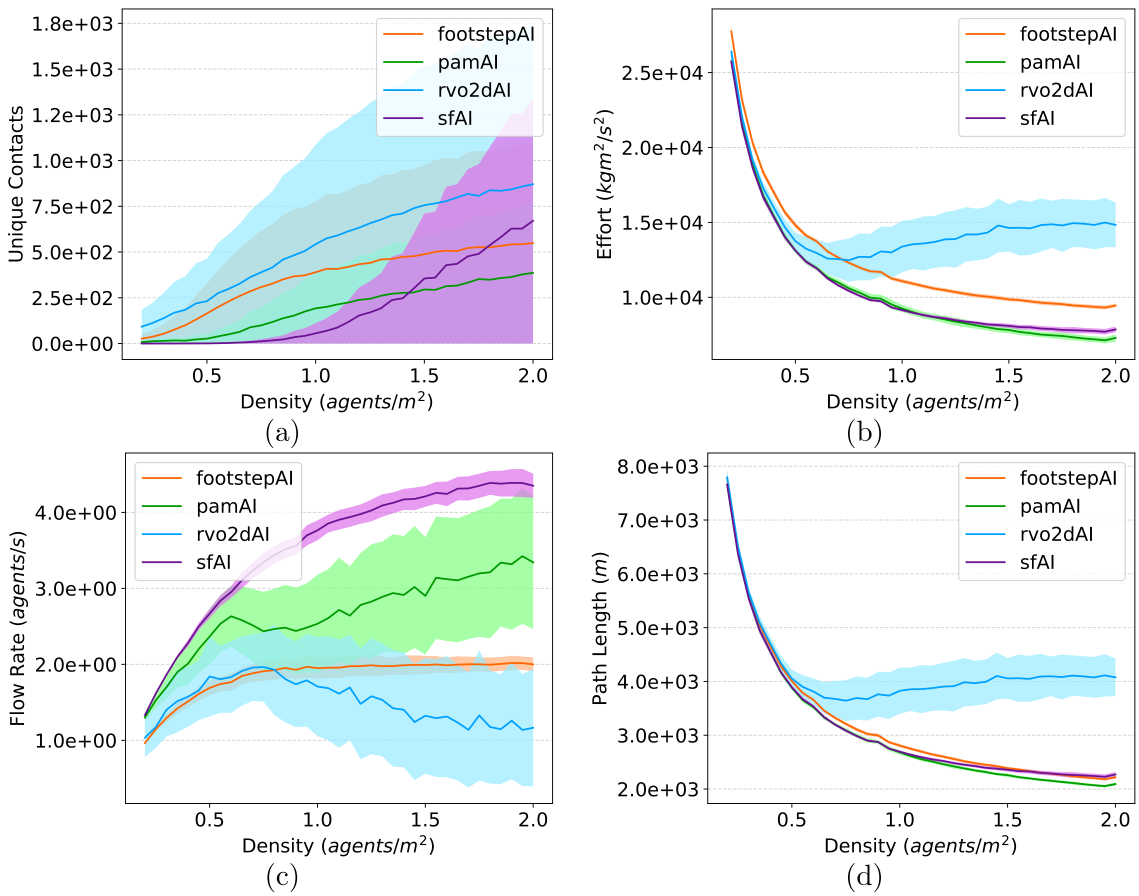


Figure 4.8. Fundamental Diagrams of (a) unique contacts, (b) effort, (c) flow rate, and (d) path length over 37 samples of increasing density [0.2,2.0] for the *footstepAI*, *pamAI*, *rvo2dAI*, and *sfAI* models in the Bi-directional Crossing Groups Scenario. The lines and shaded areas represent the means (μ) and standard deviations (σ) respectively.

point.

The effort curves for this scenario are interesting because *footstepAI*, *rvo2dAI*, and *pamAI* produce very similar curves at but at different scales, with *pamAI* performing best overall. Interestingly, *sfAI* is the top performing model in this scenario up to its critical point for both path length and effort. This is because in unhindered environments *sfAI* tends to maximize its desired speed forward, via non-oscillatory pushes. Despite outperforming in terms of effort, path length, and flow-rate, *rvo2dAI* is outperformed by *footstepAI* in terms of path length. This is a reflection of *footstepAI* fidelity, which is discussed further in the qualitative results.

The flow rate curves show some interesting outcomes behaviours. The *footstepAI* model produces the lowest flow rate and plateaus early in the density sampling. The *rvo2dAI* and *pamAI* models reach their critical points around the expected value for such a scenario w.r.t. to pedestrian literature but resulting in much higher flow rates. In contrast, the *sfAI* model reaches its critical density far before a human crowd would.

Bi-directional Crossing Groups. The collision curve is surprising in that it shows that the *rvo2dAI* has difficulty resolving high density oncoming groups without colliding.

Additionally, like the *sfAI* model in the bottleneck egress scenario, the *rvo2dAI* model struggles with all metrics beyond the density 0.9 agents/m^2 though not to such extremes.

Finally, the path lengths of *footstepAI*, *pamAI*, and *sfAI* are nearly identical across densities. This is a fascinating result because the qualitative results reveal strong differences in the emergent behaviour fo the crowds.

4.5.5 Qualitative Discussion and Failure Set

In this study it should be made clear that no particular behaviour which fundamental diagrams may measure is necessarily "correct" and/or representative of an environment's capacity. Perhaps Vanumu et. al. (Vanumu et al., 2017) explain this best:

> ..it may not be appropriate to present the [fundamental diagrams] developed for different conditions in a single diagram as they vary for various flow situations and also for different infrastructural elements. Personal and environmental characteristics influence the pedestrian motion. Moreover, capacity of the system depends on self-organization phenomena that includes zipper effect, oscillations at narrow bottlenecks, lane formations in uni-directional and bi-directional flows and so forth.

In this section, I am not examining capacity of the system (the surrounding environment design) but rather the emergent policies, behaviours, and shortcomings that result from the underlying steering model choice in a single well constrained scenario. Thus, I attempt to control for these concerns regarding the use of fundamental diagrams in comparative analysis.

Bottleneck Egress. This is a difficult scenario because the bottleneck exacerbates high density conditions. That is, after a certain spawning density the crowd no longer flows freely through the bottleneck, it causes a rapid increase in the local density near the bottleneck which the models must resolve somehow.

It is clear that the $sfAI$ model struggles to avoid unique contact conditions as it rapidly oscillates to solve the scenario, while the other models perform similarly well. In fact, beyond the density 0.9 agents/m^2 social forces performance rapidly declines across all metrics, primarily due to its oscillatory behaviour near the bottleneck point. It should be noted that these oscillations likely far exceed that of what a real crowd would produce. Without a ground truth it is difficult to concretely and quantitatively say so, but in these cases the movement is unnatural looking and occurs at very high frequencies amongst individual agents.

The flow rate metric, as well, reveals some very interesting behaviours in the models. The *footstepAI* flow rate plateaus early partially because of tighter packing available to the agent design and more importantly the sidestepping and in-place turning behaviours near the bottleneck. In cases like these, the step plans becomes like an under damped spring oscillating between side-stepping and moving forward in order to maintain the goal direction while letting others through. This non-holonomic form of oscillation appears qualitatively to be more like those oscillations found in real crowds experiments, unlike those present in *sfAI* (save for the modelling of normative versus panicked crowds respectively). The *pamAI* and *rvo2dAI* model performances are partially due to the ordered nature that predictive models produce. However, under high density conditions these particular models occasionally clip or ghost through the edges of obstacles like the walls forming the egress passageway–this is likely due to the issues with the prediction of zero-velocity obstacles. Generally, all the models produced approximates of same behaviours reaching some critical point and grouping near the bottleneck, but each model's performance differentiates with the scale of severity. This qualitative result is nicely reflected in the quantitative results.

The interesting effects of having the worst flow rate and effort, but good path length in the *footstepAI* model are reflected in the qualitative results. To resolve bottlenecks *footstepAI* agents will sidestep and allow others who have precedence to pass. Also, all other models have no real notion of free space–only *rvo2dAI* does, but this free space is in velocity space. The *footstepAI* model on the other hand will extend planning nodes into free space to find low energy path alternatives. So while other models will only compact, *footstepAI* will balance both compacting and the use of free space.

Bi-directional Crossing Groups. This experiment is important for understanding group-group interactions in a controlled environment. In real world crowds, people under similar conditions tend toward laminar flows and vortices as energy efficient solutions to counter flow scenarios like this one (Still, 2000). In this experiment, it is expected models will favour one of these strategies over the other, or succumb to a highly noisy collapse of both groups into one.

The *rvo2dAI* model favours highly laminar flow however struggles with finding free velocity space when multiple oncoming agents move towards a single agent and the enclosing spaces create no goal advancing free velocity space. In these cases, the best free velocity is behind the agent. This results in what appears to be pushing behaviour where agents will move backward until the velocity space becomes clearer of velocity obstacles. Unfortunately, the *rvo2dAI* model under higher density conditions violates the upper and lower hallway bounds to resolve a free velocity. In these cases, agents may get pushed completely through the wall boundaries and resolve the scenario from outside the hallway.

The *sfAI* model, surprisingly in contrast to its previous results, favours highly laminar flow with very thin (usually single to three agent wide) lanes. However, the agents still oscillate in somewhat unnatural looking ways near the lane boundaries. Otherwise, the *sfAI* model performs in a very ordered many before and after the group crossing points. In these cases, the model is only resolving agent-agent interactions and since the velocities are homogeneous here, the resolution is straightforward.

The *pamAI* model favours laminar flow with very thick lanes (three or more agents wide) that begin to form before the groups meet–a sign of the low-level predictive behaviour producing emergent behaviours in the larger crowd. These larger groups can lead to the pushing of a few agents, similar to the *rvo2dAI* model behaviour but on a less extreme scale. The group crossing is bounded by two large walls and these generate oscillations in the *pamAI* model, which has no predictive force for static obstacles.

The *footstepAI* model makes seemingly noisy early step decisions towards laminar flow, similar in pattern to *pamAI*, but with thinner lanes similar to *sfAI*. However, during the lane formation, as the groups begin to merge, footsteps resolves some difficult areas with micro scale vortices of just a few agents. In this way, *footstepAI* performs a mixed approach to the scenario resolution.

The interesting artefact in the path length metric outcome with respect to *rvo2dAI* appears to be due to the fact that the *footstepAI*, *pamAI*, and *sfAI* models successfully perform some variation of the laminar flow strategy. Despite the fact that these strategies take on different qualities, as described above, the path length measure remains nearly matched for these models across all densities. The deviation of the *rvo2dAI* model after approximately the 0.65 agents/m^2 critical point is due to the pushing artefact and obstacle ghosting caused by high densities.

4.6 Egress with Concavities

This experiment comparatively explores the performance of each model in an egress scenario. As noted in Section 4.5, examining evacuation scenarios at the building scale usually involves simulating the impact of signage, behavioural characteristics (like panic and confusion), and leadership (both intentional, such as evacuation leaders, and unintentional, such as group following behaviours). Here, I use a full scale building evacuation scenario to test only the root simulation method controlling local steering and collision avoidance behaviour without these higher layers. In this experiment, the nature of the scenario is meant to elicit the spectrum of difficulties crowd steering models encounter in egress scenarios. The egress itself is simple, a single exit and large connected hallways, but the environment includes both regular and irregular concavities, a static obstacle, and a 90 deg turn prior to the egress point.

4.6.1 Material and Methods

In addition to the materials and methods outlined in Section 4.2, this experiment is performed on the following scenario set.

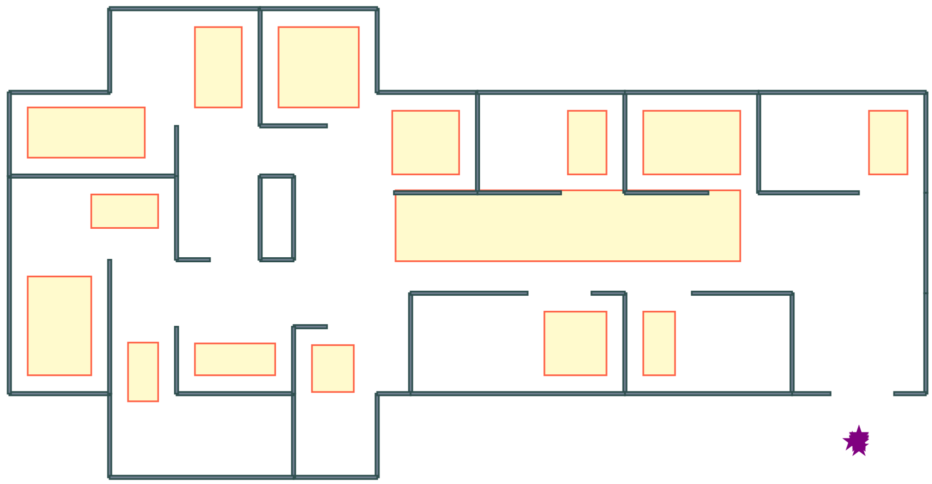
Scenario Set. In this study the environment is a simplified version of a portion of the West Building of the National Gallery of Art. The scenario goal is an evacuation of groups of patrons viewing various exhibits or 'mulling' together. Both the viewing and 'mulling' patrons are initialized with random orientations in the agent regions. The scenario space is initialized and simulated over 200 scenarios.

4.6.2 Analysis

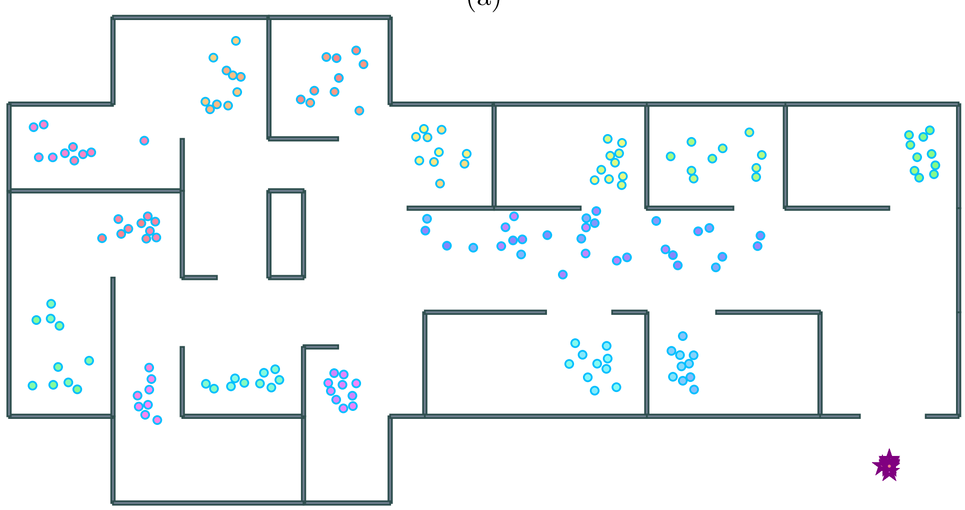
The summary statistics (median, first quartile, third quartile, IQR, max, min, and outliers) of each of the four metrics are computed over the intersection of all completed scenarios. All measures are tested using the methods outlined in Section 4.2.4 on the intersection of all completed scenarios.

4.6.3 Results

Boxplots of unique contacts, effort, flow rate, and path length statistics in Figure 4.10. There is a significant difference present amongst the models for each measure ($N = 63$, $p < 0.01$). The battery of post-hoc tests for the unique contacts measure reveal there is no significant difference between $pamAI$ and $rvo2dAI$, but all other combinations of models are significant ($p < 0.01$), see Table 4.11. All pairwise comparisons of steering models for effort are significantly different ($p < 0.01$), see Table 4.12. For the flow rate measure, there is no significant difference between $pamAI$ and $rvo2dAI$, but all other combinations of models are significant ($p < 0.01$), see Table 4.13. For the path length measure, there is no significant difference between $pamAI$ and $sfAI$, but all other combinations of models are significant ($p < 0.01$), see Table 4.14.



(a)



(b)

Figure 4.9. The Egress with Concavities scenario based on a simplified version of a portion of the West Building of the National Gallery of Art. a) The scenario space and b) an example evacuation scenario generated from (a).

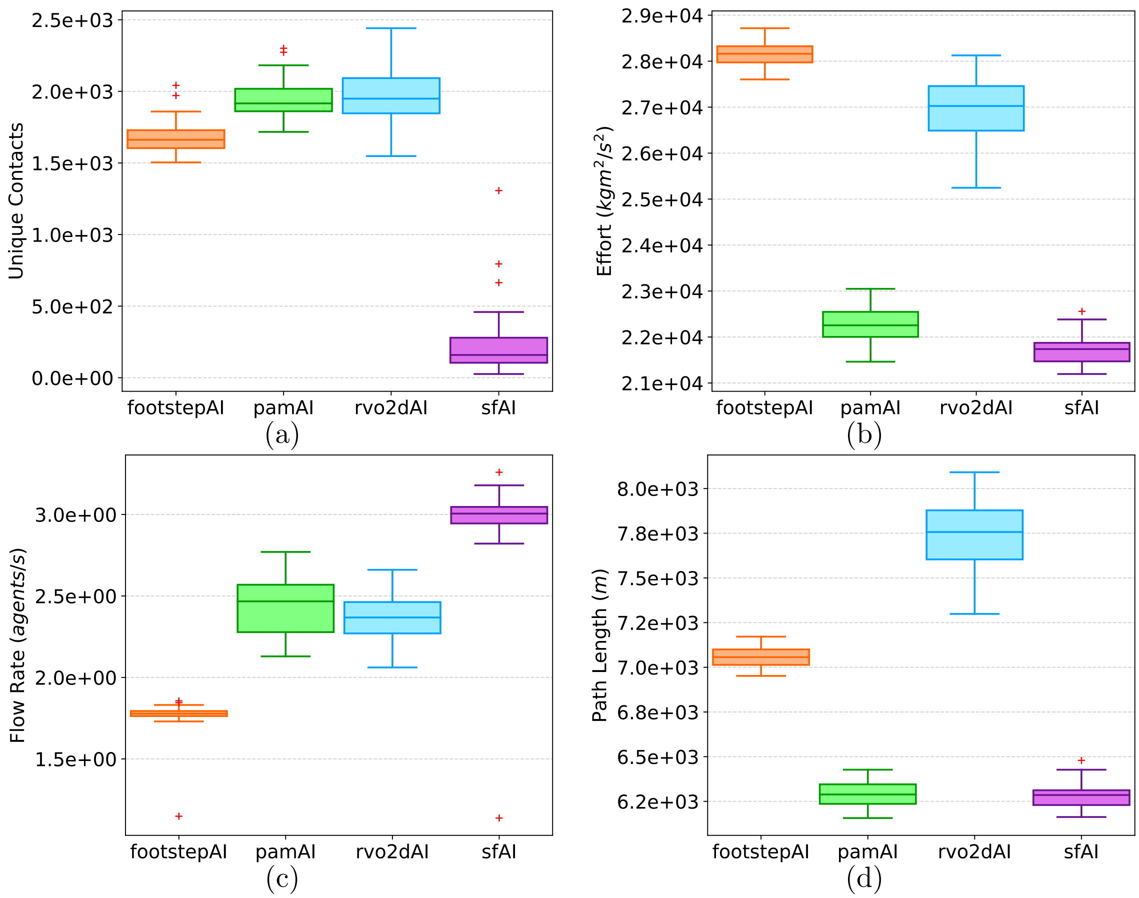


Figure 4.10. Boxplots of (a) unique contacts, (b) effort, (c) flow rate, and (d) path length statistics for the *footstepAI*, *pamAI*, *rvo2dAI*, and *sfAI* models in the egress with concavities scenario.

Table 4.11: Matrix of statistically significant results for the unique contacts metric in the egress with concavities scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	x	-
sfAI	✓	✓	✓

Table 4.12: Matrix of statistically significant results for the effort metric in the egress with concavities scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	✓	-
sfAI	✓	✓	✓

Table 4.13: Matrix of statistically significant results for the flow rate metric in the egress with concavities scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	x	-
sfAI	✓	✓	✓

Table 4.14: Matrix of statistically significant results for the path length metric in the egress with concavities scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	✓	-
sfAI	✓	x	✓

Table 4.15: Total number of completed scenarios and the corresponding success ratio for each model in the egress with concavities scenarios ($|S| = 200$).

AI	**Completed Scenarios**	**Success Ratio**
footstepAI	172	86.00%
pamAI	187	93.50%
rvo2dAI	77	38.50%
sfAI	185	92.50%

4.6.4 Discussion

This is a fascinating scenario not only because it resembles a real world environment, but also because of its size, rooms, and evacuation-like nature. In practice, this differentiates all algorithms in all metrics by scaling what are otherwise seemingly small differences in performance.

A somewhat unexpected result is the performance of the *sfAI* model across all measures. This is the result of previously discussed force based artefacts but scaled by the environments size and adherence to the conditions that produce those artefacts. These are namely: 1) single directional flow produces few if any collisions and maximizes desired speed; 2) the absence of closed space bottlenecks avoids the oscillatory behaviour; and 3) force models have highly linear paths in unhindered environments.

It appears from the quantitative results that linear collision prediction may have some impact on flow rate and unique contacts in environments such as these. While *pamAI* and *rvo2dAI* flow rate measures are significantly different, the difference between them is much smaller than their differences with the other models. Similarly, as noted above with *sfAI*, force based models have some relation to the minimization of path length and effort in environments like these.

4.6.5 Qualitative Discussion and Failure Set

This scenario forces a long high density stream of agents to navigate an environment with several concave sub-environments towards a singular egress point. The layout of the environment poses particular challenges to synthetic crowd simulators. The numerous concavities present (rooms, exhibits, and corridors), the length of the environment, and the cornering required immediately before the egress reveal certain limitations and behaviours in the models. Concavities are particularly challenging to synthetic crowd models. For a simulator to effectively handle concavities there must be higher level path planning - usually referred to as long-term path planning, as in Chapter 1. This type of path plan provides an agent with short-term goals which ideally will lead the agent out of any concavity. Otherwise, it is highly likely that agents will simply move towards the enclosing wall in the direction of their environment goal and either remain stuck there, slide out of the concavity, or slide towards a concave vertex in the solid environment geometry. This experiment revealed a deep limitation of the default long-term planning algorithm in the SteerSuite simulation system. The breadth-first search algorithm used produces unrealistic paths into concavities and must be carefully post-processed to fix this. In this study, all experiments instead used the same A* shortest path algorithm with the l^2-norm heuristic, the furthest visible waypoint culling, and local target lost long-term replanning (Kallmann & Kapadia, 2016; Hart, Nilsson, & Raphael, 1968). This ensures all models have an optimal shortest path even in adverse conditions.

There are certain fascinating artefacts in the quantitative outcomes, with respect to outcomes in the previous benchmarks, that mirror qualitative differences. The first is that the *sfAI* model incurs the least unique contact conditions. Much of what ails *sfAI* is dependent on conflicting interactions, such as head on head resolutions where *sfAI* agents simply collide and eventually slide past each other. In a long, mostly uni-directional scenario with no enclosed bottleneck, the *sfAI* model produces highly ordered flow–in contrast to its 'panicked' model derivation. As well, all models successfully reproduced the 'corners as a bottleneck' phenomenon which shows crowds fundamentally do not move like fluids (Still, 2000).

Interestingly, the *footstepAI* and *rvo2dAI* models struggle with the corner bottleneck. This is primarily because the long-term path planning places a waypoint at the corner for all agents (and all models since they are independent). The *footstepAI* model appears to struggle to balance the fact that there is a large amount of open space (this is a corner bottleneck) and the need to get close enough to the waypoint to "complete" it. The *rvo2dAI* algorithm struggles for a similar reason but the effect is more akin to pushing, in that the free velocity space is in the open space of the corridor and agents get "push" past the local waypoint.

The failure set reiterates the common thread for issues in each model. The *footstepAI* model only fails to complete scenarios when very particular spawning conditions occur. A review of the *footstepAI* model failure set shows this only occurs with one or two agents out of the total one-hundred. The *sfAI* model fails when two or more agents reach equilibrium at a corner or a local target. The *pamAI* and *rvo2dAI* models fail when an agent which ghosted an obstacle remains stuck within the obstacles. This scenario contains a relatively sizeable column (with negative space) central to several egress routes and proved challenging for these models.

4.7 Diametric Environment Goals

This experiment comparatively explores the performance of each model in a worst-case, non-egress use of an environment with a complicated concavity structure. Similar to the Egress with Concavities scenario in Section 4.6, this experiment explores the applicability and performance of algorithms under the extremes of use. In this experiment, the agents and their goals are initiated as diametric environment goals. This approach is meant to model an extreme of normal traffic use of a building, i.e. needing to traverse from one space to another in a non-emergency scenario, under the worst case, i.e. all spaces need to swap occupants at the same time. Additionally, the scenario environment is made up of concavities of varying size connected by thin hallways.

It is important to point out that these scenarios do not utilize the myriad of other considerations in evacuation simulations. These include panicked behaviours models, signage, alarms, evacuation plans, evacuation assistants (group leaders). These scenarios are provided as an example of a particular case taking place in a built environment design, focusing on only the performance of the underlying steering models.

4.7.1 Material and Methods

In addition to the materials and methods outlined in Section 4.2, this experiment is performed on the following scenario set. **Scenario Set** In this study the environment is the York University Albany Road building, a real world built environment. The scenario goal is the simultaneous diametric traversal of groups of agents. All agents are initialized

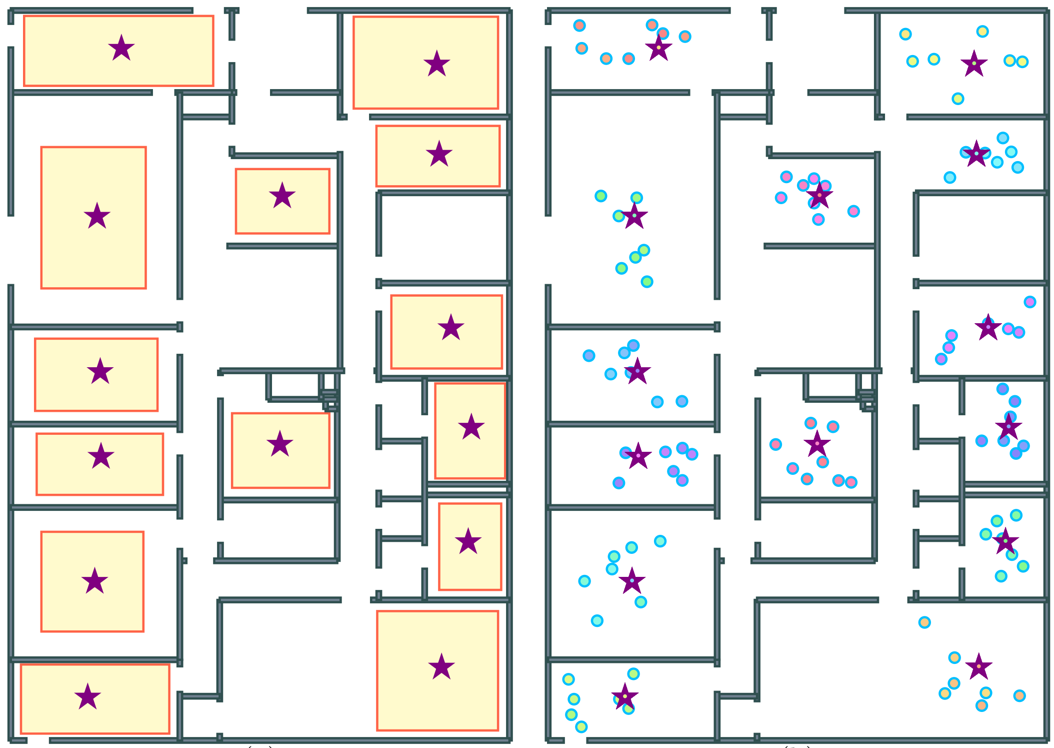


Figure 4.11. The York University Albany Road Building diametric environment goals scenario. a) The scenario space and b) an example diametric groups scenario generated from (a).

with random orientations in the agent regions. The scenario space is initialized and simulated over 200 scenarios.

4.7.2 Analysis

The summary statistics (median, first quartile, third quartile, IQR, max, min, and outliers) of each of the four metrics are computed over the intersection of all completed scenarios. All measures are tested using the methods outlined in Section 4.2.4 on the intersection of all completed scenarios.

Table 4.16: Matrix of statistically significant results for the unique contacts metric in the diametric environment goals scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	✓	-
sfAI	✓	✓	✓

Table 4.17: Matrix of statistically significant results for the effort metric in the diametric environment goals scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	✓	-	-
rvo2dAI	✓	x	-
sfAI	✓	✓	✓

4.7.3 Results

Boxplots of unique contacts, effort, flow rate, and path length statistics in Figure 4.12. The completion rates of each model are reported in Table 4.20. There is a significant difference present amongst the models for each measure ($N = 55$, $p < 0.01$). The battery of post-hoc tests for the unique contacts and effort measures reveal all combinations of models are significantly different ($p < 0.01$), see Tables 4.16 and 4.17 respectively. For the flow rate and path length measures, there is no significant difference between *footstepAI* and *pamAI*, but all other combinations of models are significant ($p < 0.01$), see Tables 4.18 and 4.19 respectively.

Table 4.18: Matrix of statistically significant results for the flow rate metric in the diametric environment goals scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	x	-	-
rvo2dAI	✓	✓	-
sfAI	✓	✓	✓

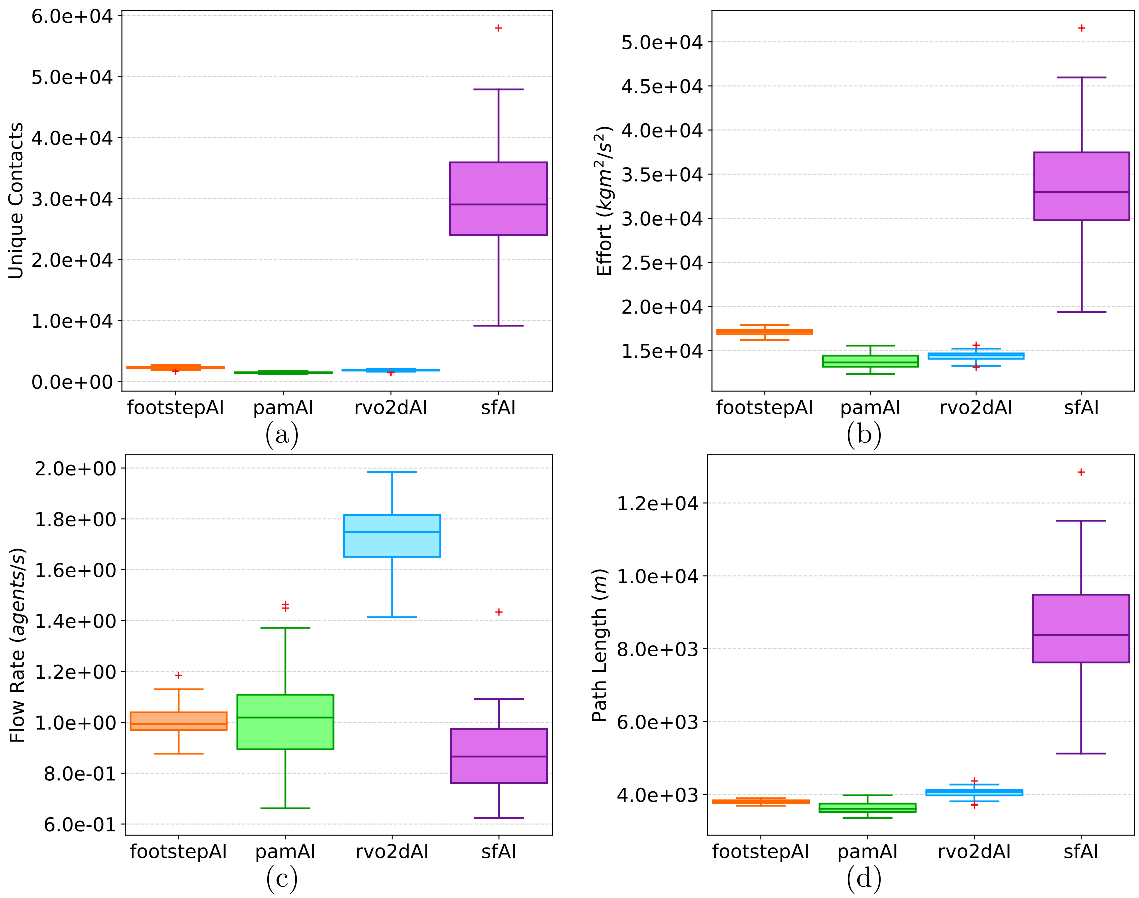


Figure 4.12. Boxplots of (a) unique contacts, (b) effort, (c) flow rate, and (d) path length statistics for the *footstepAI*, *pamAI*, *rvo2dAI*, and *sfAI* models in the diametric environment goals scenario.

Table 4.19: Matrix of statistically significant results for the path length metric in the diametric environment goals scenario. All check marks indicate significance as defined in Section 4.2.4.

	footstepAI	*pamAI*	*rvo2dAI*
pamAI	x	-	-
rvo2dAI	✓	✓	-
sfAI	✓	✓	✓

Table 4.20: Total number of completed scenarios for each algorithm in the diametric environment goals scenarios ($|S| = 200$).

AI	**Completed Scenarios**	**Success Ratio**
footstepAI	154	77.00%
pamAI	196	98.00%
rvo2dAI	79	39.50%
sfAI	183	91.50%

4.7.4 Discussion

In this scenario, the *sfAI* model stands out across all metrics for having the most unique contacts, expending the most energy, having the lowest flow-rate, and the longest paths. This scenario proves to be particularly challenging for the model.

The *rvo2dAI* model stands out as well for having a particularly high flow rate in a scenario where all other models do not perform well. Indeed, this scenario, above all others, has forced the all other models in the set into likely flow-rates for the type of scenario it represents (Fruin, 1971), while the *rvo2dAI* model falls within a normative range it is still far outperforming expectations.

4.7.5 Qualitative Discussion and Failure Set

The *sfAI* model's performance is particular low for this scenario. This is primarily because this scenario forces a variety of crowd contexts, most importantly a small high density corridor with multiple crossing groups of agents. In this context the *sfAI* begins to both oscillate and reach a sort of equilibrium where agents struggle to pass each other–combining the detrimental effects of other benchmarks at a larger scale. The result is singular agents making it through the corridor after oscillating for some time.

Here, the *footstepAI* model excels in terms of effort and less so in terms of path length. The model is able to resolve particularly high density contexts in complicated spaces. The *footstepAI* and *sfAI* models both have low flow rates but for very different reasons. The *sfAI* model has the issue as noted above but the *footstepAI* agents resolves the tight corridor crossing groups issue by sidestepping either in the thin hallways or stepping into rooms–what we might think of as ducking out of the way of a crowd. The other models, however, only achieve this behaviour through pushing and, in a non-emergency context such as this, the behaviour looks very unnatural.

The failure set reiterates the common thread for issues in each model. The reasons for failure are much the same as in Section 4.6.5, however, many of these issues, particularly for the *rvo2dAI* model, are exacerbated by the long and (relative to length) thin walls in the environment. Some spawning regions are purposefully room filling, this exacerbates the *footstepAI* model initial step spawning failure problem.

4.8 Discussion

A common issue found at different scales throughout this benchmark is particle sliding. Particle sliding is a larger general problem of particle-based crowd simulation. A particle-based steering decision may make unrealistic or biomechanically impossible movements because the particle model only represents position and radius and many models take a holonomic approach to movement. This causes outcomes like path length, unique contacts, and effort to be unrealistically inflated.

Force-based models can provide naturalistic behaviours with low computational cost, however they can also suffer from severe oscillations, or unstable cyclic movement. While oscillations are natural in crowds, especially at bottlenecks, these models can become unstable and unnatural in appearance due to holonomic particle sliding and very high frequency vibrations. However, as evidenced in this study, force-based models perform very well, both quantitatively and qualitatively, in primarily uni-directional scenarios. Though in unobstructed environments, heterogeneous parameters are required to reduce an unnatural appearance of uniformity and order, depending on the context (e.g. marching versus civilian crowds).

One approach that reduces oscillation and sliding is prediction. Both force-based and geometric-based models that layer the model with even simple prediction, like line-disc intersection for prediction, achieve excellent results (van den Berg et al., 2011; Karamouzas et al., 2009). These, however, may still fail in other ways if they do not account for obstacles carefully in their predictive forces. That is, predictive avoidance control signals (like additive force or velocity) may overcome obstacles avoidance controls or still produces oscillation at the obstacle boundary. The outcome in the best situations may create agent ghosting or clipping through walls, in the worst case the agent may become trapped inside of obstacles because of the inversion of the control signal. This issue can be largely mitigated by using physical collision constraints and resolution strategies such as those found in physics engines. The low computational overhead and use of physics engines as a corrective measure makes these models particularly well suited to games (Champandard, 2012).

Planning models, like predictive models, are particularly well suited for avoiding oscillations, step-like or jagged movement, and particle sliding. Since all future movements are planned in advance, the model can impose constraints which avoid such issues.

In addition to this, biomechanics based planning models may also avoid particle sliding and unnatural movement while affording complex interactions such as side-stepping for oncoming traffic. This problem is partially eliminated with predictive models, but these do not naturally handle side-stepping or turning, which is inherently piecewise (composed of multiple steps). Multi-particle biomechanics based planning models provide another advantage over single particle models when fidelity is needed. These models afford tighter packing, as the occupied space of the agent is represented by multiple smaller particles rather than one larger particle. As well, the foot particles can be used in space-time planning at the footstep level. Thus, steering decisions may be resolved at a very high granularity, affording high fidelity in complex scenarios. This also allows for interesting metrics such as accurate step counts which is very useful in safety-critical design applications such as hospitals, care facilities, correctional, and high security facilities.

4.9 Conclusion

This chapter has sought to both understand and reinforce the importance of biomechanics in the founding of synthetic crowd simulator steering models. To achieve this, the largest ever comparative crowds analysis benchmark was devised and undertaken. This method proved useful for several reasons. First and foremost, the benchmark solidifies the cases where steering models may fail to behave in a valid or accurate manner. Second, the use of quantitaive and qualitaitive matching of artefacts helps isolate the reasons behind issues such that they may be fixed, studied, or mitigated. Finally, the simulation results from the benchmarks provides a large amount of data for both future study and training in crowd simulation.

It is clear across the benchmarks that no singular benchmark or model is panacea. Together, these benchmarks and their results, both quantitative and qualitative, provide a strong basis on which to make decisions regarding both "ecologically valid" benchmark scenarios and applicable models. These results also help us predict the behaviour and outcomes of a particular model for a particular use case. Since we now know the combination of environment layouts and features with goal locations and models we can make informed assumptions about how a model will perform. This form of testing allows a non-expert user to make more informed decisions when seeking to apply synthetic crowds. Because of this, future works include creating an online benchmarking portal including the open source simulation suite (SteerSuite) modified to reproduce these results, with all included benchmarks and scenario files. In this way, the broader community may continuously build on rigorous analytics and afford better informed decisions amongst the users of synthetic crowds. This chapter also serves towards rectifying some extreme perceptions among users and stakeholders in various fields that synthetic crowds are either, on one hand, panacea or, on the other, not useful at all. Specifically, some models are very well suited to real time applications, scenarios where the modelling of panic is important, or high fidelity analysis, and where one excels others may not.

Another prominent result of this study is the importance of biomechanics in both fidelity and the resolution of difficult problems. In terms of fidelity, particle sliding and generally holonomic movement assumptions in steering models is a deeply important problem that affects the viability and efficacy of synthetic crowds simulations. Additionally, under ideal scenario conditions, single particle non-biomechanical models tend towards highly uniform and ordered motions that appear unnatural in some conditions. It is known that differentiating velocity even a little on the motion profile of agents induces a perception of heterogeneity (McDonnell et al., 2008). By tracing the non-linear path of the centre of mass, biomechanical models may not necessarily need to induce heterogeneous parameters to produce this result. In this same vein, in single particle non-biomechanical models it is difficult or impossible to model disordered gaits, interesting mobilities, or mobility aids without an additional higher-level model or reliance on velocity scaling as a proxy to this form of heterogeneity. This is flagged as a very important drawback of non-biomechanical methods, and the following chapters explore the importance of this issue further.

Chapter 5

Importance of Heterogeneity

This chapter examines the impact of biomechanical heterogeneity in the steering level of synthetic crowd simulations on various crowd outcomes. In particular, the chapter takes a hierarchical approach to examining the impact of these heterogeneities by examining the individual components of steering. These are the action, collision, and crowd spaces, which are described in detail in their respective studies in Sections 5.3.2, 5.4,& 5.5 respectively.

It is important to note that in much of the synthetic crowd simulation research there are two primary proxies used for heterogeneity. The first is velocity as a proxy for different walker types or as a means to diversify desired velocity to give the *appearance* of heterogeneity. The second proxy is particle diameter. As noted in Chapter 2, this dissertation deals mostly in the realm of synthetic crowd simulation which is primarily achieved using particle based multi-agent simulations. Thus, particle diameter is meant to serve as a proxy to body types. It is important to make a statement about scope here because, while this dissertation endeavours to prove the importance of locomotion biomechanics and heterogeneity, I use a limited case that is well studied in medical literature as a counter example to homogeneity and to heterogeneity proxies for locomotion. This case is limited purely to the locomotion biomechanics without reference to body types. Specifically, the focus is placed on a mobility disorder that is prevalent in elderly care institutions and hospitals, where safety-critical scenarios like evacuations present the most likely dangers.

This dissertation primarily examines the heterogeneity portion of the problem from a perspective that if the seminal velocity proxy fails to be representative even when well controlled for, then by extension the diameter proxy fails to be representative. While research in the interaction between body types and locomotion is in the same vein of this work, it is outside the scope of this particular dissertation. This work is earmarked for important and critical future work. It is equally important to note that this is not to say differing body types are disabled–reiterating the social model here that our built environment and societal organization is not designed to equitably serve the spectrum of body types, disabilities, or generally marginalized communities. Additionally, there are broader areas of shoes types, assistive devices, mobility aids, and prosthetics which are directly related to this dissertation but also outside of its scope and are earmarked as critical future work.

5.1 Background

In the literature, the scope of issues regarding heterogeneity has a range, or spectrum, of impact. At one end there is the absence of heterogeneity. In this area, methods are mostly reported as is and things like heterogeneity are not explicitly tested. In many of these cases, and in fact, in most steering models, the velocity proxy for heterogeneity is a functional one not meant as an actual proxy. For example, in the *pamAI* model, a noise force is added to the net force to reduce the occurrence of deadlocks (equilibrium of opposing net forces from multiple agents) and add variation to the avoidance strategies (Karamouzas et al., 2009). It is noted however, that more 'realism' can be introduced by varying anticipation coefficients or desired speed. Similarly, other algorithms adopt adaptive approaches to desired velocity changes which can induce between group heterogeneity, such as density adaptive desired velocity manipulation (Guo, Wang, & Wang, 2009), clustered velocity potentials (Sud et al., 2007). Other algorithms, embed velocity as a heterogeneity proxy in higher level models. As noted in Chapter 1, synthetic crowd simulation is typically a layered endeavour with behavioural layers near the highest levels. Similarly, it is possible to use personality modelling to induce heterogeneity in the crowd (Guy, Kim, Lin, & Manocha, 2011). However, as the principal components of the personality model reveal, the main contributing factor in the transformation matrix from personality model to steering parameters is radius and speed. A recent approach, focuses on overtaking and passing clearance as the primary interaction factors between normative walkers and walkers with disability (Stuart et al., 2019). Ultimately, this still maps to speed reductions or increases, and in doing so the method lumps people with visual impairment, people using canes, wheelchairs, and motorized wheelchairs, into the same model–a model predicated on giving wide berth to those with disabilities.

Beyond steering model generation, the assumptions of proxies and language gain more potential to do harm. Nearer the centre of this spectrum are studies which, as noted above with my own work, are attempting to reduce the size of the independent variable space and in doing so create studies of limited scope that may later be misinterpreted as having broader scope by non-experts. The studies typically derive a model extension or a model based study to then derive findings which use crowd simulators. A seminal and insightful paper in this area, forming the bridge between modelling and safety-critical application of multi-agent synthetic crowds in the literature, showed that panic can be modelled with a relatively straightforward weighting of a desire to follow an individual direction or a group direction (Helbing et al., 2000). This study is done in two parts, first over changing desired velocity (similar to a fundamental diagram) and then with heterogeneous panic weights, both for a crowd with heterogeneous particle radii. In both cases, because the weighting factor impacts the direction, velocity and radius are still ultimately proxies for heterogeneity. More problematic uses of this make concrete prescriptive statements about design and policy based on presumptive models of the intended users. In particular, the BUMPEE model, which modulates only velocity and radius has been used to make statements about crowds of users with several modes of locomotion (Koo, Kim, & Kim, 2012). The main problem here is that biped and wheelchair control differs in its fundamental control space. Normative biped control can, to some degree, be estimated by holonomic (controllable degrees of freedom match total degrees of freedom) particle models, however differential drive control (such as wheelchairs) are non-holonomic. It is possible to make mixed holonomic/non-holonomic simulators but not by simply attenuating the desired velocity magnitude (speed) and radius of an agent. That is, wheelchairs do not simply move slower, they move differently and this difference is likely very important in safety critical-scenarios (Geoerg, Schumann, Boltes, Holl, & Hofmann, 2018; Shimada & Naoi, 2006). Similarly, in context of the Hajj, several simulation-based studies have been proposed, and in some cases used in practice, to facilitate changes in the environments involved in the pilgrimage. An example of this, closely related to this study, breaks up the levels at which actions may impact emergent behaviours. By looking at low level actions of individuals the authors derive a large scale simulator. However, the approach uses a "Mobility Factor" which models the "impaired section of the population" with up to 30% reductions in speed (Gwynne & Siddiqui, 2013). Additionally, the paper notes conflicts between wheelchair and bipedal pilgrims, however the simulators used were not capable of producing the aforementioned dynamics, so those pilgrims are left out of the simulation.

Where language is involved, the impact of these sorts of prescriptive studies serve to potentially misrepresent marginalized genders and people with disabilities. One approach uses a Finite-State Machine to layer intention on top of a well-known velocity space geometric collision avoidance algorithm (Curtis, Guy, Zafar, & Manocha, 2011).

This paper divides a heterogeneous population into young, old, male, and female. The 'female' agents are arbitrarily assigned lower velocities and implicitly become the cause of simulation bottlenecks or slow-downs. However, the empirical data on which the study is based makes no notion of a gender/sex difference in velocities or density. Works like this serve as a useful case for study, but arbitrary assignment of properties to sex may be used in framing real world changes that impact real peoples lives negatively. Similarly, however in a much more problematic fashion, authors have assigned an explicit negative connotation to gender (particularly women) and disability, while making far-reaching claims about outcomes. In a recent paper, pedestrians are bifurcated into two groups 'weak' and 'strong' where "disabled persons, children or women" are considered 'weak' and "young men" strong (Liu, 2018). This study uses no empirical data, only presumptuous simulations, to make the very strong claim that the "weak" should be separated from the "strong" during egress to reduce both groups evacuation times.

5.2 Overview

This study is presented in addition to the biomechanics study in the last chapter as another premise towards this dissertation's conclusions. To recap, this dissertation argues that if biomechanical modelling impacts the outcomes of synthetic crowds studies (Chapter 4) and if heterogeneity in biomechanics impacts outcomes of synthetic crowds studies (this chapter) then we, the synthetic crowd community, must critically reflect on the validity and generality of previous results and incorporate more inclusive approaches moving forward. That is, given the successful argumentation of these claims, it is then not enough to assume the efficacy of proxies for biomechanical heterogeneity in the modelling and application of synthetic crowds.

This study is broken into three experiments that separate the level at which biomechanical heterogeneity is presumed to have impact. These levels follow from an understanding of the steering layer of synthetic crowds (the focus of this dissertation). This layer can be broken into: the action space, the level at which individual decisions are made; the collision space, the resultant collider corridor formed by moving the agent's colliders through space-time; and the crowd space, an infinite space of possibilities where the interactions between an arbitrary number of agents occur with an arbitrary environment (the result of actions within the collision space). Each study in this chapter motivates the next in the sequence, each serving as premises towards the conclusion that heterogeneity at the biomechanical level has both quantitative and qualitative impacts that propagate upwards to the highest level–crowds. The first study looks at the impact of heterogeneity in biomechanics at the footstep level. The second study introduces a particular area of disordered locomotion biomechanics and qualitatively looks at the impact of heterogeneity at the collision level. The third study looks at the impact of heterogeneity in biomechanics at the crowd level. In Chapter 4, I show the importance and efficacy of biomechanical modelling in synthetic crowds. In this Chapter, I exclusively look at the impact of heterogeneity in a biomechanical model. These studies explore the question, "Does heterogeneity in locomotion biomechanics matter? If so, where?", I argue that they do matter, and that indeed this follows strongly from the discussion in Chapter 3, and that they matter at all levels of the steering layer of synthetic crowds.

5.3 Action Space

This study examines the impact of heterogeneity in locomotion biomechanics on the footstep action space. This work uses the definition of the footstep action space defined by the *footstepAI* biomechanically-based synthetic crowd simulator introduced in Chapter 4 (Berseth, Kapadia, & Faloutsos, 2015). The footstep action space is a 9 dimensional feature vector describing a single footstep in a space local to the footstep and defined by the timing and shape of the centre of mass (COM) trajectory following a simple parabolic arc. These features are: Foot Angle; Stride Length; Stride Time; Stride Speed; Step Length; Step Time; Step Speed; COM Arc Length; and COM Alpha Coefficient. The parabolic trajectory model is derived by using a small-angle approximation of the hyperbola defined by the linear inverted pendulum model (Singh, Kapadia, Reinman, & Faloutsos, 2011; Kajita, Kanehiro, Kaneko, Yokoi, & Hirukawa, 2001). The result is a local space parabola whose origin is offset from the foot flat condition[2].

[2]Using this definition, the footstepAI produces a "footstep plan", a piecewise trajectory of individual footsteps that avoid collisions with other agents while moving towards the agent's current goal. This is done by minimizing the energy expenditure over time and the energy consumed by ground reaction forces (momentum dissipation and active change in momentum) while accounting for invalid footsteps (collisions or those outside of parameter ranges).

5.3.1 Material and Methods

The data used are from the CMU Motion Capture Database (La, n.d.). The data used in this study are derived from two subjects (16 & 91) who performed a mixture of normative and stylized walking trials. The normative walking trials (subject 16) include primarily straight walking tasks where the subject turns around and returns to the starting point. The stylized walking trials (subject 91) include booth normative walking tasks such as straight, figure-8, and turn in place, and stylized walking tasks such as mummy walking, and injured walking (dragging foot) trials. These particular subjects were selected to exhibit a range of different footstep types and for their capture cleanliness (fewer errors and floating form the ground plane).

5.3.2 Analysis

In this study I use clustering as a form of confirmatory data analysis. I want to show that footsteps are separable in the footstep action space and identify why/how this is so. For this analysis, I use k-Means with the squared Euclidean distance metric in the k-Means++ algorithm (Arthur & Vassilvitskii, 2007). These choices reflect the lowest common denominator (with the improvements of careful seeding) in the area of unsupervised clustering algorithms under the assumption that if k-Means returns good results (good indicated by internal evaluation and qualitative analysis) then the space can be considered separable under even the simplest conditions. Generally, motion capture datasets are not labelled by individual footstep styles or actions, so the datasets generated by the footstep action space identification process must be considered unlabelled (though each sequence has a qualitative description, which is noted in Section 5.3.1 and later utilized for clarity in Section 5.3.3). Ideally, given the nature of the step types, a labelling for such a dataset would come from expert gait analysts who could identify steps and their "character" and then be validated by a follow-up inter-rater agreement study. This form of in-depth study for the application of unsupervised and supervised learning techniques is left to future works.

In this study, considering the unlabelled nature of individual identified footsteps and that we are not interested in classification here but rather the uniqueness or separability quality of the footsteps, I rely on internal clustering criteria and the qualitative aspects of the clustering (K. Wang, Wang, & Peng, 2009; Thalamuthu, Mukhopadhyay, Zheng, & Tseng, 2006). To achieve this, the first step is to use a form of elbow analysis to find where the error of choosing K falls off below acceptable levels. This is one common way to choose a good initial value K. In this analysis each clustering sample, $k \in 1, ..., 30$, is initialized 100 times to extract the performance, in terms of variability, of choosing that K value. I estimate and use $K = 5$ as good number to capture all variations given the motion capture sequence descriptions and the qualitative review of the footsteps present in the sequences–a number which generally agrees with the elbow analysis. This is followed with a Silhouette analysis which shows the internal clustering validity. The clusters are then reviewed qualitatively using an Andrews plot, a sort of quasi Fourier analysis used to plot multidimensional data in a way that conveys underlying structure of the data or outliers. Finally, the thumbprint, or signature, of the clusters in footstep action space is shown to impress the connection between the qualitative and quantitative cluster analysis.

This internal validation approach is followed with principal component analysis (PCA) to understand what components of the footstep action space explain the clustering and separability of the footstep action space. This examination determines the "why" of the deviations in the footstep action space by identifying the action space components that define different footsteps clusters. The hypothesis being, if stylized, disordered, and normative footstep styles differentiate in the action space the heterogeneity is important at this level.

5.3.3 Results

5.3.3.1 Normative Walking

The elbow analysis for the clustering of the normative walking subject steps can be found in Figure 5.1. The silhouette analysis of the normative walking subject steps can be seen in Figure 5.2. The Andrews plot of the clustered footstep data points can be seen in Figure 5.3. To facilitate an understanding of the "shape" of the footstep clusters, the "footprint" if you will, the cluster centroids are shown in a radar plot in Figure 5.4.

The percent of variance explained for each principal component (PC) can be seen in Table 5.1. Given that the first two PCs explain %96.8 of the variance, the clustered data transformed into the space of the first two PCs can be seen in Figure 5.5. The loadings

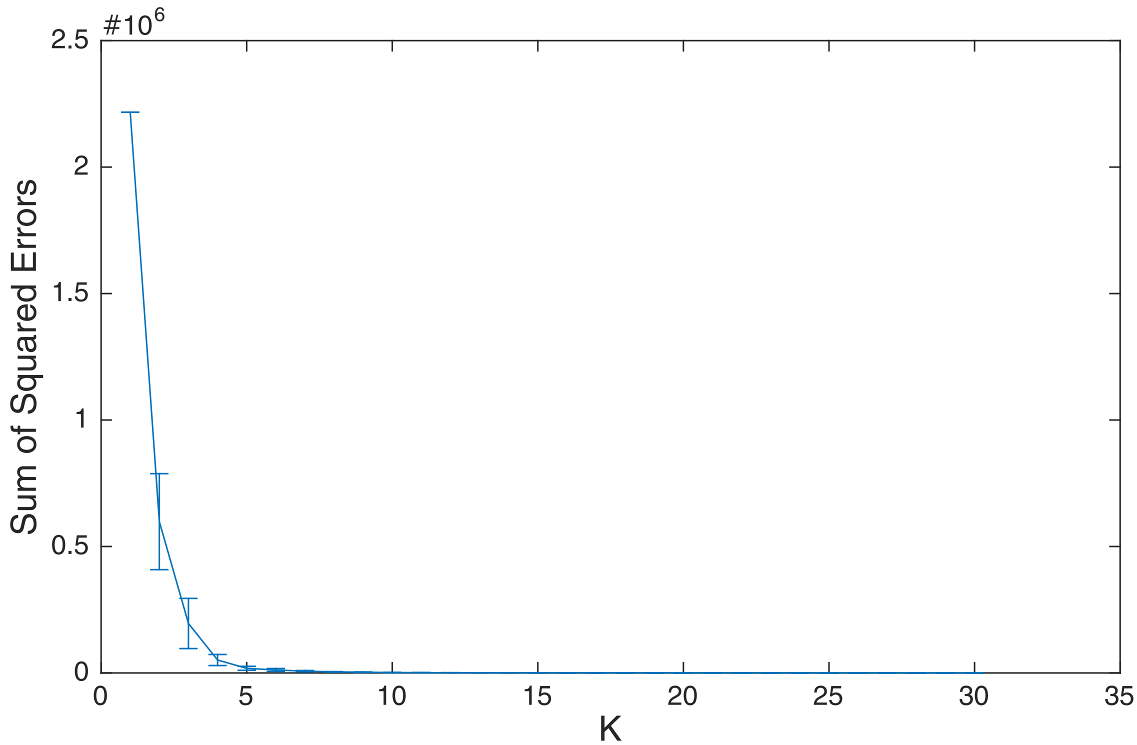


Figure 5.1. Sum of Squared Errors (SSE) elbow analysis for the normative walking subject. The K parameter of the K-Means algorithm is sampled from 1 to 30 clusters - each is initialized randomly 100 times.

for these components can be seen in Table 5.2.

5.3.3.2 Stylized Walking

The elbow analysis for the clustering of the normative walking subject steps can be found in Figure 5.6. The silhouette analysis of the normative walking subject steps can be seen in Figure 5.7. The Andrews plot of the clustered footstep data points can be seen in Figure 5.8. To facilitate an understanding of the "shape" of the footstep clusters, the "footprint" if you will, the cluster centroids are shown in a radar plot in Figure 5.9.

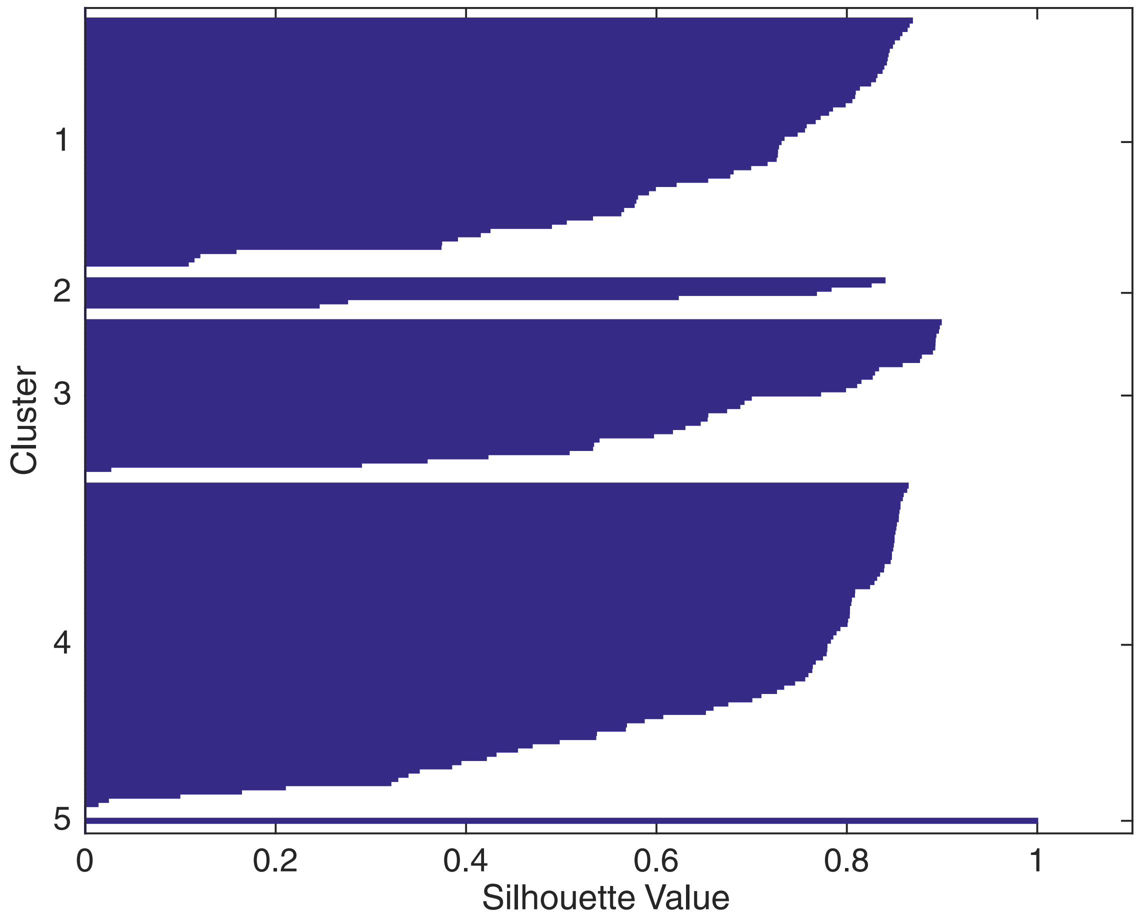


Figure 5.2. Silhouette values for the normative walking subject's footstep clusters. The singular outlier is clearly identified as being perfectly clustered on its own. Other clusters show good clustering values predominantly above 0.6 and 0.8. Cluster 4 appears to be a wide reaching cluster with approximately a quarter of footsteps not being well clustered.

Table 5.1: Percent of variance explained for each principal component of the normative walking subject steps.

PC	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
%	91.6145	5.1863	2.3622	0.3566	0.2246	0.1782	0.0591	0.0113	0.0072

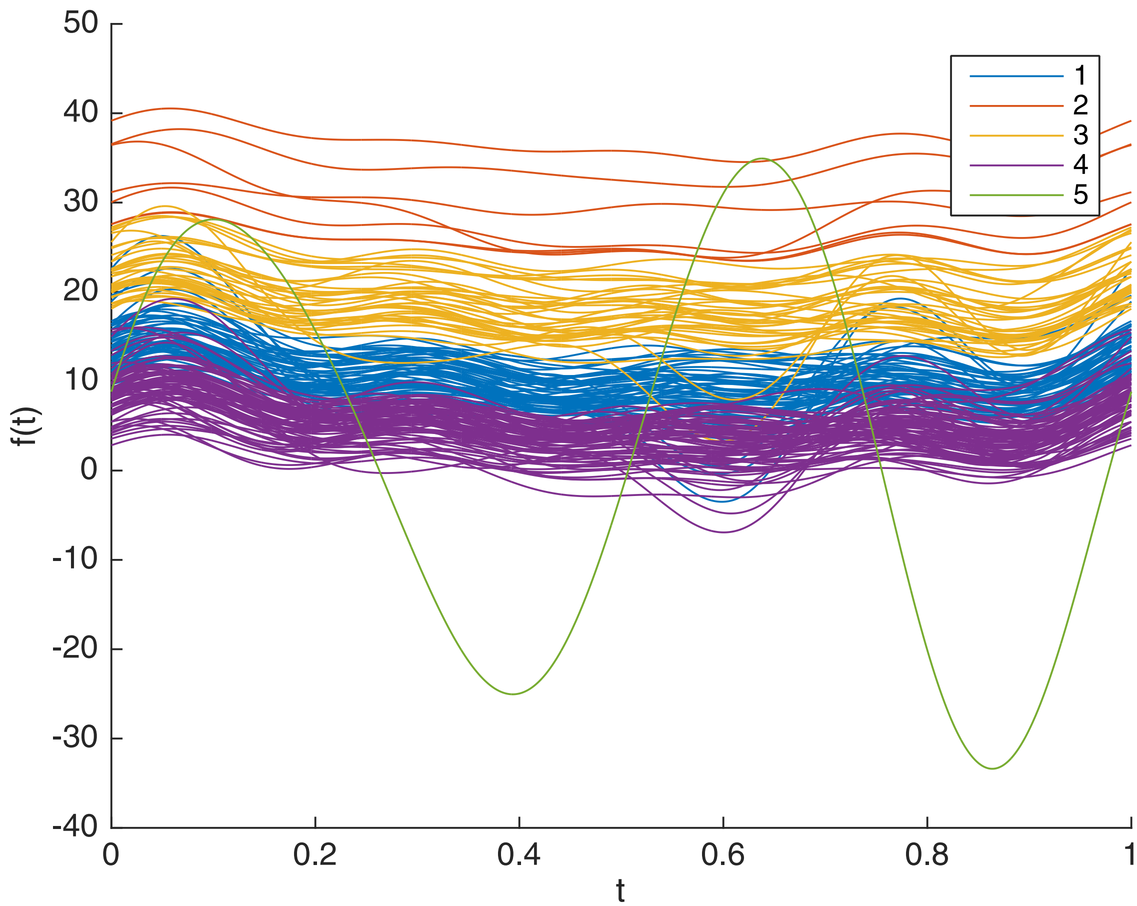


Figure 5.3. Andrews Plot of the normative walking subject's footstep clusters. Even in differently stylized normative walking footsteps there is not much separation in the Andrew's curves and the structure is regular and oscillatory.

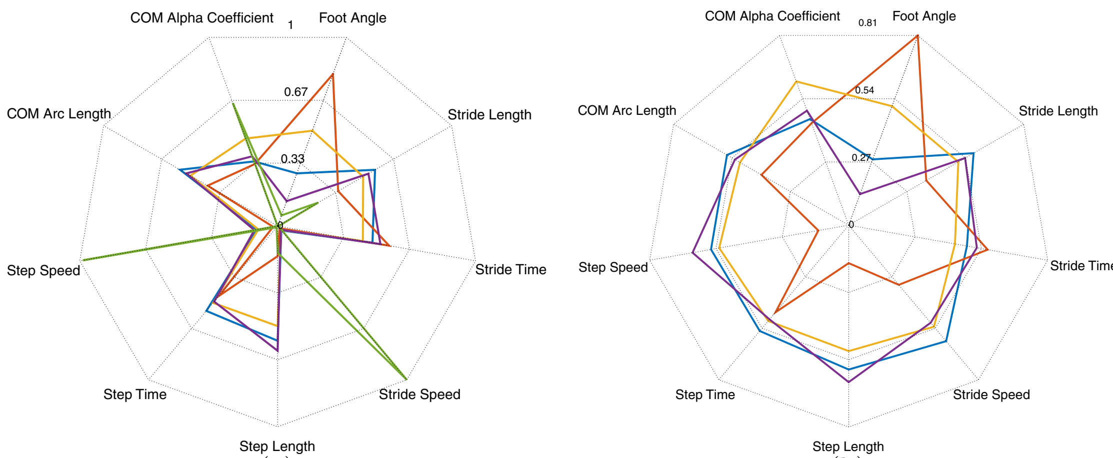


Figure 5.4. Normative Cluster Radar Plot. Because the outlier causes severe scaling, both the (a) original cluster centroids and the (b) centroids with the outlier removed are plotted for clarity and comparison. In (b) an outlier shape can be seen in high dimensions, upon qualitative review this cluster appears to be associated with the turning steps.

Table 5.2: Coefficients, or loadings, of the first two principals components of the normative walking subject's footsteps. The primary, or most significant loadings, are in boldface.

Dimension	**Feature**	**PC1**	**PC2**
1	Foot Angle	**0.9996**	0.0231
2	Stride Length	-0.0041	0.0051
3	Stride Time	-0.0041	-0.2619
4	Stride Speed	-0.0234	**0.9331**
5	Step Length	-0.0053	-0.0013
6	Step Time	-0.0006	-0.1459
7	Step Speed	-0.0123	0.1931
8	COM Arc Length	-0.0017	-0.0250
9	COM Alpha Coefficient	0.0014	-0.0275

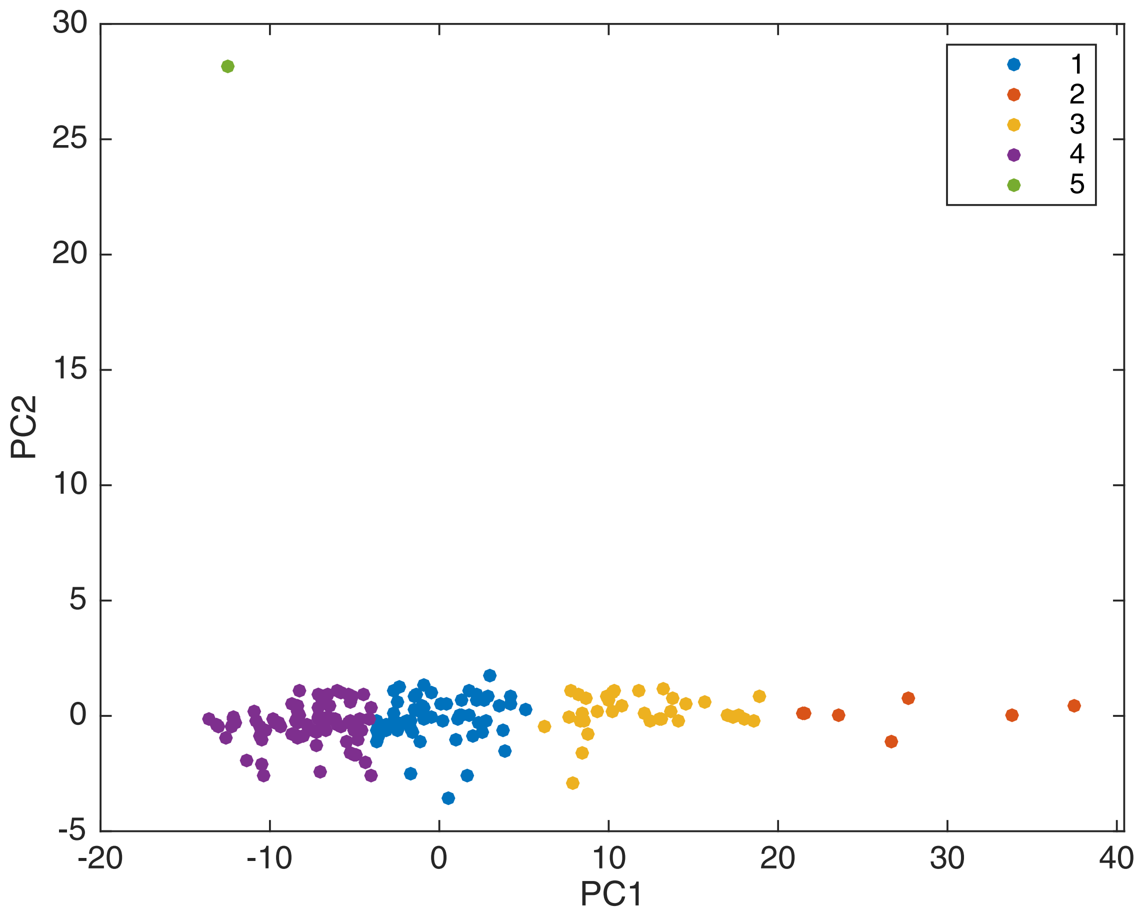


Figure 5.5. Normative PCA, PC-1 and PC-2.

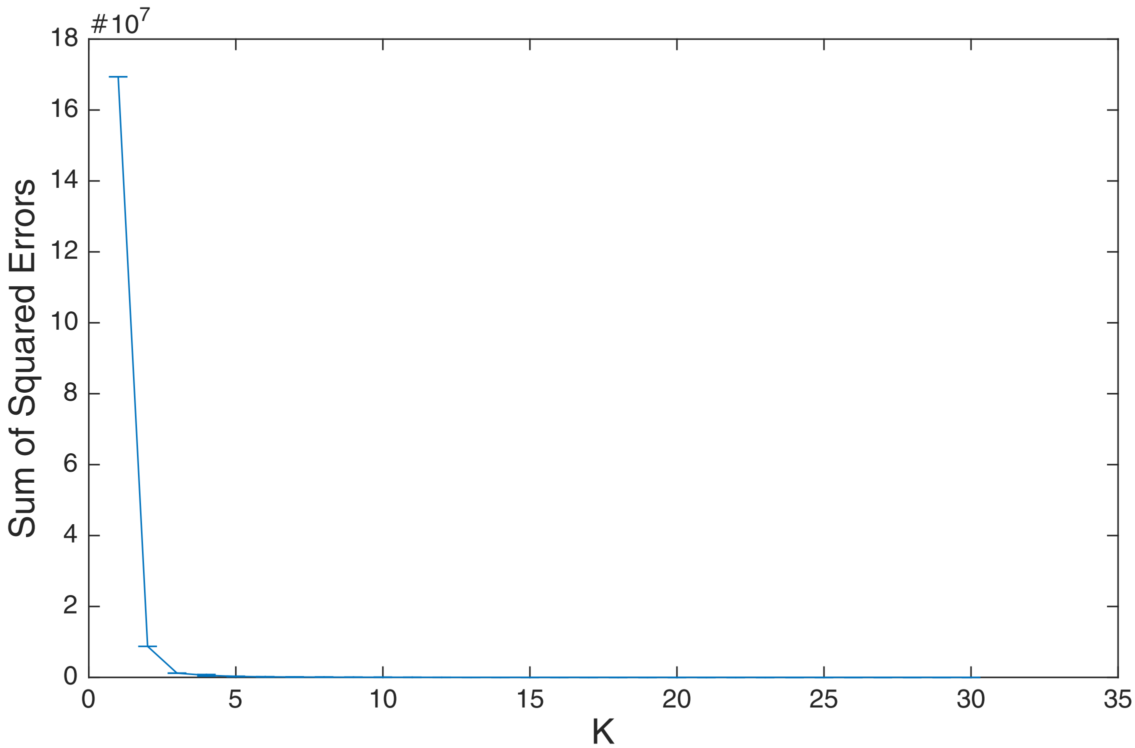


Figure 5.6. Sum of Squared Errors (SSE) elbow analysis for the stylized walking subject. The K parameter of the K-Means algorithm is sampled from 1 to 30 clusters - each is initialized randomly 100 times.

The percent of variance explained for each principal component (PC) can be seen in Table 5.3. Given that the first two PCs explain %98.9 of the variance, the clustered data transformed into the space of the first two PCs can be seen in Figure 5.10. The coefficients, or loadings, for these PCs can be seen in Table 5.4.

5.3.4 Discussion

The results of this experiment verify the general separability of footstep styles. In the following subsections, I examine each subject's results individually.

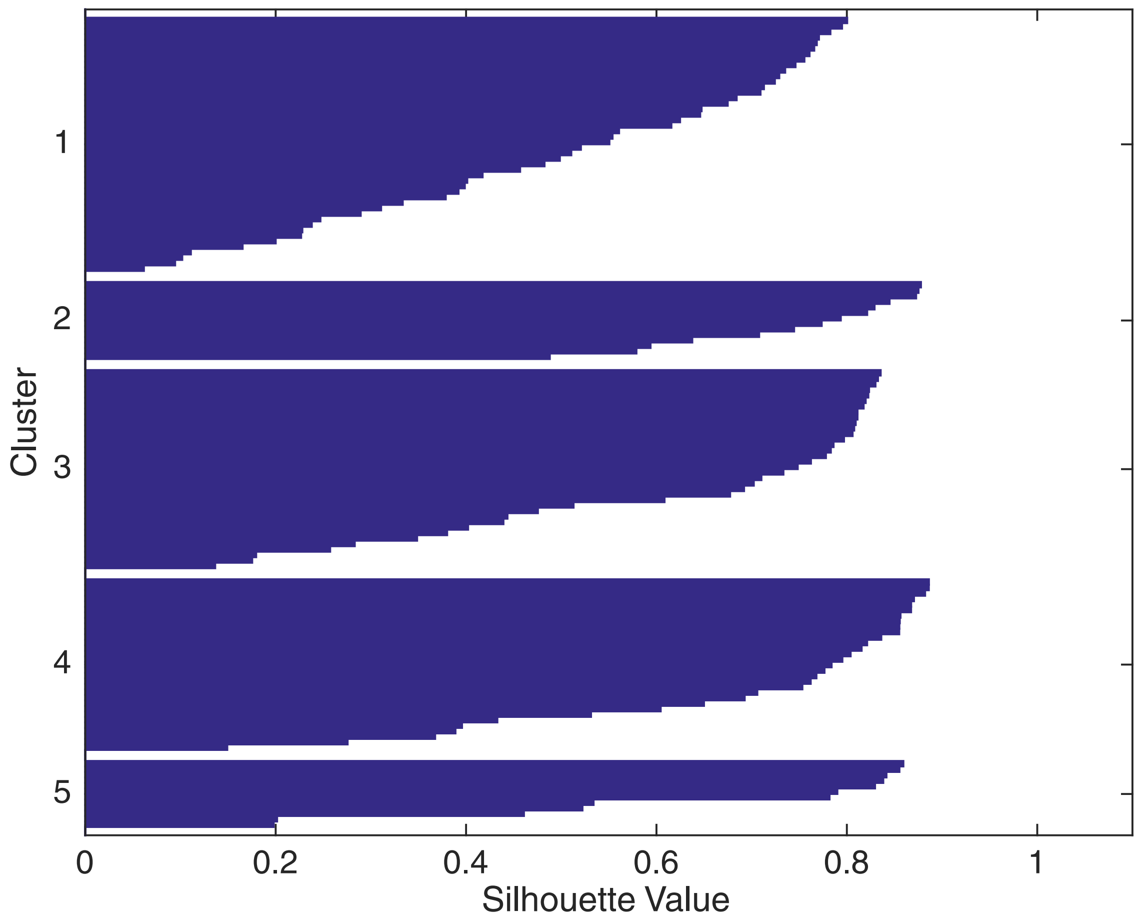


Figure 5.7. Silhouette values for the stylized walking subject's footstep clusters. The clusters show good clustering values predominantly above 0.6 and 0.8. Cluster 1 appears to be a wide reaching cluster with approximately a quarter of footsteps not being well clustered.

Table 5.3: Percent of variance explained for each principal component of the stylized walking subject steps.

PC	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
%	97.2402	1.6184	0.4826	0.3582	0.2040	0.0589	0.0208	0.0111	0.0058

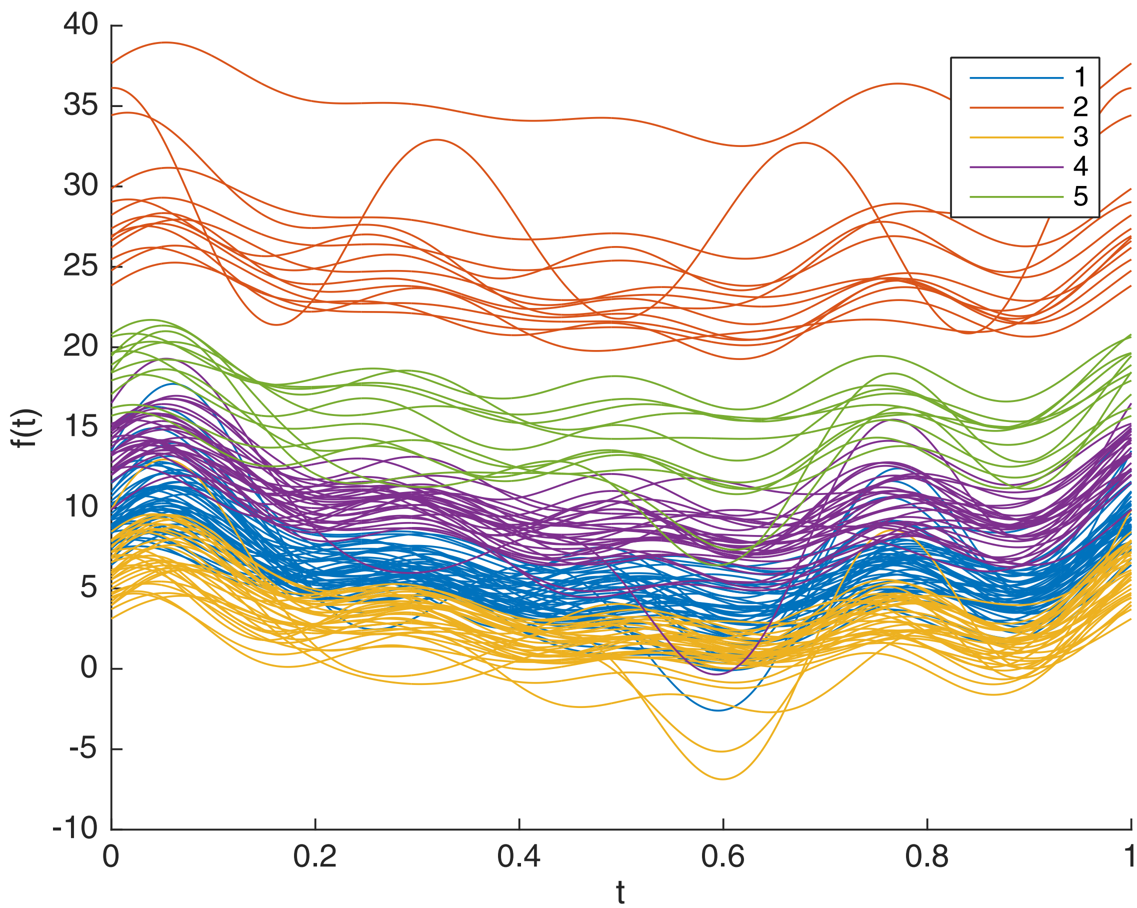


Figure 5.8. Andrews plot of the stylized walking subject's footstep clusters.

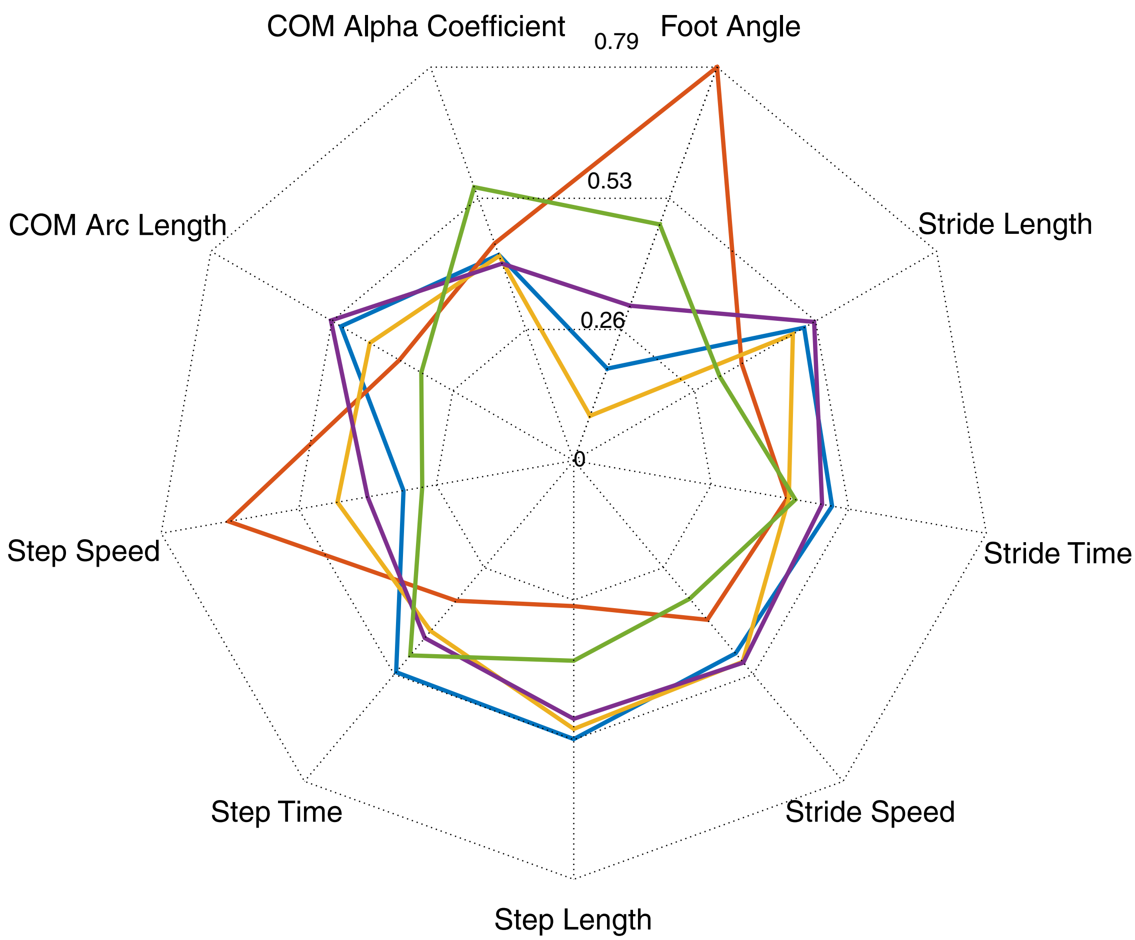


Figure 5.9. Clustered radar plot of the stylized walking subjects. An outlier shape can be seen in high dimensions, upon qualitative review this cluster appears to be associated with the limping steps.

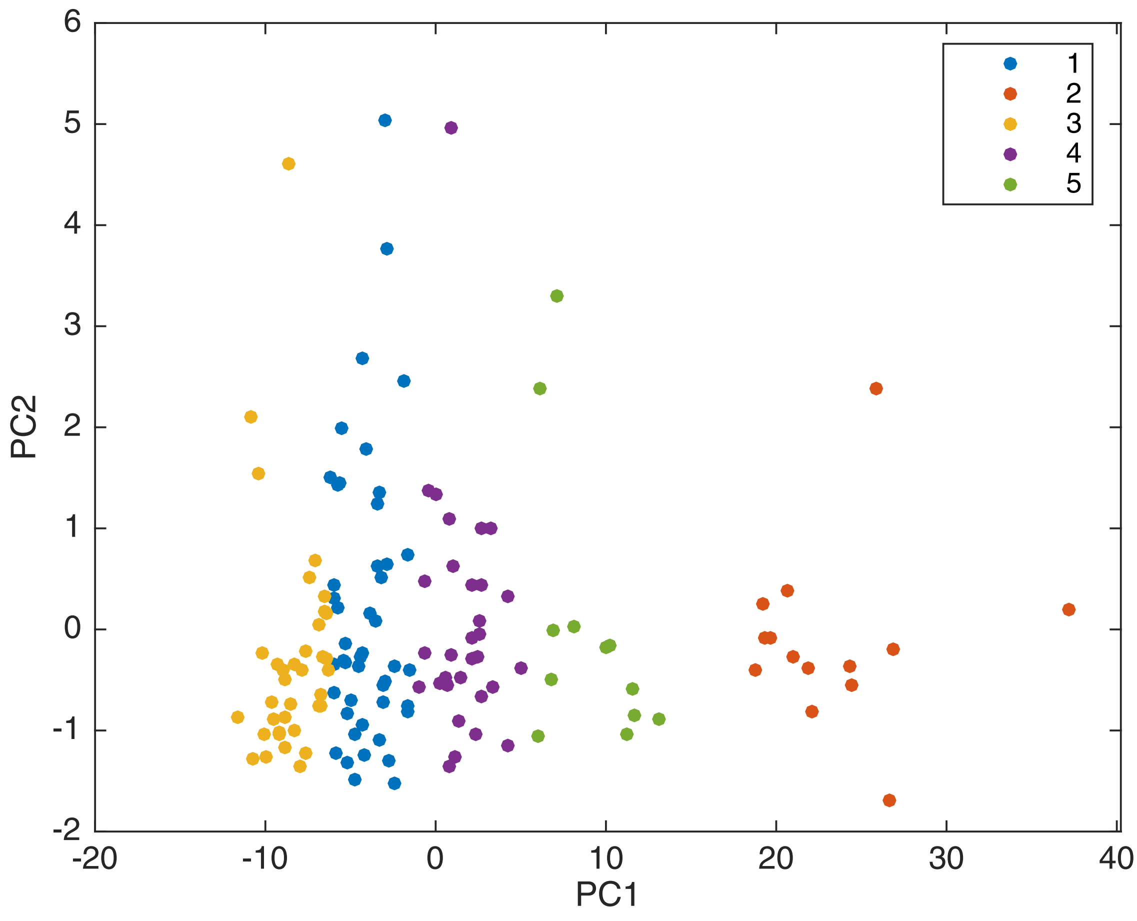


Figure 5.10. Stylized PCA, PC-1 and PC-2.

Table 5.4: Coefficients, or loadings, of the first two principals components of the stylized walking subject's footsteps. The primary, or most significant loadings, are in boldface. Interestingly, the most significant PC2 loadings are temporal features.

Dimension	**Feature**	**PC1**	**PC2**
1	Foot Angle	**0.9998**	0.0087
2	Stride Length	-0.0086	-0.0102
3	Stride Time	-0.0053	**0.8034**
4	Stride Speed	-0.0042	-0.1276
5	Step Length	-0.0053	0.0486
6	Step Time	-0.0069	**0.5627**
7	Step Speed	0.0071	-0.1107
8	COM Arc Length	-0.0052	0.0616
9	COM Alpha Coefficient	0.0038	0.0540

5.3.4.1 Normative

The results of the normative motion capture subject show an expected lack of variability in footstep action space components. There are two important outcomes of this particular experiments. The first is that the underlying system (described in Chapter 6) for reverse engineering the footstep action space from animation used to collect this data is not perfect and occasionally produces erroneously detected footsteps. This is rendered as the green coloured cluster throughout the plots which highlight the clustered data. This singular data point is the result of a false positive foot flat condition which makes the step appear to be extremely fast, flat, and short.

The second interesting and important outcome of this study is the component of footstep action space which contributes to most of the variability to normative footsteps based on the loadings in Table 5.2–the foot angle. This is entirely explained by the underlying data which comes from a subject walking in straight lines, turning around and walking back. Thus, the foot angle differentiates the turning steps from the straight walking steps. The subject makes different types of multipoint turns and it is hypothesized that this is primarily what differentiates the clusters w.r.t foot angle. Additionally, noting the lack of variability or dispersion of data points along PC2, the normative steps are strictly clustered along PC1 and do not differ much from each other in terms of PC2. Examining the loadings of PC2 reveals that the primary contributing feature of the footstep action space for PC2 is the stride speed. This can be almost entirely explained by the outlier which differentiates along PC2.

5.3.4.2 Stylized

The results of the stylized motion capture subject show an expected addition of variability in footstep action space components. There are two important outcomes of this particular experiments. The first is that the subject 91 dataset includes normative walking and these, upon an informal review, appear to have separated from the stylized walking conditions.

The second interesting and important outcome of this study is the component of footstep action space which contributes to most of the variability to stylized footsteps based on the loadings in Table 5.4. While the most significant loading in PC1 is still foot angle, the most significant loadings in PC2 are Stride Time and Step Time. This is entirely explained by the underlying data which comes from a subject walking in straight lines, turning around and walking back under stylized conditions that include temporal gait asymmetry (TGA). Thus, the foot angle differentiates the turning steps from the straight walking steps, but the stride time and step time differentiate asymmetrical stepping patterns which are a result of TGA. It is clear that these two loading have a scaling impact on PC2 unlike in the normative walking study which shows very little dispersion in PC2.

5.3.4.3 Conclusion

In conclusion, this study shows that the footstep action space is: separable over the space of footsteps (shown over a small sample of possible footsteps) and potentially a good representation for temporal gait asymmetries that produce footstep level asymmetries. Further work in this are is required in order to examine the footstep action space parameters which correspond to the vast body of conditions and disorders which cause temporal gait asymmetries. Additional further work is required to understand not only the difference between within subject footsteps in the footstep action space but also between subjects.

The radar plots of both stylized and normative subjects in the footstep action space appear to have some immediate use. By viewing the overlaid footstep patterns over the action space dimensions it is hypothesized that disordered steps and apparatus errors are much easier to identify. In the normative walking subject data, the outlier step is immediately apparent because it dominates the component scaling which appears as spikes. After removal, the radar plot clearly separates turning steps from normative ones and in place turning from curving path steps. In the stylized footstep subject data, the stylized disordered footsteps (dragging, limping) are immediately apparent because of the increased foot angle and step speed, while the stylized zombie footsteps have a short stride length and exaggerated COM Alpha Coefficient (associated with COM swaying). Further work is required to understand the efficacy of this form of visualization in for gait analysis as well as for automatic crowd analysis of ability and heterogeneity.

5.4 Collision Space

This study examines the impact of heterogeneity in locomotion biomechanics on the collision space of a synthetic crowd agent. All synthetic crowd models must facilitate and attempt to avoid collision as a fundamental aspect of the steering layer of crowd modelling. In particle-based crowd simulation, agents are represented as particles which define their collision bounds. This representation greatly simplifies the mathematics required for detecting and resolving collisions. The collision space in this experiment then is the collision shape, the combined spatial sum of the collision particles, extended through time. This can thought of as a corridor of avoidance. The hypothesis of this study is that heterogeneity in biomechanics impacts the shape of this collision space. Since the collision space is fundamental to the resolution of collision avoidance, then an affected collision space would impact combinatorial interactions between agents (the next study, Section 5.5, addresses the evidence of this implication).

5.4.1 Material and Methods

This study focuses on four different model types that span a small spectrum of mobility heterogeneities. In particular, this study focuses on spatio-temporal gait asymmetries, referred to as Temporal Gait Asymmetry (TGA) in this chapter, as a probe for exploring biomechanical heterogeneity at the steering level. TGA is a common symptom of several conditions and medical disorders from injuries to neurodegenerative diseases and is often described characteristically as 'limping' though it may take many forms and appear at different intervals. Specifically, TGA is the occurrence of asymmetries in temporal gait parameters, such as step timing (more details in Section 5.4.1.2). In this study, I use the *footstepAI* steering model with various forms of higher level parametrization control to induce different TGA conditions in the footstep planning. Much of the information regarding TGA in this section is extracted from Dr. Kara Kathleen Patterson's PhD Dissertation on measures of gait asymmetry post-stroke, and is recommended reading for those interested in these conditions (Patterson, 2010).

5.4.1.1 Normative Locomotion Biomechanics

This condition uses the *footstepAI* model as is with its Principle of Least Effort (PLE) parametrization of the robust *footstepAI* steering model (Berseth, Kapadia, & Faloutsos, 2015; Berseth et al., 2014). That is, this model matches the parametrization of the *footstepAI* model in Chapter 4. The full set of parameters and their default values can be seen in Table 5.5. Note that, *Step Duration Scale* and *Desired Velocity Scale* are

Table 5.5: Least effort parametrization of the footstepsAI model for step affecting parameters.

footstepAI **Footstep Parameters**	**Values**
Preferred Step Angle	0.58 rad
COM Height	0.93m
Min. Step Length	0.16m
Max. Step Length	0.79m
Min. Step Time	0.03m
Max. Step Time	0.6m
Max. Speed	1.33m/s
Time Cost Weight	0.58
Trajectory Cost Weight	0.03
Shoulder Comfort Zone	0.39m
Shoulder Comfort Zone 2	0.03m
Step Duration Scale	1
Desired Velocity Scale	1

additional parameters used to support the following models (thus their default values for the normative model are 1). For the remainder of this chapter, this model will be referred to as "normative".

5.4.1.2 Temporal Gait Asymmetry

This conditions the robust *footstepAI* model with a higher level model that produces the TGA symptoms in the steering model by modulating the parameters during footstep planning. This higher level parametrization is required to bifurcate the foot step parameters (necessary for creating asymmetry). To focus this study, and derive empirical parameters from real data, I use data and analysis of TGA in patients post-stroke. The prevalence of Stroke is well studied, and is considered a significant health issue (Hodgson, 1998; Hakim, Silver, & Hodgson, 1998). Furthermore, spatiotemporal gait asymmetries are prevalent in community-ambulating (independent walking ability within $200m$ (Brown et al., 2010)) chronic stroke survivors. For example, of 54 participants post-stroke, 55%

and 33% exhibited significant temporal and spatial asymmetries respectively (Patterson et al., 2008). Gait asymmetries in patients post-stroke have impacts on walking speed, falls, and energy costs, all of which, in turn, impact patients' quality of life, health, and autonomy (Lauziere, Betschart, Aissaoui, & Nadeau, 2014).

This higher level model modulates the step time duration for each footstep planned. When the model is initialized the paretic side is chosen randomly with no preference. Then for each potentially planned step, when the planning foot matches the paretic side, the model loads and modulates the step duration scale. On the paretic side, the step duration scale is 1.42385, which is the average of the swing time asymmetry range reported as a ratio of paretic/non-paretic side in the literature (Dettmann, Linder, & Sepic, 1987; Brandstater, Gowland, Clark, et al., 1983). Note that it is unclear what effect post-stroke swing time asymmetry has on step length, with patients having either longer paretic or non-paretic steps (Kim & Eng, 2003; Hsu, Tang, & Jan, 2003; Dettmann et al., 1987; Balasubramanian, Bowden, Neptune, & Kautz, 2007). The impact of the step duration scaling can be seen spatially in the results, Section 5.4.3, of this study, Figure 5.11. This is the result of the internal optimization resolving the footstep parameter constraint given the step duration scaling, and may reflect certain real world cases. The accuracy of this modelling for patients post-stroke is outside the scope of this dissertation and is earmarked as important future work. For the remainder of this chapter, this model will be referred to as "TGA ".

5.4.1.3 Temporal Gait Asymmetry Velocity Reduction

The relationship between velocity and asymmetry in gait is complicated. It is not clear whether the two represent the same changes in post-stroke walking ability. For example, asymmetry will likely worsen over time, velocity may not (Patterson, Gage, Brooks, Black, & McIlroy, 2010a). Post-stroke gait velocity is characterized by a non-linear reduction with age (El Haber, Erbas, Hill, & Wark, 2008). However, there is extensive variation in the measures of preferred velocity across the literature. See Section 2.2.1 of Patterson's Dissertation for an excellent review of this literature (Patterson, 2010).

Much like the aforementioned TGA model, this model induces asymmetry by bifurcating and modulating the foot planning parameters. In this model, step duration scale is modulated for the paretic side, while the desired velocity scale, max. speed, min. step length, max. step length are altered generally. This general change is required to help the internal optimization algorithm find footsteps which fall within these constraints (otherwise steps would fall outside the min. and max. values, forcing the optimization into an endless search). The step duration scale is set at 1.42385 as in the aforementioned TGA model. The desired velocity scale is meant to scale preferred velocities down to those found in the literature. That is, the highest reported preferred velocity of chronic patients post-stroke is 0.76m/s (Nadeau, Arsenault, Gravel, & Bourbonnais, 1999) and the normative average preferred velocity is 1.33m/s, so the desired velocity scale is 0.571428. The max. speed is 1.09m/s, the highest reported speed of chronic patients post-stroke (Nadeau et al., 1999). The min. and max. step lengths of 0.08m and 1.027m were found empirically under a battery of tests with the planning model. Adjustments were made to the step length range until the planner could successfully create plans under all local conditions. For the remainder of this chapter, this model will be referred to as "TGA_velocity".

5.4.1.4 Velocity Reduction

As explored in Section 5.1, the literature on applied synthetic crowd models seeking to cover diverse mobilities has used velocity as a proxy for people with gait disorders and people using assistive devices. In the review of literature, velocity was used in several instances where the modelled agent would have had Temporal Gait Symmetries. In this model, I match the velocity scale, max. speed, min. step length, max. step length with the aforementioned TGA_velocity model in Section 5.4.1.3. However, this model has no bifurcation of parameters in the step planning phase. In this way, this model serves as a good control for the TGA_velocity model. For the remainder of this chapter, this model will be referred to as "Velocity" (in the context of models not to be confused with the concept of velocity itself).

5.4.2 Analysis

This study is primarily a qualitative one to show that collider corridors of the various models differ from each other in shape. Recording taken over 7 normative steps (3.5 strides), or approximately 2.66m, are compared to each other. This distance was chosen as all models approximately reach a *beat* at the same distance, a typically physical phenomenon where two similar frequencies interfere periodically. Here the term is applied loosely since the foot side or amplitude is not taken into account, simply that feet spatially hit their foot flat, or loading response phase, conditions in approximately the same place. Knowing the approximate distance travelled also affords the quantitative measure of spatial frequency (here, cycles per m, or $strides/distance$). All measures are approximated from the collider corridors. Previous methods in animal gait characterization and motion path editing have used the Froude number to characterize the relationship between spatial and temporal attributes (Alexander, 1976; Lockwood & Singh, 2011). The Froude number is useful in characterizing walking patterns between animals of varying size. In these studies, I control for body size and mass as a form of heterogeneity to focus on changes at the action level in the biomechanical model[3]. Thus, the Froude number reduces to step frequency. As well, this frequency is separated between left and right side to note asymmetry.

5.4.3 Results

The corridors of each model in the simple straight walking task can be seen in Figure 5.11. The normative model produces a regular symmetric pattern where the mean stride length is 0.762m (step lengths are then approximately 0.381m). The spatial stride frequency is approximately 1.32Hz. The velocity model produces a regular symmetric pattern of stride lengths approximately 0.508m (step lengths of approximately 0.254m). The spatial stride

[3]These particular parameters are purposefully homogeneous in this study, but are earmarked as important forms of heterogeneity to be explored with respect to biomechanics and crowds. Past works have used the particle radius parameter to represent different body sizes.

frequency is approximately 2.068Hz. The TGA model produces asymmetrical stepping pattern with shorter paretic side step. The mean length of paretic side steps is 0.431m, while the for non-paretic side the mean length is 0.504m. The spatial stride frequency is approximately 0.9398Hz. The TGA_velocity model produces asymmetrical stepping pattern with shorter paretic side step. The mean length of paretic side steps is 0.28m, while the for non-paretic side the mean length is 0.299m. The spatial stride frequency is approximately 1.6917Hz.

5.4.4 Discussion

The qualitative difference between the model's collider corridors is highlighted in Figure 5.11 by demarcating the approximate left and right foot flat, or loading response phase, conditions. None of the corridors is like the other. The implication here is that if an agent's steering behaviour is predicated on the collision space of nearby agents, then changes in collision space would change the agent's steering behaviours. It is clear in this study that the step sizes, spatial frequency, and ultimately the shape of the collider corridors is impacted by the control of temporal parameters related to TGA. While the velocity model may capture symmetric shuffling behaviours, several features of the collider corridors are not. This is in addition to the findings of Chapter 4, which show that these sorts of non-linear movements are not even possible without either locomotion biomechanics or some form of non-linear kinematics modelling.

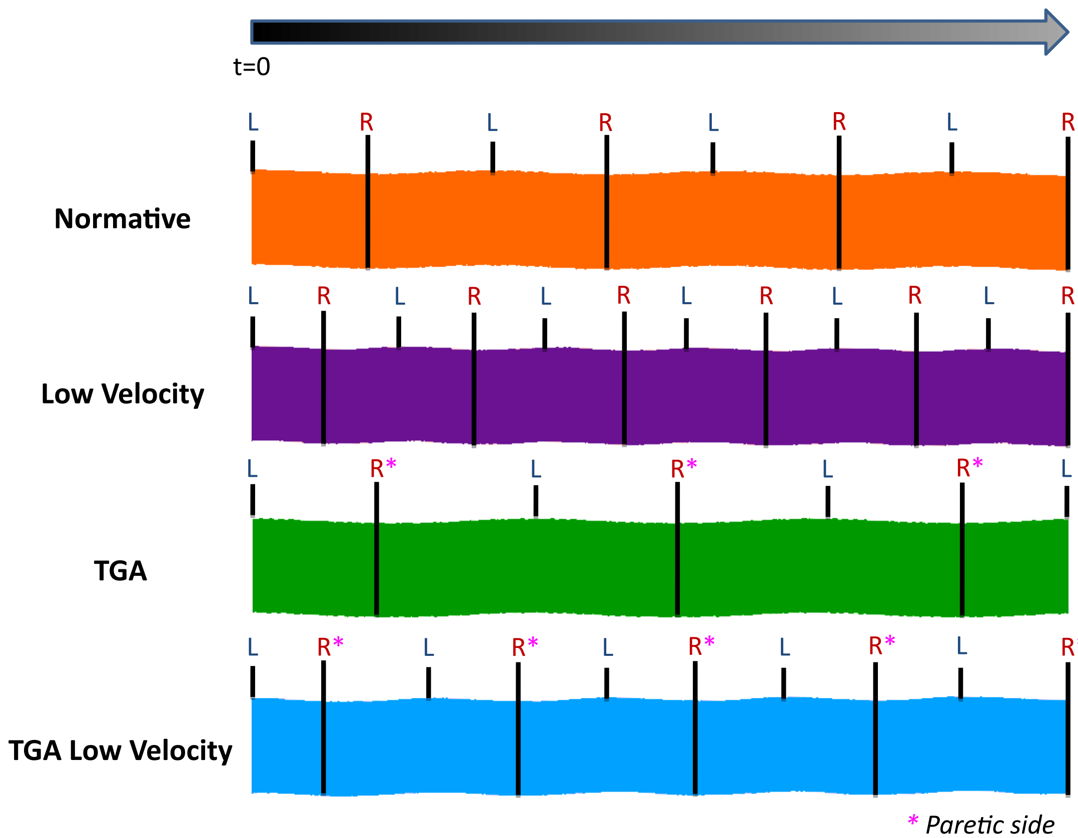


Figure 5.11. Agent level collider corridors. Each agent's collider space is defined by its particle(s). The collider corridors shown here are the collider space over time in a simple straight walking task from left to right. The loading response, or foot flat, phase conditions are annotated in each model for each foot. In the disordered conditions the paretic side is annotated with an asterisk.

5.5 Crowd Space

This study examines the impact of heterogeneity in locomotion biomechanics on common crowd outcomes measures. The hypothesis of this study, given the results of Sections 5.3 & 5.4, is that heterogeneity in locomotion biomechanics (in this case, the non-linearity of TGA) within a crowd changes common crowd outcomes measures. The focus is on outcomes which are used in applied situations (policy decisions, design decisions, and evaluation frameworks), and in particular flow rate which, while it has no singular definition, is common in the analysis of safety critical scenarios.

5.5.1 Material and Methods

The definitions of the outcome measures used in this study match the descriptions in Chapter 4 Section 4.2.2. The study examines these crowd outcome measures in the context of a common scenario in the literature, and sub scenario of applied synthetic crowds - the bottleneck egress (see Figure 4.5). There is a rich history of this type of scenario's use covered in Chapter 4 Section 4.5 and Chapter 2 Section 2.2. The models used in this study are those described in Section 5.4.1.

In the previous study, all sources of heterogeneity were carefully removed and the models normalized before comparison. In this chapter, heterogeneity is carefully introduced in a purposefully limited way to understand the impact without confounding factors. To do this, the study samples multiple levels of heterogeneity, namely: 1/50 (2%), 5/50 (10%), 10/50 (20%), 20/50 (40%), 30/50 (60%), 40/50 (80%), 50/50 (100%); where the ratio is *agents with the specified model/total agents* and the remaining agents from the total are using the normative effort optimal parametrization. In each crowd mixture, there are a number of controls. First, each mixture includes a comparison to a normative baseline where all agents in the crowd are homogeneous and have the default effort optimal parameters. Conversely, the TGA condition is a control for all other conditions including the normative, so that the impact of TGA can be analysed in crowds generally. Finally, and perhaps most importantly for this dissertation, the velocity condition is a control for the TGA and TGA_velocity conditions. This last control is important because it will tell us if the use of velocity, as is done across the history of synthetic crowds, is an adequate proxy for heterogeneity in synthetic crowds. This way of controlling the study will tell us at which point, if any, the mixture of agents produces a significant impact on outcomes.

It should be noted however that within each group, that is *agents with the specified model* and *remaining agents* the within group agents are homogeneous. The only caveat to this is the paretic side selection in the TGA conditions. There is no preference for side across conditions, so in these groups the paretic side is randomized and considered approximately symmetric over the distribution of scenarios. That is, the average left and right paretic sides across mixtures and simulations is equal. The importance of this point is mainly that, I hypothesize, that in these very controlled conditions, if significant outcome differences appear, then within group heterogeneity would only increase these differences in outcomes. However, this study is outside the scope of this dissertation and requires an in-depth sampling of condition dependent distributions of symptoms.

5.5.2 Analysis

In this study, I extend the concept of fundamental diagram to compare the changes in outcome measures over levels of heterogeneity mixtures. For each heterogeneity mixture, agents are randomly initialized and simulated 100 times. For each heterogeneity mixture level and each outcome measure, a Kruskal Wallis test is performed to determine if there is a significant difference amongst the ranks of each of the models. Repeated pairwise post-hoc Conover's tests (with false discovery rate (FDR) corrections in the presence of ties) are used to identify the significant differences among the models $p < 0.05$.

The qualitative analysis includes heatmaps of aggregate movement, or flux, over the best performing (in terms of flow rate) scenarios for each of 30/50 (60%), 40/50 (80%), 50/50 (100%) mixtures. The scenario is divided into sections to focus the qualitative analysis.

5.5.3 Results

The fundamental diagram for the total crowd effort metric can be seen in Figure 5.12 and the statistical test results can be found in Table 5.6. The fundamental diagram for the crowd flow rate can be seen in Figure 5.13 and the statistical test results can be found in Table 5.7. The fundamental diagram for the total crowd path length can be seen in Figure 5.14 and the statistical test results can be found in Table 5.8. The aggregate flux maps for the best performing (by flow rate) scenarios for each model is shown in Figure 5.15.

Table 5.6: Matrix of statistically significant test results for total crowd effort. All check marks are significant as per Section 5.5.2. The N value is the intersection of completed scenarios between all models.

	1	**5**	**10**	**20**	**30**	**40**	**50**
normative/TGA	x	x	✓	✓	✓	✓	✓
normative/TGA_Velocity	✓	✓	✓	✓	✓	✓	✓
normative/velocity	✓	✓	✓	✓	✓	✓	✓
TGA/TGA_velocity	✓	✓	✓	✓	✓	✓	✓
TGA/velocity	✓	✓	✓	✓	✓	✓	✓
TGA_velocity/velocity	x	x	x	✓	✓	✓	✓
N	*96*	*97*	*95*	*87*	*83*	*84*	*77*

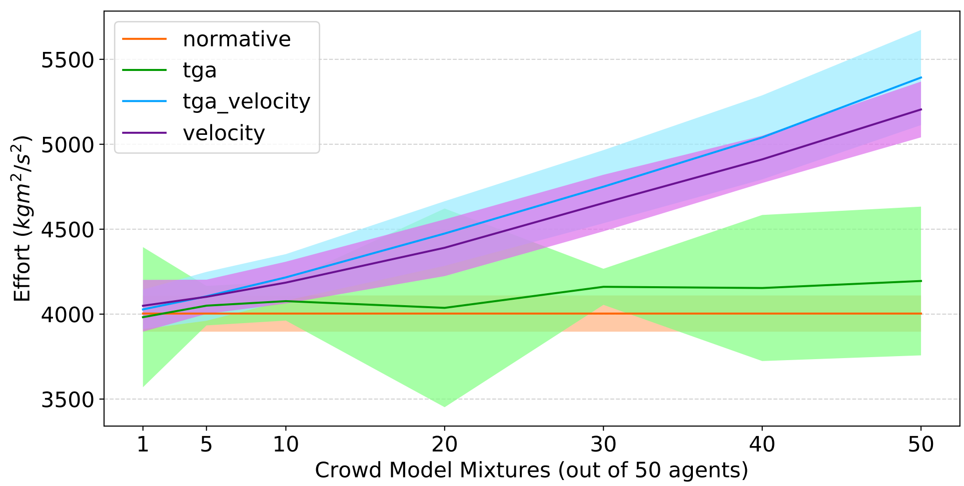


Figure 5.12. Fundamental diagram of total crowd effort over model mixtures. The lines and shaded areas represent the means (μ) and standard deviations (σ) respectively. Detailed results can be found in Appendix B.

Table 5.7: Matrix of statistically significant test results for the crowd flow rate. All check marks are significant as per Section 5.5.2. The N value is the intersection of completed scenarios between all models.

	1	**5**	**10**	**20**	**30**	**40**	**50**
normative/TGA	x	✓	✓	✓	✓	✓	✓
normative/TGA_Velocity	x	✓	✓	✓	✓	✓	✓
normative/velocity	✓	✓	✓	✓	✓	✓	✓
TGA/TGA_velocity	✓	✓	✓	✓	✓	✓	✓
TGA/velocity	✓	✓	✓	✓	✓	✓	✓
TGA_velocity/velocity	x	x	x	✓	✓	✓	✓
N	*96*	*97*	*95*	*87*	*83*	*84*	*77*

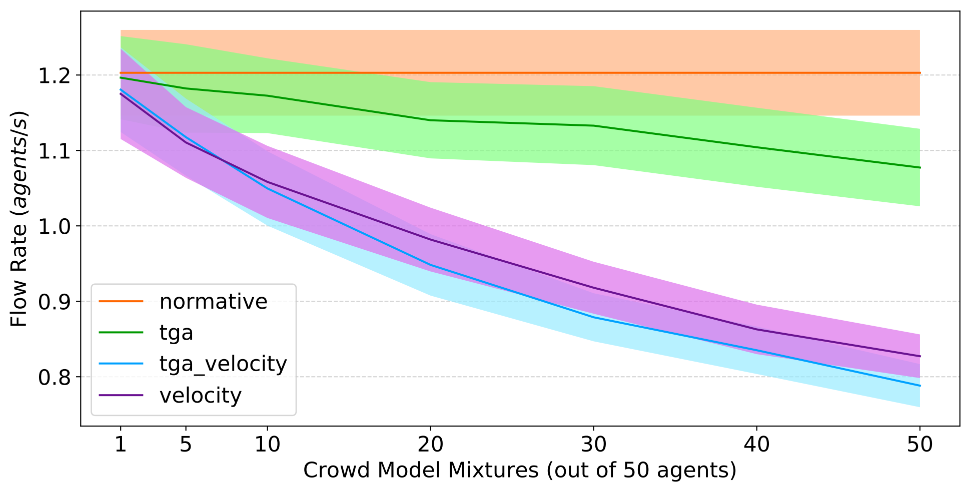


Figure 5.13. Fundamental diagram of total crowd flow rate over model mixtures. The lines and shaded areas represent the means (μ) and standard deviations (σ) respectively. Detailed results can be found in Appendix B.

Table 5.8: Matrix of statistically significant test results for the crowd path length. All check marks are significant as per Section 5.5.2 (no difference amongst models at the 1/50 (2%) level). The N value is the intersection of completed scenarios between all models.

	1	**5**	**10**	**20**	**30**	**40**	**50**
normative/TGA	-	x	✓	✓	✓	✓	✓
normative/TGA_Velocity	-	x	x	✓	✓	✓	✓
normative/velocity	-	x	✓	✓	✓	✓	✓
TGA/TGA_velocity	-	✓	✓	✓	✓	✓	✓
TGA/velocity	-	✓	✓	✓	✓	✓	✓
TGA_velocity/velocity	-	x	x	x	x	x	✓
N	*96*	*97*	*95*	*87*	*83*	*84*	*77*

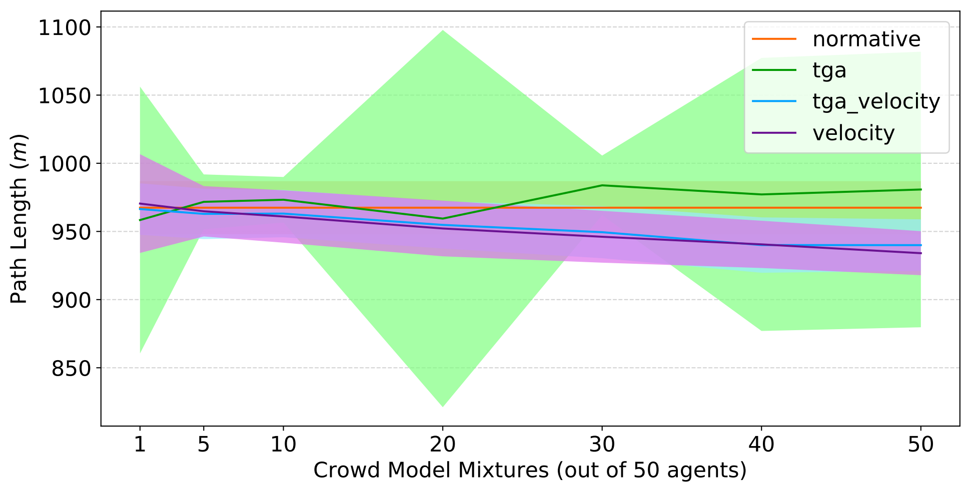
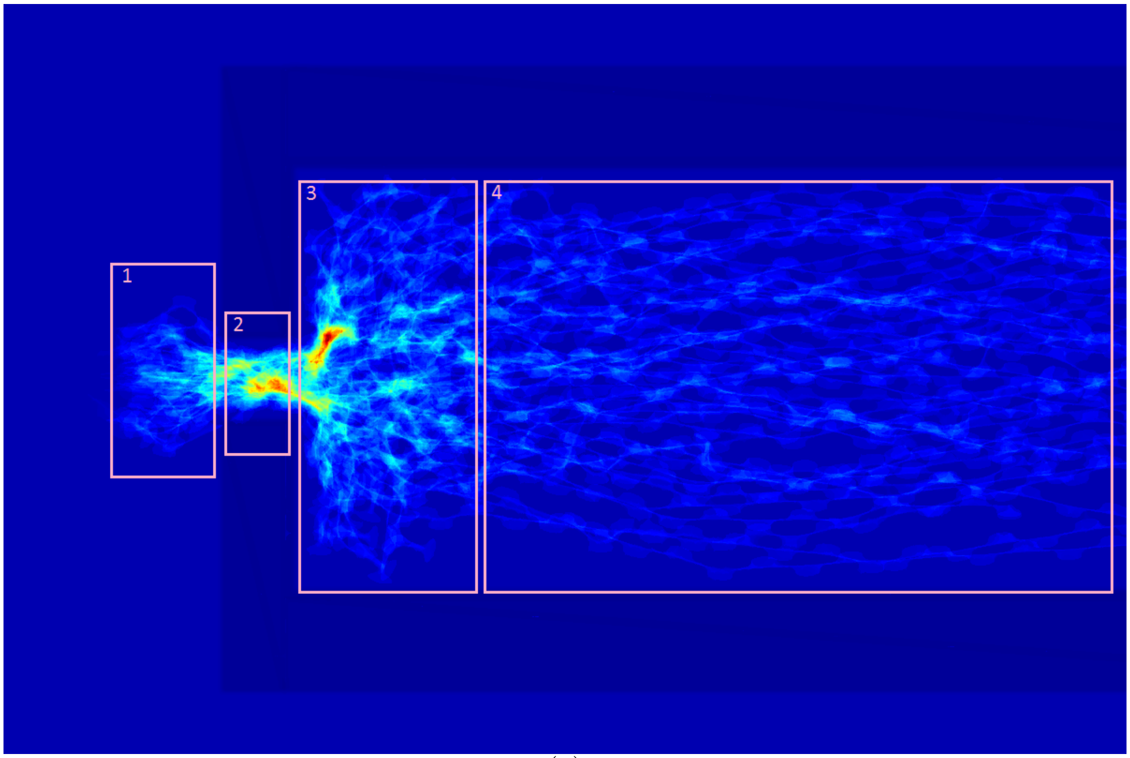


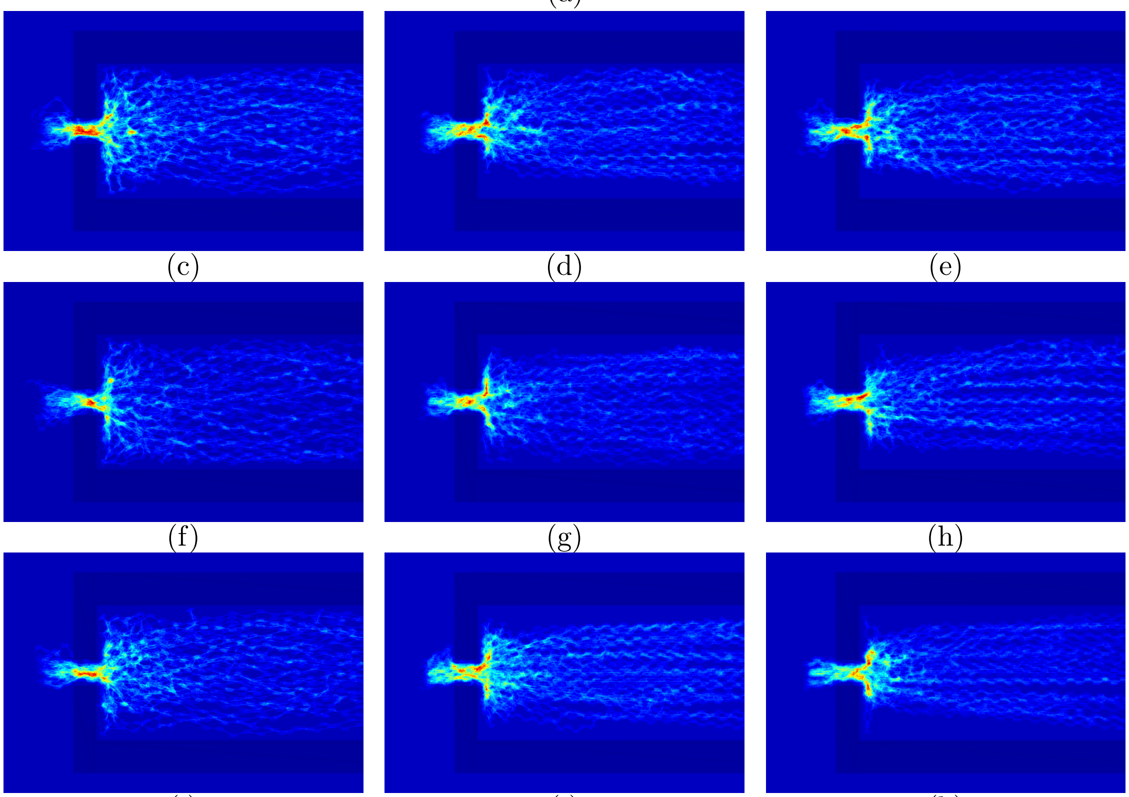
Figure 5.14. Fundamental diagram of total crowd path length over model mixtures. The lines and shaded areas represent the means (μ) and standard deviations (σ) respectively. Detailed results can be found in Appendix B.

5.5.4 Discussion

This experiment completes the series of experiments investigating the impact of heterogeneity in locomotion biomechanics. There are two notable outcomes of this study which impact the area of safety-critical analysis using synthetic crowds. The first is the behaviour of the effort metric over the heterogeneity mixture levels with respect to the models. The velocity reduction models (TGA_velocity & velocity) differ significantly from the normative velocity models (normative & TGA) in their effort outcomes. This is expected as the effort measure is predicated on velocity changes over the path integral of each agent, thus scaling velocity would scale the effort metric overall. However, the velocity model and the TGA_velocity model differ significantly between 10 and 20 agents, or 20% and 40% respectively, heterogeneity mixture levels. This result is very interesting as it is likely rooted in normative assumptions in the measure about the agent's being simulated. The effort metric assumes the values of two biological constants which are generally set to empirical averages for normative human walking. In contrast, it is known that TGA is an inefficient gait strategy, or at least leads to higher effort. In particular, post-stroke TGA (Finley & Bastian, 2017; Platts, Rafferty, & Paul, 2006; Michael, Allen, & Macko, 2005), post-amputation TGA (Mahon, Darter, Dearth, & Hendershot, 2019), and even perturbed normative walking (Ellis, Howard, & Kram, 2013) lead to higher metabolic energy costs. This result then shows a need for defining the biological constants of effort on a per agent basis, or, in the case of non-linear energy consumption, redefining the effort path integral.



(a)



(c) (d) (e)

(f) (g) (h)

(i) (j) (k)

Figure 5.15. Aggregate flux heatmaps for the best (by flow rate) egress scenarios (crowd initialized at right, egress bottleneck at left). The default model is shown in (a) to illustrate areas of interest. The first column (c, f, i) show the TGA model results for the 30 (60%), 40 (80%), and 50 (100%) mixture conditions. Similarly, the TGA_velocity (d, g, j) and the velocity (e, h, k) model results are shown in the middle and right columns respectively.

The second notable result of this study is the differentiation between the TGA model and the normative, TGA_velocity, and velocity models as well as the differentiation of the TGA_velocity and velocity models with respect to the flow rate measure. The TGA model and the normative, TGA_velocity, and velocity model differentiation becomes significant between 1 (2%) and 5 (10%) levels. The TGA_velocity and velocity model differentiation becomes significant between the 10 (20%) and 20 (40%) levels. The means that under these conditions, and considering these synthetic models of TGA, that velocity attenuation alone is not an efficacious proxy for the simulation of populations with TGA symptoms.

The qualitative results highlighted in Figure 5.15, reveal interesting patterns within models and across mixture levels. In Figure 5.15 areas for salient behaviours are delineated, with 4 being the rightmost and earliest point in the scenario and 1 being the leftmost and last point in the scenario. In order from beginning to end: 4 is both crowd instantiation and movement toward egress point; 3 the area before the egress point which form a salient region because of the effects of the bottleneck; 2 the bottleneck itself, a doorway 1.37m wide; and 1 the egress point and goal area within 2m of the bottleneck. It should be noted that these are the qualitative results of synthetic crowds in an environment and scenario structured to be problematic. The qualitative analysis will focus on the emergent crowd behaviours with respect to model, but the scenario itself is the problem (that is, an element of building design under a capacity it potentially can not handle well). Without this type of analysis, with the scenario under diverse model conditions, one may make conclusions about means to fix or improve the egress design (for example optimizing a flow altering obstacle). However, these improvements would only be predictive under a limited range of uses, *read: normative users*.

In the TGA model results, there is a general broader use of space in area 4 that increases with the mixture. Because of this, there appears to be more uniform spread of space use in area 3, which appears as radiant lines of constructive interference in the aggregate flux. Agents with the TGA model are moving to the edges of the available space and entering a state similar to under damped PD controllers. This appears to be due to the wide non-paretic step length in the face of an undersized doorway for this capacity. Most of the space use over the scenario is clearly concentrated within the bottleneck in area 2. In area 1, several agents appear to make circuitous exits to their goal to avoid other emerging agents. In general, TGA appears to produce noisy paths, space filling behaviours, under damped oscillations, and wide avoidance paths.

In the TGA_velocity and velocity results, there is generally less noise (oscillations, and non-linear paths) than in the default and TGA models. In area 4, most paths are very straight and low noise. In both models, there appears to be less utilization of the entire hallway width than the TGA and default models. In area 3 and 2, velocity and TGA_velocity methods differ, in a subtle but serious way. While both models produce a funnel like flux pattern, with strong symmetric utilization at the edges of the entrance to the bottleneck in area 3 connected to a strong movement in the bottleneck in area 2. Similar to the TGA model, but at a smaller scale, the TGA_velocity model produces oscillations and side stepping in the mixture of agents. This produces high frequency noise patterns in the flux and ultimately reduce flows at key points in the bottleneck. At the exit point, in area 1, the TGA_velocity model is somewhat chaotic, while the velocity model is very symmetric and mostly linear.

An interesting artefact that motivates future work is that the noise in the paths appears to lessen as the mixture becomes more and more homogeneous. The asymmetry of in the TGA_velocity model appears to exacerbate this behaviour. This may be evidence that the differing of models, the mixture of different types of synthetic gaits, produces numerous micro-conflicts. The resolution of these micro conflict between models appears to be oscillations and non-linear paths. This may be where the impact on path and flow rate stems from. Future work will explore this impact in real crowds and how to build environment and safety-critical building elements to support these types of crowds.

5.6 Discussion

In Canada, as of 2012, an estimated 7.2% of the population has a mobility disability, and this prevalence increases with age (Statistics Canada, 2015). In neurological hospitals, upwards of 60% of patients may have disordered gait (Stolze et al., 2005), and the prevalence in community-residing older adults is upwards of 35% with the chance of incidence increasing with age (Verghese et al., 2006). These statistics reveal how important it is to build inclusive environments, make inclusive policies, and develop inclusive media. If our current proxies are not adequate for these purposes, as this study shows in a limited and controlled way, then we must develop better models when representing these populations.

5.7 Conclusion

This chapter sought to examine if heterogeneity in locomotion biomechanics has a discernible impact on the layers involved in steering models of synthetic crowds. These are namely: the action space; the collision space; and the emergent crowd space. It is clear from the studies in this chapter that heterogeneity has an impact at each level.

Limitations of this study are the scope of both scenario and locomotion biomechanics. The experiments are purposefully limited to control for confounds and to derive the models from well studied conditions that are both prevalent and impact the given scenario or class of scenarios. in This case, I focus on patients post-stroke with TGA in a common safety-critical scenario, the bottleneck egress. This means the results have limited generality, however, it is strongly implied (given the impact in all three layers of the steering model) that results similarly differentiate under other conditions. Additionally, the models built to simulated TGA conditions are not validated with real world data. Instead, the parameters are carefully constructed from analyses coming from a long history of investigations in to the effects of TGA in patients post-stroke. The velocity model, as well, is unvalidated, but uses these same parameters without the bifurcation of paretic and non-paretic side steps. The modelling approach is described in Chapter 6, however, direct validation of the models (fitting and error analysis between actual walkers and the simulated agents) is left for future work.

A common theme in these studies has been the importance of heterogeneity at different scales. What this sort of language erases here is that the end result of this direction of thinking is not heterogeneity of synthetic agents–it is diversity, and specifically diversity in representation. The importance of this is: if you as a person are not captured by a synthetic crowd model; if that synthetic model is used to make decisions; then those decisions may impact your quality of life and the equity of your access to things associated with those decisions. That is to say, diverse mobilities exist, and they must be accounted for and done so in a way that captures their intricacies. People with non-normative gaits and mobilities know that they exist, and I, as a researcher investigating this dimension of synthetic crowds, am certainly not 'discovering' them. My hope with this study is to reinforce, in the way required by the sciences, the humanity and importance of these people who are so often overlooked or even hidden from view to the detriment of everyone. That is, this study takes a critical standpoint regarding diversity in crowds and from this standpoint we see that diversity impacts outcomes at all levels of steering modelling. Thus, we must consider a critical standpoint in the adoption of synthetic crowds as an analytic methodology, and likely a content creation methodology. This means at least recognizing the limitations of our approaches and ideally developing more representative ones.

Chapter 6

Toward Capturing Heterogeneity

This chapter reviews the findings of and methods used in the dissertation to build a picture of the direction synthetic crowds is heading. Specifically, I delineate: the method used in Chapter 5 Section 5.3 for extracting the features of the footstep action space; the method used in Chapter 5 Sections 5.4 & 5.5 for modelling temporal gait asymmetries in a space-time planning steering model; the findings and possible resolutions for measuring effort of agents with asymmetric gaits from Chapter 5 Section 5.5.

6.1 Seeing the Footstep Action Space

In Chapter 5 Section 5.3, I use a method devised for "reverse-engineering", or seeing, the footstep action space. In step planning based crowd simulation the footstep action space is used to generate footstep plans. In this case, a reverse process is devised as a means of extracting the action space as feature vectors for each step from a data source. This method is a relatively straightforward set of assumptions and transformations with some fitting procedures to produce the 9-dimensional footstep action space for every footstep in a given data stream–in this case the skeleton of motion capture based data (La, n.d.). This tool has multiple uses. The first is simply extracting step information in the form of footstep action space feature vectors. The second use unsupervised machine learning in the form of clustering, and exploring the clustering of, this type of data. The third use is exploring and visualizing the footstep action space of data source.

The process, outlined in Figure 6.1, requires some input data source of human skeletal motion (including feet) which, in turn, requires some basic labelling. The most important labels are the foot front and back, as well as the Centre of Mass (COM). The COM may be: derived from direct measurements as part of the data source (a per person COM); approximated by assuming a uniform distribution of weight in the void space of a 3D model and computing the mean mass point; or approximated by hand using the skeleton of the recording. Generally, these labelled points become child transformations of the skeleton such that their motions are the result of the data sources motions.

Once the key points are labelled, the data source is tracked throughout its frames. In the system, this becomes an animation of the skeleton. As the animation proceeds, the COM point is projected onto the ground plane and the foot markers are monitored for foot down conditions. These conditions include: *foot flat*–both front and back foot labels are within some epsilon of each other; *foot down*–the front and back foot labels are both below a height value (ground plane); and *foot moved*–the current foot has moved more than some delta value. When all of these conditions have been met a foot flat condition is recorded. The parametrization of each condition allows the user to account for noisy data sources or badly tracked feet. In the poor cases, a number of methods may help increase accuracy and detection. These include: manual clean-up of the data source; filtering or smoothing of the relevant label positions; and, in the case of missed footsteps, using first order derivative approximation methods, like central differencing to find the vertex point of the COM arc. In the latter method the foot features may be estimated directly using some basic transformations, or may informed by data or models to approximate the relative foot placement and angle.

At each successful foot flat condition a number of conditions are checked to ensure there is enough information to extract the footstep action space parameters. These conditions are based on footsteps accumulated in to strides–there must be: more than two steps, the previous foot can not be the same as the current; and the current foot must be the same as the step before the previous. Clearly hopping or uni-pedal movement strategies are not captured here, but for the studies in this dissertation, these are adequate. Generally, these conditions can easily be changed or adapted for use with a wide range of locomotion strategies.

Once steps are detected and enough information is available, the foot, step, and stride features are trivial to extract. The COM features require processing. The COM features are defined by a parabola in local space. First the relevant COM trajectory points for the current foot are extracted based on foot flat condition timings. The first feature, COM arc length, can be estimated by summing the polyline segment lengths formed by the segmented COM trajectory points. The accuracy of this measure is then dependent on the frequency of the COM sampling and the speed of the person being tracked. In these studies, the data source participants walked at or less than the average human walking speed, approximately 1.3m/s, and the sampling frequency of 200Hz proved more than adequate. The second COM feature, the COM parabola α parameter, requires the COM trajectory to be placed in a local space. This requires a set of assumptions and transformations based on the shape of the COM trajectory. This process is outlined in detail in Figure 6.2.

Finally, the tool provides an interface to several unsupervised machine learning methods. The current version provides: kMeans, kModes, Guassian Mixture Model, Mean Shift, and Binary Split. Clustering analysis in this tool is typically done in two passes. The first pass involves extracting all the footstep feature vectors from whatever data source is provided. Once data is collected, the user may choose the clustering method and its parameters (typically k). The system will internally cluster all captured feature vectors. The second pass is optional, when the user may play back this data to see what footsteps are grouped together, or use the plotting tool to understand the distribution of footstep action space parameters in the clusters.

It is important to note that the data source for this method may be anything that tracks bipedal human skeletons and can transform them into world space. For pure computer vision based methods this usually means some scene understanding as well for the world space transformation. However, these methods are becoming more and more

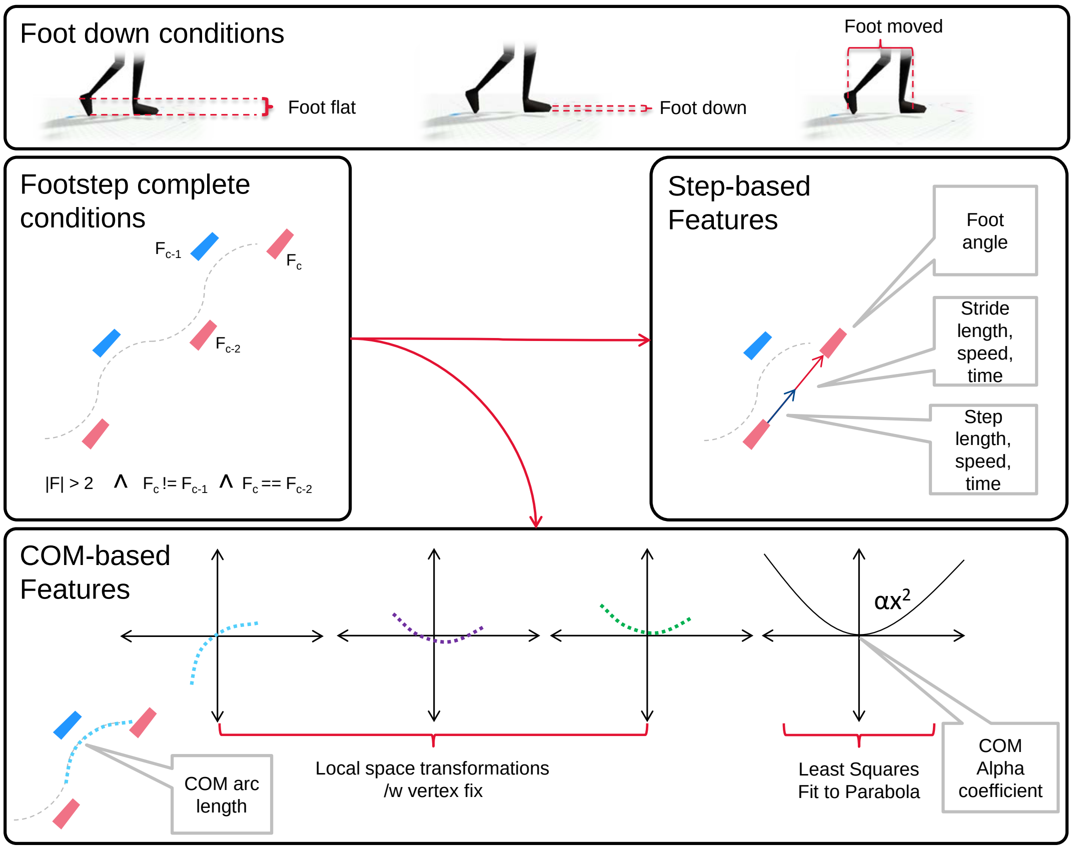


Figure 6.1. Overview of the process for footstep action space feature extraction. Foot down conditions provide enough information to trigger step and stride completion conditions. At each stride completion, step-based features are extracted. These features inform the COM trajectory segmentation and afford COM-based feature extraction. The segmented COM trajectory undergoes transformations to place it in local space. It is then fitted with a simple curve $y = \alpha x^2$ to approximate α which determines the shape of the COM's parabolic projection on the ground plane. Figure 6.2 outlines the details of the COM trajectory processing.

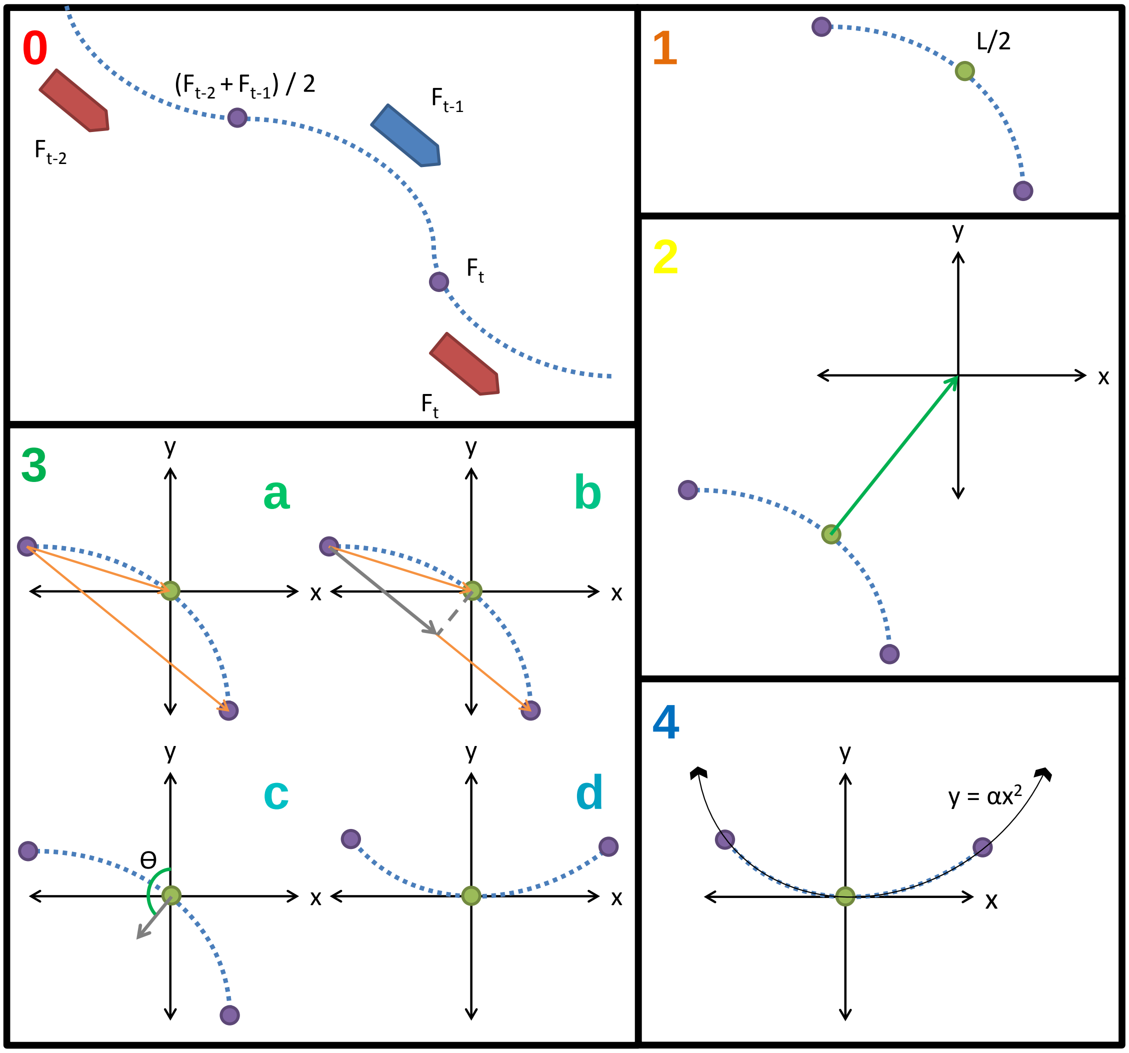


Figure 6.2. Overview of the COM trajectory process. In panel 0, a typical example of a complete stride. In panel 1, the centroid of the COM parabola is estimated using the arc length. In panel 2, a translation is performed to place the estimated vertex at the origin. In panel 3, the opening direction of the parabola is estimated by: (a) forming the vectors to the estimated vertex and the arc end points; (b) projecting the vertex vector on to the endpoint vector and subtracting the vertex vector from the projected one; (c) finding the smallest rotation to align the vector from (b) with "up"; and (d) rotating all arc points (this step can be augmented with a corrective translation to fix badly estimated vertices). Finally, the parabola, now in local space, is fitted with a simple curve $y = \alpha x^2$ to approximate α. Overview of the steps for the feature extraction process in Figure 6.1.

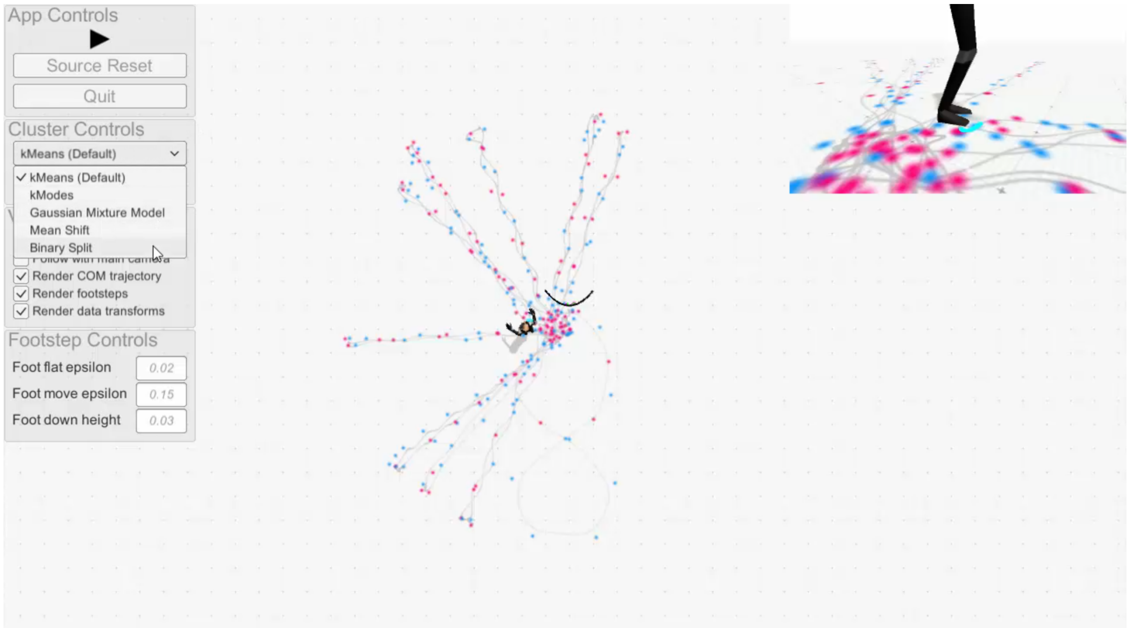


Figure 6.3. Interface for footstep action space feature extraction and clustering.

ubiquitous. For example, the OpenPose (Cao, Hidalgo, Simon, Wei, & Sheikh, 2018; Cao, Simon, Wei, & Sheikh, 2017; Simon, Joo, Matthews, & Sheikh, 2017; Wei, Ramakrishna, Kanade, & Sheikh, 2016) and ArtTrack (Insafutdinov et al., 2017) projects represent the current state-of-the-art in this area. The value of OpenPose and projects like this, includes the annotation and training of foot detection models–an often overlooked portion of motion capture.

6.2 Measuring Heterogeneity

This section expands on some speculations regarding measurement in the areas of this dissertation. The first subject is the direct measurement of heterogeneity of a data source. The second subject is the modification or redefinition of previously used metrics which have been shown to be less than adequate.

6.2.1 Locomotion Heterogeneity as Diversity

The direct measurement of heterogeneity poses and interesting problem while opening a large space of exploration in futures works. Using the method outlined in Section 6.1, it is possible to see heterogeneity. If this method is applied over a more than mocap, as suggested by the use of methods like OpenPose, in the setting of crowds, it is possible to see the heterogeneity of a population of environment users. The applications of this are numerous, but there are critical questions to ask about this approach. The applications include: understanding the mobility of a population (smart city and urban planning; adaptive routing and signage for diversely mobile populations (hospitals and care homes); event planning; accessibility validation. However, there are also critical questions we must ask, and I motivate these for the reader with an anecdote recounted from my time working abroad, a colleague said,

> When visiting [country name] a few years ago I noticed that there were many people in wheelchairs moving about the city. I asked myself what had [country name] done to these people, why was there so much disability in their population. I noticed this elsewhere as well as I travelled. When I returned home, I began to notice the sidewalks, the stairs, the places people in wheelchairs could not go. I realized these countries did not produce more disabilities, I realized my own was hiding them–forcing people with disabilities to only access small portions of the city by not making the city accessible.

This anecdote was spurned by a conversation we were having about the work in this dissertation and it has stuck with me. It suggests that our environment can skew our perception of underlying systemic issues. Even the most well-meaning people may be shielded from the realities of the lived experiences that do not match their own. In this suggested application, if we are to, say, measure heterogeneity, we must remember that we are only measuring the endpoint of a system. This system may hide or obscure an underlying series of assumptions and barriers. This is what much of this dissertation is about–visibility. If the representation is lacking or absent then our decisions are at minimum biased and problematic and at worse wrong and harmful. So while many of these applications hold great promise, they must be taken with care. Perhaps the most interesting is accessibility validation because we can examine the distributions of mobilities in a population to see what is *missing*. If an environment or event does not have a diverse mobility population in an expected proportion to known values, then this is strong evidence that there is a need to assess what and where barriers still exist. The other application, and perhaps closely related to this thesis, is data diversity verification. If a data source (crowd simulation, motion capture database, etc) is being used in the application of synthetic crowds (built environment testing/optimization, policy development, or digital animation), it can be analysed to find whether the source has good coverage of expected mobilities or diversity prior to use.

There are several possibilities for heterogeneity, or diversity, metrics that range from basic high level statistics to machine learning based methods. For a data set of footstep action space feature vectors, the variance in features may be a high level indicator. More specifically, the per person variance of per foot features may be an indicator of either environment details (like foot angle variance in an environment that has a corner) or diversity in mobility (like asymmetrical variance or mean values in foot angle and step time features). Separating person-environment interactions from within and between person mobility interactions and mobility-environment interactions is a non-trivial task. For a clustered data set, there are even more possibilities for analysis and visualization. Beyond simple summary statistics or cluster centroids, there are several metrics for clustering quality with respect to diversity. These include: separation (e.g. between cluster Sum of Squared Errors, graph node weights); joint cohesion and separation (Silhouette Coefficient (Rousseeuw, 1987)); richness (akin to the K value in kMeans, fund using Silhuoette Coefficient, Elbow Methods, Gap Statistic (Tibshirani, Walther, & Hastie, 2001), etc); Shannon index (Tuomisto, 2010; Shannon, 1948); and numerous other indices or diversity statistics. Generally speaking, many of these methods have either been applied broadly as internal cluster validations or in ecology where understanding species distributions is important.

6.2.2 Normative Crowd Measures

An important outcome of the experiment outlined in Chapter 5 Section 5.5 is strong evidence that the total metabolic energy expenditure, or effort, of non-symmetrical gait patterns is more than that of symmetrical gaits. The literature strongly supports this result, specifically, that asymmetry in hemiplegic patients post-stroke induces a higher energy expenditure (Kramer, Johnson, Bernhardt, & Cumming, 2016). Handling this difference accurately in the application of synthetic crowds is very important. With respect to mobility, the more important factor between speed and distance in community functioning of post-stroke patients is distance (Combs et al., 2013). This is likely the role of the community design and the built environment in the disablement process(Renalds et al., 2010; Clarke et al., 2008; Clarke & George, 2005). The energy cost of walking is considered one of the primary factors in the distance a patient post-stroke, both younger and older, can walk (Platts et al., 2006; Zamparo et al., 1995). This warrants a closer investigation of the effort metric.

The definition of the effort metric (single agent version of Equation 4.3 from Chapter 4 Section 4.2.2) is the path integral for total metabolic energy expenditure during walking:

$$E = m \int (e_s + e_w |\mathbf{v}|^2) dt \tag{6.1}$$

where m is the mass of the agent, e_s and e_w are biological constants set to empirical averages for human walking, 2.23 J/Kg·s and 1.26 Js/Kg·m^2 respectively, and v_a is the velocity of the agent. Herein lies a normative assumption, both e_s and e_w are based on averages for normative walking over level ground (Whittle, 2014). The value e_s is the rate of base metabolic energy expenditure per second and grows depending only on time. The value e_w is the rate metabolic energy expenditure of walking and grows depending on the agent walking velocity and time. Assuming that e_s remains fixed, which may not be true, the average percent difference of the metabolic energy expenditure is 12.76% so e_s becomes 1.420776 for post stroke patients with hemiplegia (Zamparo et al., 1995). However, this assumes that: 1) e_s can accurately remain fixed for patients post-stroke; 2) that the derivation of measurements of the energy expenditures can be compared; and 3) that this average difference is suitable (note that the energy expenditure cost is directly related to speed). To further this conversation, it is clear from past experiments that the energy expenditure is not simply an average scalar value. In fact, there appears to be a power law relationship for both normative and hemiplegic energy expenditure with respect to speed (Zamparo et al., 1995). This can be seen in Figure 6.4. Further work in this area is required to define more appropriate and potentially adaptive metabolic energy expenditure path integrals for use in the analysis of diverse synthetic crowds. This is particularly important for safety-critical scenarios as there may be large variations in speed as bottlenecks form, panic ensues, and differences in walking patterns are resolved via social norms (sidestepping, assistance, etc) or safety practice (mobility routing, care staff, etc). Additionally, as noted in Chapter 4 Section 4.3.4, the measure as defined does not necessarily capture the energy cost of torques.

6.3 Modelling Heterogeneity

This section briefly delineates the key implementation details of the *footstepAI* model used in this dissertation, and then approach used in Chapter 5 Sections 5.4 & 5.5 as a means to further the conversation on modelling in this space. The *footstepAI* model is implemented with an interesting use of the A* shortest path planner. To provide solutions to the problem of planning of multiple sequential footsteps with multiple objectives, as described in the original paper, the author's convert the problem into a just-in-time graph

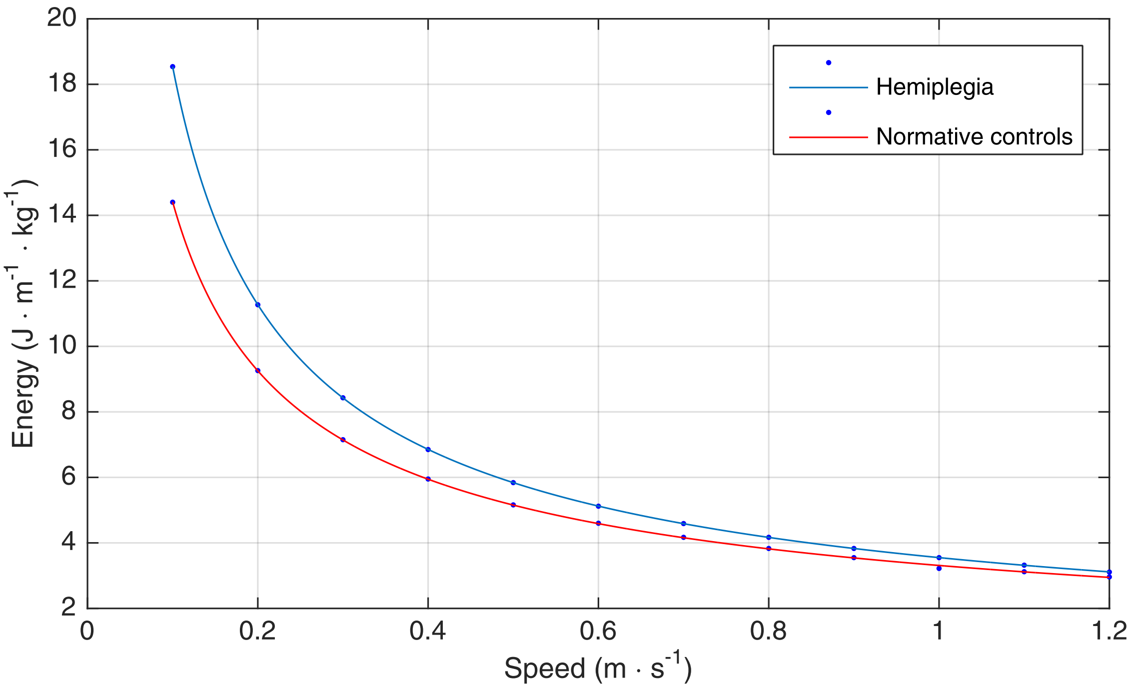


Figure 6.4. Estimated power law relationships between speed and energy expenditure in normative controls and patients with hemiplegia–data from (Zamparo et al., 1995).

exploration. From a high level, the footsteps planner, given one agent, plans a path for the agent as a piecewise footstep plan, where each piece is a complete footstep. The number of steps to plan is a sort of time horizon, where typically the value is around 10 steps. Each step, as the plan is built must minimize three energy cost functions: 1) the cost of a fixed energy expenditure (this is to reduce the total time spent trying to reach a goal); 2) the cost of counteracting the instantaneous loss of momentum during the heel strike phase (this is to measure the effort required to maintain a desired speed); 3) the change of momentum over the step trajectory (this makes straighter steps preferential at the desired speed and slower steps preferential when changing direction). In the just-in-time graph solution of this problem, at each step, a series of nodes are created which represent various types of possible steps from the discretized footstep action space–the footstep action space is described in detail in Chapter 5 Section 5.3. To simplify this search, the implementation uses a large set of normative footstep action prototypes during the search. Each of these footstep actions is tested for collisions (either culled or incur a high cost penalty if they occur along the step trajectory) and the total of the cost function. The resultant graph after the number of nodes to expand and number of steps to plan are reached is then searched using the A* path planner.

To implement the models in Chapter 5 sections 5.4 & 5.5, I interject at the point of the just-in-time node expansion of the implementation. Each time a footstep is evaluated for its cost and collisions, a higher level model is passed the current state of the agent. This state includes the current foot flag and the parameters of the footstep being evaluated. Primarily the focus of the modelling in that study is the representation of hemiplegia in community ambulating patients post-stroke. This means the models primary function is to bifurcate the parameters such that the paretic side and unaffected aside differ. This difference produces spatiotemporal gait asymmetries reflected in both the stepping patterns and th COM trajectories. It is important to note that the main reason for using scaling factors for temporal parameters rather than concrete values or parameter bounds is that any given step passed from the representative normative set may fulfil a particular function, such as turning, back stepping, sidestepping, etc. Scaling factors based on empirical values solve the issue of robustly handling the various step types. For the TGA models the higher level model simply checks the current foot state and switches out the parameters noted in Chapter 5 sections 5.4.1.2 & 5.4.1.3.

While this model provides a high fidelity and highly modular means of producing biomechanically based synthetic crowds including agents modelling spatiotemporal gait asymmetries, it has several key drawbacks as implemented. One important issue to note is that in the models presented in Chapter 5 Section 5.5, I am primarily aiming to model the temporal parameters of asymmetry symptoms found post-stroke. However, as noted in that study, to make the optimization tractable with the same underlying footstep planner model, it was necessary to constrain certain spatial parameters, such as minimum & maximum step length. There is some evidence that the spatial (length) and temporal (time) parameters are correlated, but the correlations are weak to modest (Patterson, Gage, Brooks, Black, & McIlroy, 2010b). Additionally, there is evidence that patients post-stroke can adapt spatial and temporal asymmetries independently (Malone & Bastian, 2014; Malone, Bastian, & Torres-Oviedo, 2012). This is to say the model, as presented, serves the purpose of presenting possible walkers based on empirical evidence, but that there is strong reason to create new footstep planners which separate the mechanisms controlled by these parameters. The future of this area of modelling includes such approaches as using probabilistic models for irregular stepping patterns; space-time planning models for differential drive controllers (for footsteps + wheelchairs); physical models for mobility devices and persons/environment interactions; and machine learning based models to remove the dependence on an underlying crowd model. A particularly important goal would be to produce a sort of fidelity agnostic panacea physical model that is flexible enough to capture the range of not only gait disorders and interesting mobilities but also integrates physically robust simulations of mobility devices.

6.4 Conclusion

In this chapter I explore many of the questions that arise from the findings in the dissertation. I make preliminary attempts to either address them or provide further evidence from the literature of their importance. My hope with this chapter is not to provide solutions to issues in crowds but to strongly suggest important avenues the future of synthetic crowds may take. In doing so, I have covered: a method and system for extracting the footstep action space features from a a data source and performing unsupervised learning on the data set; the importance of capturing and measuring the diversity of mobilities including the diversity of a population from a pattern recognition and information theoretic approaches derived from ecology to the importance of metric derivation such as effort when measuring diverse crowds; and finally, the modelling diverse mobilities and the limits if current modelling through the inner workings of the *footstepAI* models used throughout this dissertation.

Chapter 7

Conclusion

This dissertation has endeavoured to examine the underpinnings of steering model selection and how representation impacts outcomes we find when we apply steering models. The dissertation is arranged as a critical standpoint and informal proof for both endeavours. That is, assuming the implied (by way of the literature) universal quantifications that: 1) steering models capture human behaviour and thus reproduce the same human behaviour; and 2) low-cost (computational and implementation complexity) heterogeneity proxies capture the outcomes of diverse crowds; then there exists counterexamples which falsify these statements. This dissertation provides these counterexamples by way of extensive studies each framed by a critical review of the literature and assumptions made. This chapter reviews what these studies have found, and how those findings challenge common practices in the application of steering models. Following these discussions, I explore and delineate the limitations of the studies and their findings. Finally, I follow-up with future works to address these limitations and explore directions which this area of work could take.

7.1 Findings

The dissertation is primarily broken into two extensive studies. Since the topic of the dissertation is the importance of biomechanical heterogeneity in the space of synthetic crowds and their applications, I build toward the major findings of the dissertation by separating the two core concepts–biomechanics and heterogeneity. In the first study, I examine the importance of biomechanics by constructing, both from the literature and new real world examples, the largest single comparative synthetic crowds study ever conducted. This study builds over the course of five experiments which increase in scenario complexity. The study examines four ubiquitous steering models which are commonly found in games, animation, research, and analytics which employee synthetic crowds. By normalizing the models beforehand under the same conditions and for the same fitness function, they can be directly compared to understand how their underlying models and built-in assumptions perform under the battery of scenarios. The key finding in this study is that no one model is panacea. Even when normalized the models produce different quantitative and qualitative results. The next major finding is that the models underlying mathematics and assumptions produce shortcomings under specific conditions–no one model is capable of 100% success across all scenarios. These shortcomings naturally delineate the applications of where these models are successful. Some models are perfect for gaming as they are computationally inexpensive, produce nice qualitative results, and would overcome there shortcomings in gaming environments where other systems or mechanisms fix certain assumptions (such as physics engines, navmeshes, higher level AI etc). Some models are excellent for rapid analysis as they produce emergent behaviours found in real crowds. However, the key insight here was that fidelity had a large impact on outcomes, both quantitative and qualitative. In particular, biomechanical modelling affords a level of analytics not provided by other approaches, namely that across scenario complexities they produce expected direct and emergent behaviours and quantitative outcomes which reflect the underlying human nature of certain scenarios.

In the second study, I examine the impact of heterogeneity on outcomes derived from synthetic crowds by examining the three levels of outcomes where steering decisions have an impact. These levels of impact are the actions space, the collision space, and the emergent crowd space. At the action space level I examine real world data of normative and stylized walking in the form of motion capture and derive footstep action vectors. The main finding at the action space level was that footsteps are separable into distinct regions of footstep action space and that this is especially true of stylized, non-normative steps–thus these actions produce action spaces of distinctly different character. At the collision space level, I reuse the biomechanical model from the first study and introduce a higher level models which perturbs the footstep action planning. I specifically focus on patients post-stroke as several prior studies make claims in the space of evacuation analysis of hospitals and care homes whose populations would include these people, but use velocity reduction as a proxy for their mobility modelling. This focus also allows for the empirically based derivation of footstep parameters and constraints as there is extensive literature on the analysis of spatiotemporal gait asymmetries of patients post-stroke. Using the previous literature as a backbone, I develop three new models which model: temporal gait asymmetry, velocity reduction, and temporal gait asymmetry with velocity reduction. The main finding of this study is that the collision space of all models differs in shape, amplitude, and frequency. Since all steering models attempt to produce collision free trajectories for agents and this necessarily relies on the collisions space, the finding implies that they emergent collision avoidance strategies, and thus emergent crowd behaviours and outcomes, would differ as well. At the highest level, crowds outcomes, I test this implication in a carefully controlled study of a ubiquitous synthetic crowds scenario. The scenario is important as it is not only used across crowd simulation literature but is also a sub-feature common to many buildings and has been responsible for the deaths of people in real world tragedies–the bottleneck egress. By examining common crowds outcome measures over increasing levels of heterogeneity (mixing in more of a particular agent model), I show that the biomechanical heterogeneity of agents in the makeup of a crowd significantly impacts the quantitative and qualitative outcomes of a safety-critical crowd scenario.

7.2 Implications

The implications of the dissertation can be summed up simply: diversity matters. If we are rendering media, producing analytics, drafting policies, or generally making decisions which impact peoples lives, and we are using synthetic crowds to do so, then the diversity of those people we are simulating must be taken into account.

In study one, I devise a novel hybrid approach to comparative analysis of synthetic steering models. The approach combines optimization as a means of normalization between models of arbitrary parameter dimensions (Wolinski et al., 2014a) and the use of optimal ratio metric minimization (or maximization) (Berseth et al., 2014). Additionally, this approach combines and sorts prior and new benchmarks by scenario difficulty into the largest, non-procedural, benchmark set for in-depth analysis of crowds. The implications of this are two-fold: first models can now be comparatively analysed in ecologically valid scenarios without ground truth; second the result of comparative analysis can feed future research results (see Section 7.4).

In study two, I show in a very controlled case that the diversity of simulated population impacts crowd outcomes at every level of steering. Since the heterogeneity is introduced at the very lowest level, the action space, and the results show significant difference between all controls, the implication is that the non-linear modelling of gaits is important in crowd simulation outcomes. This means that, particularly in safety-critical analysis, that an analysis of expected population must take place before and during the use of synthetic crowds to inform decisions.

7.3 Efficacy & Limitations

The studies presented in this dissertation are primarily limited by scope. The first study examines only four steering models. There are several models available to the synthetic crowds research scientist and most commercial software have one in particular they use. The approach in this dissertation was to select several common models which are ubiquitous in a few key areas. The first study also carefully removes heterogeneity from the models. This step proceeds down to the individual agent level, and is artefact of experimental control for this dissertation. That is, in certain cases where agents would have heterogeneous desired locomotives (again this is the primary heterogeneity proxy), I have homogenized the desired velocity. I wanted to explore biomechanics separately from biomechanical heterogeneity to build towards a particular research arc. In practice, this benchmark would not do this. It is important to keep this alteration of desired speed when comparing models. Additionally, I removed scenarios from prior benchmarks which relied solely on this–namely the *overtake* scenarios of the Simple Interactions experiment (Singh et al., 2009).

The second study is limited in scope because there are numerous sources of interesting gait patterns. Temporal Gait Asymmetry is a nice case study because it specifically relates to linearity of step and COM patterns at the lowest level. In this way, biomechanical heterogeneity can be introduced in the lowest level of footstep planning. Similarly, I analyse the footsteps of two particular subjects in a large motion capture dataset. This is partially because direct analysis of motion capture can be difficult without professional cleaning, but also motion capture of people with disordered gaits is difficult to obtain because of the need for facilities and the control of participants with co-morbid conditions, medication, availability, and varying comfort. In this dissertation, I use data collected, analysed, and published to make models of a particular condition.

7.4 Future Work

The future work stemming from the findings in this dissertation are numerous. There are several areas involved including: comparative crowds analysis; crowd and agent outcome measures; and diverse crowd modelling.

Within the area of comparative crowds analysis the main future work would be to create a benchmarking portal for both researcher and practitioners to make use of. This would include supporting software for simulation, data acquisition, and analytics as well as interfaces for exploring comparative results.

Within the area of crowd and agent outcomes measures, there is an entire area of research to be explored further. Both the selection and definition of crowd measures produce biases in outcomes. At the selection stage, it has to be clear why a particular "thing" is being measured over another. Generally, the approach here would be to measure as much as possible and explore high dimensional datasets for artefacts and interesting findings. This area of work requires the application of findings in big data, high dimensional analytics, and the visualizations of high dimensional datasets. At the definition stage it is important to be critical of outcomes metrics. Many metrics have no inherently built in biases, but may have biases when paired with a particular model (e.g. note that the path length has no inherent bias, but some models have linear movements and some more accurate models have non-linear movements). Then there are more powerful metrics, but they may have biases, such as effort, heat production, etc. The metrics afford in sight into biological effects of crowds but may utilize constants or values derived from medical literature focused on normative controls.

Within the area of diverse crowd modelling the future works are endless. There is an opportunity for modelling all sorts of interesting gaits, mobilities, devices, and interactions which are simply not captured by simplified models. Much of crowd simulation makes a singular approach towards robust emergent crowd behaviours. IN different areas there are different focuses for the approach. In large scale simulations, often the focus is produces basic emergent behaviours with very many agents. In robotics the focus is on producing provably collision free trajectories. In building analytics the focus is on reproducing real world scenarios and emergent behaviours with enough fidelity to assume some generality of the results. In safety-critical analytics the focus is on produce midscale emergent behaviours captured in safety critical scenarios. Some areas, such as games and media focus on panacea models where several layers are involved to produce believable animations. This latter approach has applications in all the other areas, however, it is often strongly driven by human artists and can be costly to produce high fidelity results. In each area, the focus is often on model and its supposed affordances. So the future work in this area could conceivably be to devise a model-free approach. Or, in other words, a model which makes no assumptions about how crowds or people move, but rather 1) learns movement and 2) is robust and general enough in definition that multiple steering modalities may be added and interact such that large diverse crowds may be constructed without the need for costly human intervention.

7.5 Conclusion

This dissertation has covered a particular area of synthetic crowd simulation, the steering model, and revealed through a series of studies the impact of model decisions made at this, the lowest, level of synthetic crowds. The studies open up new areas of research in the field of synthetic crowds and supports future works in these particularly difficult but rich areas from a critical standpoint.

References

Alexander, R. M. (1976). Estimates of speeds of dinosaurs. *Nature*, *261*(5556), 129.

Allen, B., & Faloutsos, P. (2009a). Complex networks of simple neurons for bipedal locomotion. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems* (pp. 4457–4462).

Allen, B., & Faloutsos, P. (2009b). Evolved controllers for simulated locomotion. In *Lecture notes in computer science: Lecture notes in artificial intelligence and lecture notes in bioinformatics* (Vol. 5884 LNCS, pp. 219–230).

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual acm-siam symposium on discrete algorithms* (pp. 1027–1035).

Balasubramanian, C. K., Bowden, M. G., Neptune, R. R., & Kautz, S. A. (2007). Relationship between step length asymmetry and walking performance in subjects with chronic hemiparesis. *Archives of physical medicine and rehabilitation*, *88*(1), 43–49.

Banerjee, B., & Kraemer, L. (2011). Evaluation and comparison of multi-agent based crowd simulation systems. In *Agents for games and simulations ii* (pp. 53–66). Springer.

Benjamini, Y., & Hochberg, Y. (1995). 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.

Bentley. (2019). *Legion.* Retrieved 2019-05-10, from `https://www.bentley.com/en/products/brands/legion`

Berseth, G., Kapadia, M., & Faloutsos, P. (2013). Steerplex: Estimating scenario complexity for simulated crowds. In *Proceedings of the 6th ACM SIGGRAPH conference on motion in games* (pp. 67–76).

Berseth, G., Kapadia, M., & Faloutsos, P. (2015). Robust space-time footsteps for agent-based steering. *Computer Animation and Virtual Worlds*.

Berseth, G., Kapadia, M., Haworth, B., & Faloutsos, P. (2014). SteerFit : Automated Parameter Fitting for Steering Algorithms. In *Proceedings of the ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 1–10).

Berseth, G., Usman, M., Haworth, B., Kapadia, M., & Faloutsos, P. (2015). Environment optimization for crowd evacuation. *Computer Animation and Virtual Worlds*, *26*(3-4), 377–386.

Bertram, J. E., & Ruina, A. (2001). Multiple walking speed–frequency relations are predicted by constrained optimization. *Journal of theoretical Biology*, *209*(4), 445–453.

Boatright, C. D., Kapadia, M., Shapira, J. M., & Badler, N. I. (2013). Context-sensitive Data-driven Crowd Simulation. In *Proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry* (pp. 51–56). New York, NY, USA: ACM.

Brandstater, M. E., Gowland, C., Clark, B., et al. (1983). Hemiplegic gait: analysis of temporal variables. *Archives of physical medicine and rehabilitation*, *64*(12), 583–587.

Brown, C. J., Bradberry, C., Howze, S. G., Hickman, L., Ray, H., & Peel, C. (2010). Defining community ambulation from the perspective of the older adult. *Journal of geriatric physical therapy*, *33*(2), 56–63.

Bruggeman, H., Zosh, W., & Warren, W. H. (2007). Optic flow drives human visuo-locomotor adaptation. *Current Biology*, *17*(23), 2035–2040.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2018). *Openpose: realtime multi-person 2d pose estimation using part affinity fields.*

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Cvpr.*

Champandard, A. J. (2012, April). *Velocity obstacle algorithms and locomotion integration.* Retrieved from `http://aigamedev.com/open/teaser/velocity-obstacle-integration/` ([Online; posted 30-April-2012])

Charalambous, P., Karamouzas, I., Guy, S. J., & Chrysanthou, Y. (2014). A data-driven framework for visual crowd analysis. *Computer Graphics Forum*, *33*(7), 41–50.

Chraibi, M., Ensslen, T., Gottschalk, H., Saadi, M., & Seyfried, A. (2016). Assessment of models for pedestrian dynamics with functional principal component analysis. *Physica A: Statistical Mechanics and its Applications*, *451*, 475–489.

Clarke, P., Ailshire, J. A., Bader, M., Morenoff, J. D., & House, J. S. (2008). Mobility disability and the urban built environment. *American journal of epidemiology*, *168*(5), 506–513.

Clarke, P., & George, L. K. (2005). The role of the built environment in the disablement process. *American Journal of Public Health*, *95*(11), 1933–1939.

Combs, S. A., Van Puymbroeck, M., Altenburger, P. A., Miller, K. K., Dierks, T. A., & Schmid, A. A. (2013). Is walking faster or walking farther more important to persons with chronic stroke? *Disability and rehabilitation*, *35*(10), 860–867.

Conover, W. J., & Iman, R. L. (1979). *On multiple-comparisons procedures* (Tech. Rep.). NM, USA: Los Alamos Scientific Lab.

Conway, J. (1970). The game of life. *Scientific American*, *223*(4), 4.

CORE CROWD LLC. (2019). *Simx.* Retrieved 2019-05-10, from `http://corecrowd.net/simx.html`

Curtis, S., Guy, S. J., Zafar, B., & Manocha, D. (2011). Virtual tawaf: A case study in simulating the behavior of dense, heterogeneous crowds. In *Proceedings of the*

IEEE international conference on computer vision workshops (pp. 128–135).
Dettmann, M. A., Linder, M. T., & Sepic, S. B. (1987). Relationships among walking performance, postural stability, and functional assessments of the hemiplegic patient. *American journal of physical medicine*, *66*(2), 77–90.
Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241–252.
Dupre, R., & Argyriou, V. (2017). A human and group behavior simulation evaluation framework utilizing composition and video analysis. *Computer Animation and Virtual Worlds*, e1844.
El Haber, N., Erbas, B., Hill, K. D., & Wark, J. D. (2008). Relationship between age and measures of balance, strength and gait: linear and non-linear analyses. *Clinical Science*, *114*(12), 719–727.
Ellis, R. G., Howard, K. C., & Kram, R. (2013). The metabolic and mechanical costs of step time asymmetry in walking. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1756), 20122784.
Faloutsos, P., Van de Panne, M., & Terzopoulos, D. (2001). Composable controllers for physics-based character animation. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques* (pp. 251–260).
Feng, T., Yu, L.-F., Yeung, S.-K., Yin, K., & Zhou, K. (2016). Crowd-driven mid-scale layout design. *ACM Transactions on Graphics*, *35*(4), 132–1.
Finley, J. M., & Bastian, A. J. (2017). Associations between foot placement asymmetries and metabolic cost of transport in hemiparetic gait. *Neurorehabilitation and neural repair*, *31*(2), 168–177.
Fiorini, P., & Shiller, Z. (1993). Motion planning in dynamic environments using the relative velocity paradigm. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 560–565).
Fiorini, P., & Shiller, Z. (1998). Motion planning in dynamic environments using velocity obstacles. *The International Journal of Robotics Research*, *17*(7), 760–772.
Fruin, J. J. (1971). *Pedestrian planning and design* (Tech. Rep.).
Garland-Thomson, R. (2005). Feminist disability studies. *Signs: Journal of Women in Culture and Society*, *30*(2), 1557–1587.
Ge, S. S., Li, Z., & Yang, H. (2012). Data driven adaptive predictive control for holonomic constrained under-actuated biped robots. *IEEE Transactions on Control Systems Technology*, *20*(3), 787–795.
Geijtenbeek, T., van de Panne, M., & van der Stappen, A. F. (2013). Flexible muscle-based locomotion for bipedal creatures. *ACM Transactions on Graphics*, *32*(6), 206.
Geng, T., Porr, B., & Wörgötter, F. (2006). A reflexive neural network for dynamic biped walking control. *Neural Computation*, *18*(5), 1156–96.
Geoerg, P., Schumann, J., Boltes, M., Holl, S., & Hofmann, A. (2018). The influence of physical and mental constraints to a stream of people through a bottleneck. In *Pedestrian and evacuation dynamics* (p. 57).

Golaem. (2017). *Gallery — golaem.* Retrieved 2017-06-27, from `http://golaem.com/content/customers/gallery`

Guo, D., Wang, C., & Wang, X. (2009). A hierarchical pedestrians motion planning model for heterogeneous crowds simulation. In *Proceedings of the international conference on information and automation* (pp. 1363–1367).

Guy, S. J., Chhugani, J., Curtis, S., Dubey, P., Lin, M., & Manocha, D. (2010). Pledestrians: a least-effort approach to crowd simulation. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 119–128).

Guy, S. J., Kim, S., Lin, M. C., & Manocha, D. (2011). Simulating heterogeneous crowd behaviors using personality trait theory. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 43–52).

Guy, S. J., van den Berg, J., Liu, W., Lau, R., Lin, M. C., & Manocha, D. (2012). A statistical similarity measure for aggregate crowd dynamics. *ACM Transactions on Graphics*, *31*(6), 1.

Gwynne, S. M. V., & Siddiqui, A. A. (2013). Understanding and simulating large crowds. In V. V. Kozlov, A. P. Buslaev, A. S. Bugaev, M. V. Yashina, A. Schadschneider, & M. Schreckenberg (Eds.), *Traffic and granular flow '11* (pp. 217–239). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hakim, A. M., Silver, F., & Hodgson, C. (1998). Organized stroke care: A new era in stroke prevention and treatment. *CMAJ: Canadian Medical Association Journal*, *159*(6), S1.

Hart, P., Nilsson, N., & Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, *4*(2), 100–107.

Haworth, B. (2016). Towards Computer Assisted Crowd Aware Architectural Design. In *Proceedings of CHI extended abstracts on human factors in computing systems* (pp. 2119–2125).

Haworth, B., Usman, M., Berseth, G., Kapadia, M., & Faloutsos, P. (2015). Evaluating and optimizing level of service for crowd evacuations. In *Proceedings of the 8th ACM SIGGRAPH conference on motion in games* (pp. 91–96).

Haworth, B., Usman, M., Berseth, G., Kapadia, M., & Faloutsos, P. (2017). On density–flow relationships during crowd evacuation. *Computer Animation and Virtual Worlds*, *28*(3-4).

Haworth, B., Usman, M., Berseth, G., Khayatkhoei, M., Kapadia, M., & Faloutsos, P. (2016). Using synthetic crowds to inform building pillar placements. In *Proceedings of the IEEE virtual humans and crowds for immersive environments* (pp. 7–11).

Haworth, B., Usman, M., Berseth, G., Khayatkhoei, M., Kapadia, M., & Faloutsos, P. (2017). Code: Crowd-optimized design of environments. *Computer Animation and Virtual Worlds*.

Helbing, D. (1992). A fluid-dynamic model for the movement of pedestrians. *Complex Systems*, *6*, 391–415.

Helbing, D., Buzna, L., Johansson, A., & Werner, T. (2005). Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science*, *39*(1), 1–24.

Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature*, *407*(6803), 487–490.

Helbing, D., Johansson, A., & Al-Abideen, H. Z. (2007). Dynamics of crowd disasters: An empirical study. *Physical Review E*, *75*(4), 046109.

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, *51*(5), 4282.

Hochberg, J., & Tamhane, A. (1987). *Multiple comparison procedures* (Tech. Rep.). John Wiley & Sons,.

Hodgson, C. (1998). Prevalence and disabilities of community-living seniors who report the effects of stroke. *CMAJ: Canadian Medical Association Journal*, *159*(6), S9.

Hoyet, L., Olivier, A.-H., Kulpa, R., & Pettré, J. (2016). Perceptual effect of shoulder motions on crowd animations. *ACM Transactions on Graphics*, *35*(4), 53.

Hsu, A.-L., Tang, P.-F., & Jan, M.-H. (2003). Analysis of impairments influencing gait velocity and asymmetry of hemiplegic patients after mild to moderate stroke. *Archives of physical medicine and rehabilitation*, *84*(8), 1185–1193.

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

INCONTROL Simulation Solutions. (2017). *Pedestrian dynamics.* Retrieved 2017-06-27, from `http://http://www.pedestrian-dynamics.com/`

Insafutdinov, E., Andriluka, M., Pishchulin, L., Tang, S., Levinkov, E., Andres, B., & Schiele, B. (2017). Arttrack: Articulated multi-person tracking in the wild. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 6457–6465).

Integrated Environment Solutions. (2019). *Simulex.* Retrieved 2019-05-10, from `https://www.iesve.com/software/ve-for-engineers/module/Simulex/480`

Jablonski, K., Argyriou, V., Greenhill, D., & Velastin, S. A. (2015). Evaluation framework for crowd behaviour simulation and analysis based on real videos and scene reconstruction. In *Proceedings of the 6th latin-american conference on networked and electronic media.* IET.

Jiang, L., Li, J., Shen, C., Yang, S., & Han, Z. (2014). Obstacle optimization for panic flow - Reducing the tangential momentum increases the escape speed. *PLoS ONE*, *9*(12), e115463.

Johansson, A., & Helbing, D. (2007). Pedestrian flow optimization with a genetic algorithm based on Boolean grids. In *Pedestrian and evacuation dynamics* (pp. 267–272). Springer.

Johansson, A., Helbing, D., Al-Abideen, H. Z., & Al-Bosta, S. (2008). From crowd dynamics to crowd safety: a video-based analysis. *Advances in Complex Systems*, *11*(04), 497–527.

Ju, E., Choi, M. G., Park, M., Lee, J., Lee, K. H., & Takahashi, S. (2010, December). Morphable crowds. *ACM Transactions on Graphics*, *29*(6), 140:1–

140:10. Retrieved from `http://doi.acm.org/10.1145/1882261.1866162` doi: 10.1145/1882261.1866162

Junior, J. C. S. J., Musse, S. R., & Jung, C. R. (2010). Crowd analysis using computer vision techniques. *IEEE Signal Processing Magazine*, *27*(5), 66–77.

Kajita, S., Kanehiro, F., Kaneko, K., Yokoi, K., & Hirukawa, H. (2001). The 3d linear inverted pendulum mode: A simple modeling for a biped walking pattern generation. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems* (Vol. 1, pp. 239–246).

Kallmann, M., & Kapadia, M. (2016). *Geometric and discrete path planning for interactive virtual worlds*. Morgan & Claypool Publishers.

Kapadia, M., Singh, S., Allen, B., Reinman, G., & Faloutsos, P. (2009). Steerbug: an interactive framework for specifying and detecting steering behaviors. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 209–216).

Kapadia, M., Singh, S., Hewlett, W., & Faloutsos, P. (2009). Egocentric affordance fields in pedestrian steering. In *Proceedings of the 2009 symposium on interactive 3d graphics and games* (Vol. 1, p. 215). New York, NY, USA: ACM.

Kapadia, M., Singh, S., Hewlett, W., Reinman, G., & Faloutsos, P. (2012). Parallelized egocentric fields for autonomous navigation. *The Visual Computer*, *28*(12), 1209–1227.

Kapadia, M., Wang, M., Reinman, G., & Faloutsos, P. (2011). Improved benchmarking for steering algorithms. In *Proceedings of the 4th international conference on motion in games* (pp. 266–277).

Kapadia, M., Wang, M., Singh, S., Reinman, G., & Faloutsos, P. (2011). Scenario space: characterizing coverage, quality, and failure of steering algorithms. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 53–62).

Karamouzas, I., Heil, P., Van Beek, P., & Overmars, M. H. (2009). A predictive collision avoidance model for pedestrian simulation. In *Lecture notes in computer science - lecture notes in artificial intelligence and lecture notes in bioinformatics* (Vol. 5884 LNCS, pp. 41–52).

Karamouzas, I., & Overmars, M. (2012). Simulating and evaluating the local behavior of small pedestrian groups. *IEEE Transactions on Visualization and Computer Graphics*, *18*(3), 394–406.

Kim, C. M., & Eng, J. J. (2003). Symmetry in vertical ground reaction force is accompanied by symmetry in temporal but not distance variables of gait in persons with stroke. *Gait & posture*, *18*(1), 23–28.

Koo, J., Kim, Y. S., & Kim, B.-I. (2012). Estimating the impact of residents with disabilities on the evacuation in a high-rise building: A simulation study. *Simulation Modelling Practice and Theory*, *24*, 71–83.

Kramer, S., Johnson, L., Bernhardt, J., & Cumming, T. (2016). Energy expenditure and cost during walking after stroke: a systematic review. *Archives of physical medicine* *and rehabilitation*, *97*(4), 619–632.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, *47*(260), 583–621.

Kulpa, R., Olivier, A.-H., Ondřej, J., & Pettré, J. (2011). Imperceptible relaxation of collision avoidance constraints in virtual crowds. *ACM Transactions on Graphics*, *30*(6), 1.

Kun, A., & Miller III, W. T. (1996). Adaptive dynamic balance of a biped robot using neural networks. In *Proceedings of the IEEE international conference on robotics and automation* (Vol. pages, pp. 240–245).

Kuo, A. D. (2001). A simple model of bipedal walking predicts the preferred speed–step length relationship. *Journal of biomechanical engineering*, *123*(3), 264–269.

Kuo, A. D. (2002). The relative roles of feedforward and feedback in the control of rhythmic movements. *Motor Control*, *6*(2), 129–145.

Kwon, T., & Hodgins, J. (2010). Control systems for human running using an inverted pendulum model and a reference motion capture sequence. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 129–138).

La, C.-M. U. G. (n.d.). *Carnegie-Mellon Mocap Database.* http://mocap.cs.cmu.edu/.

Lauziere, S., Betschart, M., Aissaoui, R., & Nadeau, S. (2014). Understanding spatial and temporal gait asymmetries in individuals post stroke. *Int J Phys Med Rehabil*, *2*(3), 201.

Lee, K., Choi, M., Hong, Q., & Lee, J. (2007). Group Behavior from Video: A Data-Driven Approach to Crowd Simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 109 – 118).

Lee, Y., Kim, S., & Lee, J. (2010). Data-driven biped control. In *ACM transactions on graphics* (Vol. 29, p. 1).

Lerner, A., Chrysanthou, Y., & Lischinski, D. (2007). Crowds by example. *Computer Graphics Forum*, *26*(3), 655–664.

Lerner, A., Chrysanthou, Y., Shamir, A., & Cohen-Or, D. (2009). Data driven evaluation of crowds. In *Proceedings of the 2nd international workshop on motion in games* (pp. 75–83). Springer.

Lerner, A., Chrysanthou, Y., Shamir, A., & Cohen-Or, D. (2010). Context-dependent crowd evaluation. , *29*(7), 2197–2206.

Liu, Q. (2018). The effect of dedicated exit on the evacuation of heterogeneous pedestrians. *Physica A: Statistical Mechanics and its Applications*, *506*, 305–323.

Lockwood, N., & Singh, K. (2011). Biomechanically-inspired motion path editing. In *Proceedings of the 2011 acm siggraph/eurographics symposium on computer animation* (pp. 267–276).

Mahon, C. E., Darter, B. J., Dearth, C. L., & Hendershot, B. D. (2019). The relationship between gait symmetry and metabolic demand in individuals with unilateral transfemoral amputation: A preliminary study. *Military medicine*.

Maïm, J., Yersin, B., & Thalmann, D. (2009). Unique character instances for crowds. *IEEE Computer Graphics and Applications*, *29*(6), 82–90.

Malone, L. A., & Bastian, A. J. (2014). Spatial and temporal asymmetries in gait predict split-belt adaptation behavior in stroke. *Neurorehabilitation and neural repair*, *28*(3), 230–240.

Malone, L. A., Bastian, A. J., & Torres-Oviedo, G. (2012). How does the motor system correct for errors in time and space during locomotor adaptation? *Journal of neurophysiology*, *108*(2), 672–683.

Maslow, A. H. (1943). A theory of human motivation. *Psychological review*, *50*(4), 370.

MASSIVE. (2017). *Massive software :: Film.* Retrieved 2017-06-27, from `http://www.massivesoftware.com/film.html`

McDonnell, R., Larkin, M., Dobbyn, S., Collins, S., & O'Sullivan, C. (2008). Clone attack! Perception of crowd variety. *ACM Transactions on Graphics*, *27*(November), 1.

McDonnell, R., Larkin, M., Hernández, B., Rudomin, I., & O'Sullivan, C. (2009). Eye-catching crowds: saliency based selective variation. In *ACM transactions on graphics* (Vol. 28, p. 55).

Mehran, R., Oyama, A., & Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 935–942).

Michael, K. M., Allen, J. K., & Macko, R. F. (2005). Reduced ambulatory activity after stroke: the role of balance, gait, and cardiovascular fitness. *Archives of physical medicine and rehabilitation*, *86*(8), 1552–1556.

Miller III, W. T. (1994). Real-time neural network control of a biped walking robot. *Control Systems, IEEE*, *14*(1), 41–48.

Minetti, A., & Alexander, R. M. (1997). A theory of metabolic costs for bipedal gaits. *Journal of Theoretical Biology*, *186*(4), 467–476.

Mordatch, I., De Lasa, M., & Hertzmann, A. (2010). Robust physics-based locomotion using low-dimensional planning. *ACM Transactions on Graphics*, *29*(4), 71.

Musse, S. R., Cassol, V. J., & Jung, C. R. (2012). Towards a quantitative approach for comparing crowds. *Computer Animation and Virtual Worlds*, *23*(1), 49–57.

Musse, S. R., Jung, C. R., & Cassol, V. J. (2016). A quantitative approach to comparing real and synthetic crowds. *Simulating Heterogeneous Crowds with Interactive Behaviors*, 141.

Musse, S. R., & Thalmann, D. (1997). A model of human crowd behavior. In *Proceedings of computer animation and simulation* (pp. 39–51).

Nadeau, S., Arsenault, A. B., Gravel, D., & Bourbonnais, D. (1999). Analysis of the clinical factors determining natural and maximal gait speeds in adults with a stroke1. *American Journal of Physical Medicine & Rehabilitation*, *78*(2), 123–130.

Narang, S., Randhavane, T., Best, A., Shapiro, A., & Manocha, D. (2016). *Fbcrowd: Interactive multi-agent simulation with coupled collision avoidance and human motion synthesis* (Tech. Rep.). Technical report, Department of Computer Science, UNC Chapel Hill.

Nemenyi, P. (1963). *Distribution-free multiple comparisons* (Unpublished doctoral dissertation). Princeton University.

Oasys Limited. (2017). *Massmotion.* Retrieved 2017-06-27, from http://www.oasys-software.com/

Ondřej, J., Pettré, J., Olivier, A.-H., & Donikian, S. (2010). A synthetic-vision based steering approach for crowd simulation. *ACM Transactions on Graphics*, *29*(4), 1.

Paris, S., Pettré, J., & Donikian, S. (2007). Pedestrian reactive navigation for crowd simulation: a predictive approach. *Computer Graphics Forum*, *26*(3), 665–674.

Patterson, K. K. (2010). *Gait asymmetry post-stroke* (Unpublished doctoral dissertation). University of Toronto.

Patterson, K. K., Gage, W. H., Brooks, D., Black, S. E., & McIlroy, W. E. (2010a). Changes in gait symmetry and velocity after stroke: a cross-sectional study from weeks to years after stroke. *Neurorehabilitation and Neural Repair*, *24*(9), 783–790.

Patterson, K. K., Gage, W. H., Brooks, D., Black, S. E., & McIlroy, W. E. (2010b). Evaluation of gait symmetry after stroke: a comparison of current methods and recommendations for standardization. *Gait & posture*, *31*(2), 241–246.

Patterson, K. K., Parafianowicz, I., Danells, C. J., Closson, V., Verrier, M. C., Staines, W. R., . . . McIlroy, W. E. (2008). Gait asymmetry in community-ambulating stroke survivors. *Archives of physical medicine and rehabilitation*, *89*(2), 304–310.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2007). Controlling individual agents in high-density crowd simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 99–108).

Pelechano, N., Stocker, C., Allbeck, J., & Badler, N. (2008). Being a part of the crowd: towards validating vr crowds using presence. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems* (pp. 136–142).

Peng, X. B., Berseth, G., Yin, K., & van de Panne, M. (2017). Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*, *36*(4).

Pettré, J., Ondrej, J., Olivier, A.-h., Cretual, A., & Donikian, S. (2009). Experiment-based modeling, simulation and validation of interactions between virtual walkers. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics symposium on computer animation* (Vol. 2009, p. 189).

Platts, M. M., Rafferty, D., & Paul, L. (2006). Metabolic cost of over ground gait in younger stroke patients and healthy controls. *Medicine and science in sports and exercise*, *38*(6), 1041–1046.

PTV Group. (2019). *Ptv viswalk.* Retrieved 2019-05-10, from http://vision-traffic.ptvgroup.com/en-us/products/ptv-viswalk/

Renalds, A., Smith, T. H., & Hale, P. J. (2010). A systematic review of built environment and health. *Family & community health*, *33*(1), 68–78. doi: 10.1097/FCH.0b013e3181c4e2e5

Reynolds, C. (1995). *Boids: Background and update.* Retrieved 2017-06-27, from http://www.red3d.com/cwr/boids/

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, *21*(4), 25–34.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53–65.

Sargent, R. G. (2009). Verification and validation of simulation models. In *Proceedings of the 2009 winter simulation conference* (pp. 162–176).

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*(2), 143–186.

Severiukhina, O., Voloshin, D., Lees, M., & Karbovskii, V. (2017). The study of the influence of obstacles on crowd dynamics. *Procedia Computer Science*, *108*, 215–224.

Seyfried, A., Boltes, M., Kähler, J., Klingsch, W., Portz, A., Rupprecht, T., ... Winkens, A. (2010). Enhanced empirical data for the fundamental diagram and the flow through bottlenecks. In *Pedestrian and evacuation dynamics* (pp. 145–156). Springer Berlin Heidelberg.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379–423.

Shao, W., & Terzopoulos, D. (2005). Autonomous pedestrians. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 19–28).

Shimada, T., & Naoi, H. (2006). An experimental study on the evacuation flow of crowd including wheelchair users. *Fire Science and Technology*, *25*(1), 1–14.

Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In *Cvpr.*

SIMWALK. (2019). *Simwalk pro.* Retrieved 2019-05-10, from `http://www.simwalk.com/simwalk_pro/index.html`

Singh, S., Kapadia, M., Faloutsos, P., & Reinman, G. (2009). Steerbench: A benchmark suite for evaluating steering behaviors. *Computer Animation and Virtual Worlds*, *20*(5-6), 533–548.

Singh, S., Kapadia, M., Hewlett, B., Reinman, G., & Faloutsos, P. (2011). A modular framework for adaptive agent-based steering. In *Proceedings of the symposium on interactive 3d graphics and games* (pp. 1–9).

Singh, S., Kapadia, M., Reinman, G., & Faloutsos, P. (2011). Footstep navigation for dynamic crowds. *Computer Animation and Virtual Worlds*, *22*(2-3), 151–158.

Snape, J., Guy, S. J., Vembar, D., Lake, A., Lin, M. C., & Manocha, D. (2012). Reciprocal collision avoidance and navigation for video games. In *Proceedings of the game developers conference.*

Sok, K. W., Kim, M., & Lee, J. (2007). Simulating biped behaviors from human motion data. *ACM Transactions on Graphics*, *26*(3), 107.

Statistics Canada. (2015). *Disability in canada: Initial findings from the canadian survey on disability* (Fact Sheet for the Canadian Survey on Disability (CSD) No. 89-654-X). Statistics CanadaStatistics Canada.

Still, G. K. (2000). *Crowd dynamics* (Unpublished doctoral dissertation). University of Warwick.

Still, G. K. (2007). Review of pedestrian and evacuation simulations. *International Journal of Critical Infrastructures*, *3*(3-4), 376–388.

Stolze, H., Klebe, S., Baecker, C., Zechlin, C., Friege, L., Pohle, S., & Deuschl, G. (2005). Prevalence of gait disorders in hospitalized neurological patients. *Movement disorders: official journal of the Movement Disorder Society*, *20*(1), 89–94.

Stuart, D. S., Sharifi, M. S., Christensen, K. M., Chen, A., Kim, Y. S., & Chen, Y. (2019). Crowds involving individuals with disabilities: Modeling heterogeneity using fractional order potential fields and the social force model. *Physica A: Statistical Mechanics and its Applications*, *514*, 244–258.

Stüvel, S. A., de Goeij, M., van der Stappen, A. F., & Egges, A. (2015). An analysis of manoeuvring in dense crowds. In *Proceedings of the 8th ACM SIGGRAPH conference on motion in games* (pp. 85–90).

Sud, A., Gayle, R., Guy, S., Andersen, E., Lin, M., & Manocha, D. (2007). *Real-time simulation of heterogeneous crowds* (Tech. Rep.). Tech. rep., University of north california.

Taga, G., Yamaguchi, Y., & Shinizu, H. (1991). Self-organized control of bipedal locomotion by neural oscillators in unpredicatable environments. *Biological Cybernetics*, *65*(3), 147–159.

Thalamuthu, A., Mukhopadhyay, I., Zheng, X., & Tseng, G. C. (2006). Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics*, *22*(19), 2405–2412.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423.

TraffGo HT. (2019). *Pedgo.* Retrieved 2019-05-10, from `https://www.traffgo-ht.com/en/pedestrians/products/pedgo/index.html`

Treuille, A., Cooper, S., & Popović, Z. (2006). Continuum crowds. *ACM Transactions on Graphics*, *25*(3), 1160.

Tsai, Y. Y., Lin, W. C., Cheng, K. B., Lee, J., & Lee, T. Y. (2010). Real-time physics-based 3D biped character animation using an inverted pendulum model. *IEEE Transactions on Visualization and Computer Graphics*, *16*(2), 325–337.

Tuomisto, H. (2010). A diversity of beta diversities: straightening up a concept gone awry. part 1. defining beta diversity as a function of alpha and gamma diversity. *Ecography*, *33*(1), 2–22.

van den Berg, J., Guy, S. J., Lin, M., & Manocha, D. (2011). Reciprocal n-body collision avoidance. In *Robotics research* (Vol. 70, p. 3-19).

Van den Berg, J., Lin, M., & Manocha, D. (2008). Reciprocal velocity obstacles for real-time multi-agent navigation. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 1928–1935).

Vanumu, L. D., Rao, K. R., & Tiwari, G. (2017). Fundamental diagrams of pedestrian flow characteristics: a review. *European Transport Research Review*, *9*(4), 49.

Verghese, J., LeValley, A., Hall, C. B., Katz, M. J., Ambrose, A. F., & Lipton, R. B. (2006). Epidemiology of gait disorders in community-residing older adults. *Journal of the American Geriatrics Society*, *54*(2), 255–261.

Viswanathan, V., Lee, C. E., Lees, M. H., Cheong, S. A., & Sloot, P. M. (2014). Quantitative comparison between crowd models for evacuation planning and evaluation. *The European Physical Journal B*, *87*(2), 27.

Von Neumann, J., Burks, A. W., et al. (1966). Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, *5*(1), 3–14.

Wang, H., Ondřej, J., & O'Sullivan, C. (2017). Trending paths: A new semantic-level metric for comparing simulated and real crowd data. *IEEE Transactions on Visualization and Computer Graphics*, *23*(5), 1454–1464.

Wang, K., Wang, B., & Peng, L. (2009). Cvap: validation for cluster analyses. *Data Science Journal*, 0904220071–0904220071.

Wei, S.-E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In *Cvpr.*

Wendell, S. (1996). *The rejected body: Feminist philosophical reflections on disability.* Routledge.

Whittle, M. W. (2014). *Gait analysis: an introduction.* Butterworth-Heinemann.

Wilkie, D., van den Berg, J., & Manocha, D. (2009). Generalized velocity obstacles. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems* (pp. 5573–5578).

Wolinski, D., Guy, S. J., Olivier, A.-H., Lin, M. C., Manocha, D., & Pettré, J. (2014a). Optimization-based pedestrian model calibration for evaluation. *Transportation Research Procedia*, *2*, 228–236.

Wolinski, D., Guy, S. J., Olivier, A.-H., Lin, M. C., Manocha, D., & Pettré, J. (2014b). Parameter estimation and comparative evaluation of crowd simulations. *Computer Graphics Forum*, *33*(2), 303–312.

Wolinski, D., Lin, M. C., & Pettré, J. (2016). Warpdriver: context-aware probabilistic motion prediction for crowd simulation. *ACM Transactions on Graphics*, *35*(6), 164.

Wrotek, P., Jenkins, O. C., & McGuire, M. (2006). Dynamo. In *Proceedings of the 2006 ACM SIGGRAPH symposium on videogames* (Vol. 1, pp. 61–70).

Yin, K., Loken, K., & van de Panne, M. (2007). Simbicon: Simple biped locomotion control. *ACM Transactions on Graphics*, *26*(3), Article 105.

Zamparo, P., Francescato, M., De Luca, G., Lovati, L., & di Prampera, P. (1995). The energy cost of level walking in patients with hemiplegia. *Scandinavian journal of medicine & science in sports*, *5*(6), 348–352.

Zarrugh, M., Todd, F., & Ralston, H. (1974). Optimization of energy expenditure during level walking. *European journal of applied physiology and occupational physiology*, *33*(4), 293–306.

Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A., & Xu, L.-Q. (2008). Crowd analysis: a survey. *Machine Vision and Applications*, *19*(5-6), 345–357.

Zhao, Y., Li, M., Lu, X., Tian, L., Yu, Z., Huang, K., ... Li, T. (2017). Optimal layout design of obstacles for panic evacuation using differential evolution. *Physica A: Statistical Mechanics and its Applications*, *465*, 175–194.

Appendices

Appendix A

Effort Optimal Parameter Sets

Table A.1: *rvo2dAI* Effort optimal parameters. *rvo2dAI* has 4 parameters that are involved in the optimization process.

Parameter Name	Value
neighbor distance	12.08
time horizon	2.72
time horizon obstacles	11.81
max neighbors	15.03

Table A.2: *sfAI* Effort optimal parameters. *sfAI* has 11 parameters that are involved in the optimization process.

Parameter Name	Value
acceleration	0.05
personal space threshold	0.1
agent repulsion importance	0.11
query radius	10
body force	500
agent body force	4027.4
sliding friction force	10000
agent b	0.11
agent a	53.24
wall b	0.08
wall a	61.65

Table A.3: *footstepAI* Effort optimal parameters. *footstepAI* has 19 parameters that are involved in the optimization process.

Parameter Name	**Value**
preferred step angle	0.58
default com height	0.93
default min step length	0.16
default max step length	0.79
default min step time	0.03
default max step time	0.6
default time cost weight	0.58
default trajectory cost weght	0.03
shoulder comfort zone	0.39
shoulder comfort zone	0.03
ped reached target distance threshold	1
furthest local target distance	71.72
next waypoint distance	26.37
ped max num waypoints	26.19
ped query radius	10.3
ped num steps before forced plan	6.28
ped initial step variation	0.45
ped reached footstep goal threshold	1.05
max nodes to expand	910.07

Table A.4: *pamAI* Effort optimal parameters. *pamAI* has 12 parameters that are involved in the optimization process.

Parameter Name	**Value**
max acceleration	22.22971
fov	187.5081
ksi	0.4
neighbour distance	8.208917
max neighbours	7
time horizon	4.027235
agent distance	0.05
wall distance	0.05
d mid	3.405647
agent strength	0.9
wall steepness	1.784718
w factor	0.760493

Appendix B
Heterogeneity Results Details

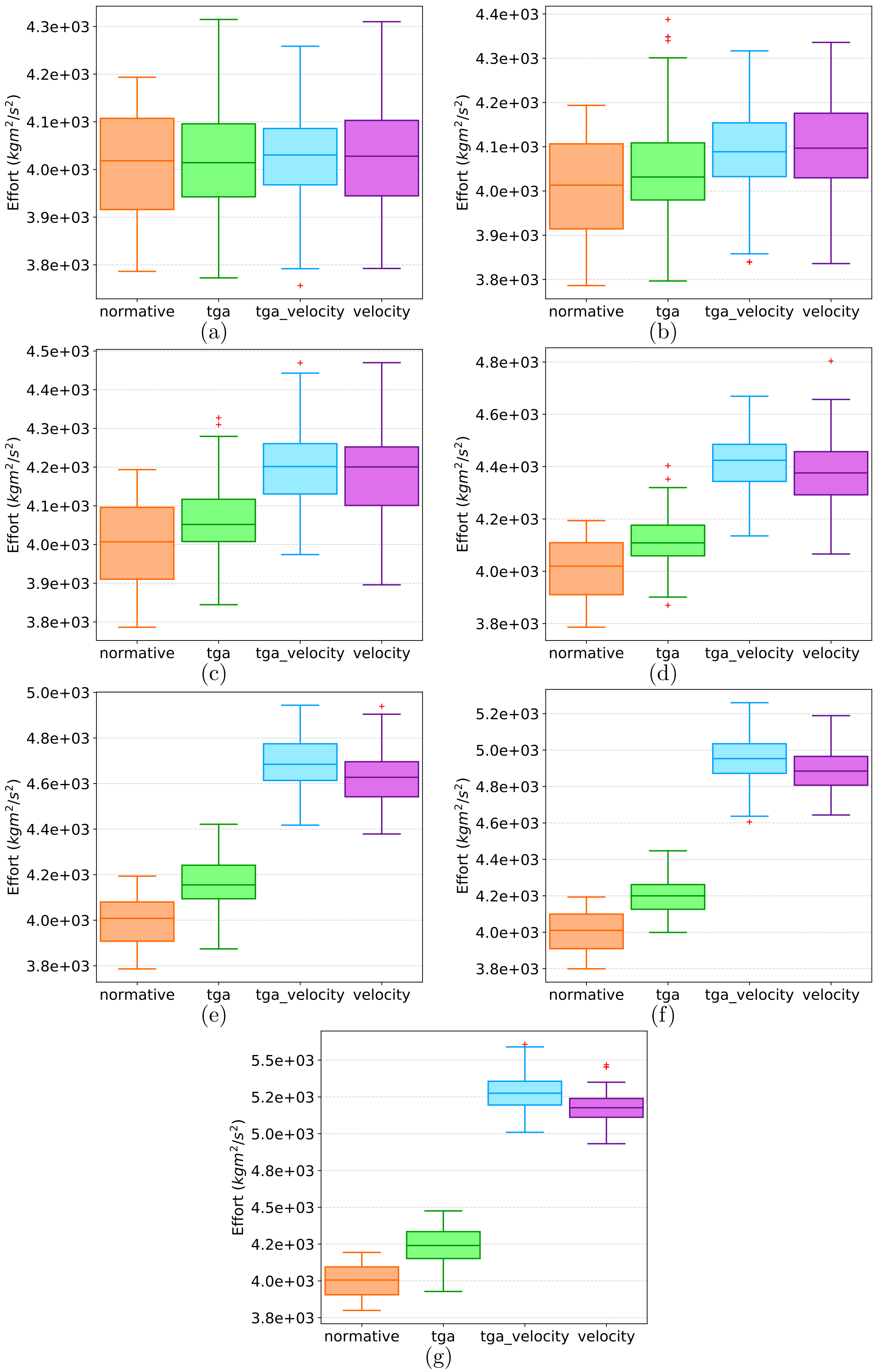


Figure B.1. Detail boxplots of effort for each crowd model mixture from the experiment in Section 5.5 of Chapter 5: (a) 1/50 (2%), (b) 5/50 (10%),(c) 10/50 (20%),(d) 20/50 (40%),(e) 30/50 (60%),(f) 40/50 (80%),(g) 50/50 (100%).

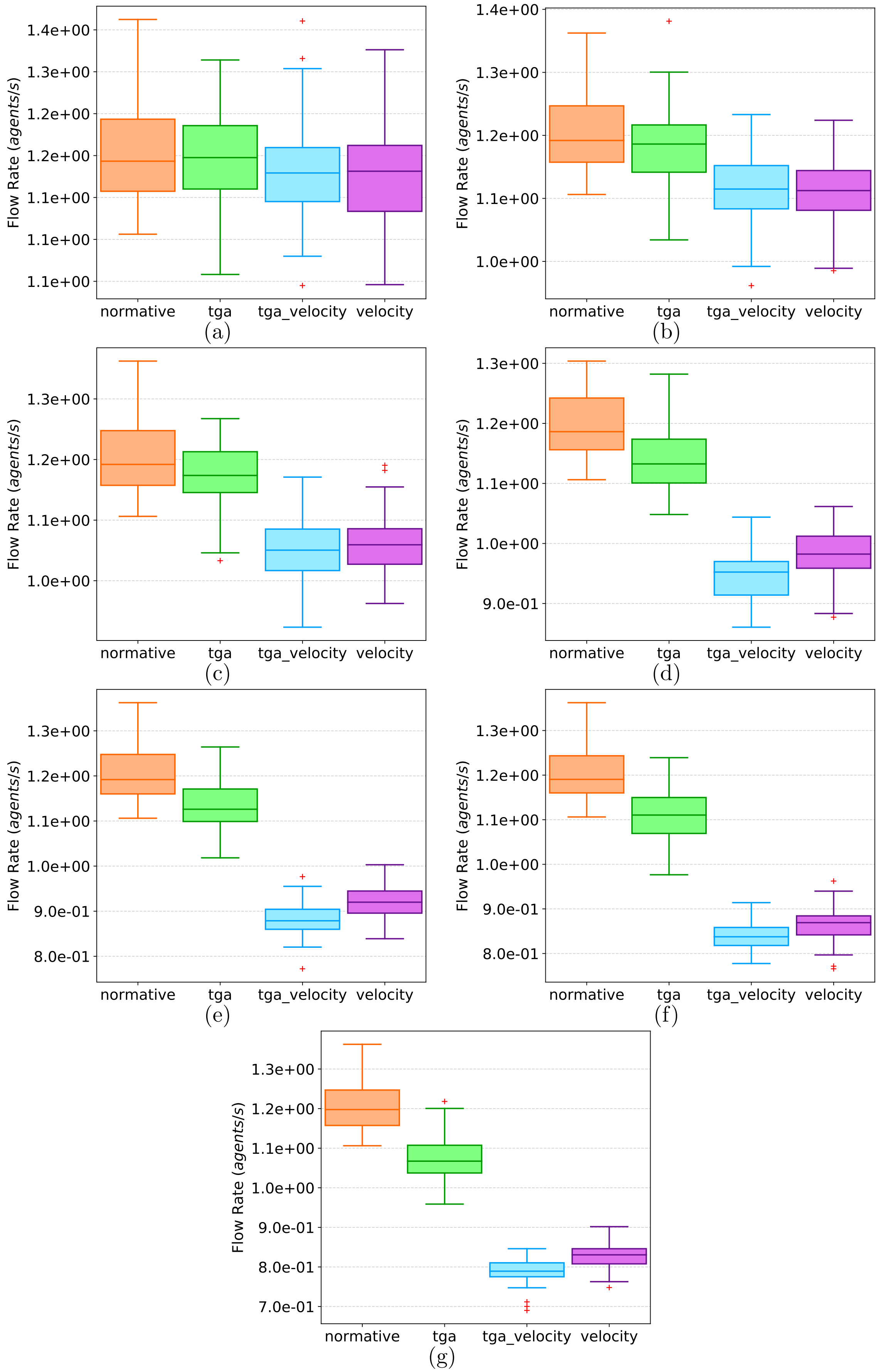


Figure B.2. Detail boxplots of flow rate for each crowd model mixture from the experiment in Section 5.5 of Chapter 5: (a) 1/50 (2%), (b) 5/50 (10%),(c) 10/50 (20%),(d) 20/50 (40%),(e) 30/50 (60%),(f) 40/50 (80%),(g) 50/50 (100%).

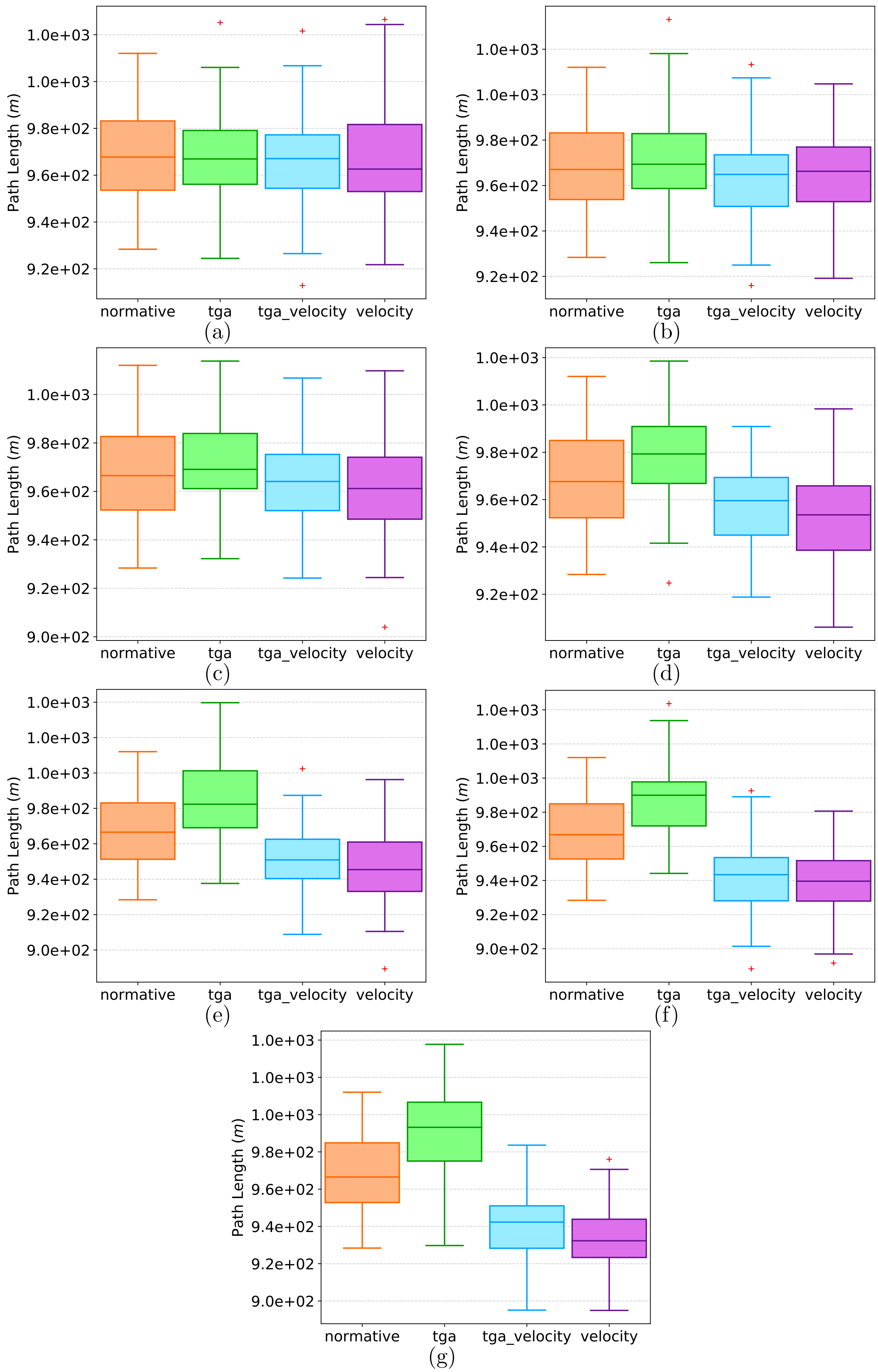


Figure B.3. Detail boxplots of path length for each crowd model mixture from the experiment in Section 5.5 of Chapter 5: (a) 1/50 (2%), (b) 5/50 (10%),(c) 10/50 (20%),(d) 20/50 (40%),(e) 30/50 (60%),(f) 40/50 (80%),(g) 50/50 (100%).