

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

PhD Thesis

Trinity College Dublin

The University of Dublin

School of Computer Science & Statistics

Investigating Motion Perception and Physics-based Methods for Body Shape Diversity in Virtual Avatars

Author:
Bharat Vyas

Supervisor:
Prof. Carol O'Sullivan

Submitted
May 20, 2025

A thesis submitted in partial fulfillment of the degree of
Doctor of Philosophy

Declaration of Work

I hereby declare that this dissertation is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

I agree to deposit this thesis in the University's open-access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Date: May 20, 2025



Signature

Abstract

Creating virtual environments populated with realistic humanoid avatars is a challenging task, especially when the aim is to represent the human diversity found in the real world. One significant factor in this diversity is body shape, which not only impacts the visual appearance of virtual avatars but also influences their physical motion characteristics. The goal of this work was to explore the perception of different body shapes in virtual avatars and investigate methods for incorporating these variations into character animations.

We first investigated whether body shape influences motion perception in virtual avatars, specifically for walking and jogging animations. We conducted a user study in which participants were shown avatars that either matched or differed from the body shape of the motion's original actor. Participants were asked to rate the consistency between the avatar's body shape and motion. The Results showed that avatars within the same BMI (Body Mass Index) group were perceived as more consistent. We also conducted a case study on the impact of self-representation (users were embodied in a full-body avatar or not) on presence in virtual reality. We found that even minimal representations, such as animated hands, were sufficient for users to feel a sense of presence in head-mounted display systems. As the case study indicated that the avatar body did not significantly impact presence in VR, we instead focussed on the findings from the first study, which clearly demonstrated the importance of body shape in maintaining motion consistency. These results motivated us to explore methods for incorporating body shape diversity into virtual environments.

We explored two distinct methods for generating body shape-aware character animations, mainly focusing on physics-based avatars. The first approach involved a deep reinforcement learning framework, where the task was to animate a rigid body skeleton and mimic the given motion file. We introduced some body shape parameters which modified the task reward for individual characters, thereby incorporating the effect of mass during training. This resulted in altered final motion and variation of motion based on the body shape. While the generated motions aligned with the anthropometric data of human movement, this pipeline was not easily com-

patible with traditional animation workflows or game engines. This prompted us to investigate some alternative methods.

The second approach was mainly based on adjusting the inertial parameters of physics-driven avatars. We devised a physics-based controller framework that utilises motion, target body weight, and height as inputs to generate retargeted motions. The approach explored whether postural control parameters, i.e., mass and inertial properties like the centre of mass location, can generate realistic motions across different BMIs. This was achieved by controlling limb balance points using a PD control system. To evaluate the quality of the retargeted motions, we then developed a perceptual framework that compared the generated motions against an *average* motion, which had been perceived as most attractive and least distinctive in previous studies.

In addition to individual animations, we explored potential applications of these physics-based avatars in small-size crowd simulations. Specifically, we examined whether body shape affects the detection of motion clones, which are common in crowd scenarios. Contrary to our hypothesis, body shape had no significant effect on motion clone detection. Interestingly, the inclusion of distinct motions caused motion clones to be more easily detectable.

In conclusion, our work explored the role of body shape in the perception and generation of virtual avatar motion. We primarily explored two methodologies for generating body shape-aware animations and developed a perceptual framework for evaluation. While both methods have certain limitations, the significance of body shape in enhancing diversity among virtual avatars cannot be overlooked. Our findings contribute to advancing the representation of diversity in virtual environments and provide practical insights for future avatar animation techniques. Further exploration into factors such as gender, race, or age and their influence on motion perception could deepen our understanding of how best to retarget motion across diverse body types. Our work offers valuable foundations for generating more natural and realistic human motions in virtual settings.

Acknowledgements

It's hard to believe that it's been four years already, but it surely feels like the end of a long journey. So much has happened in these years, including some small contributions compiled in this thesis. Of course, no journey is interesting without the amazing people who join you along the way. First of all, I would like to thank my supervisor, Carol. Thank you for showing trust and confidence in my ideas throughout these four years, and introducing me to the world of research. I also want to extend my thanks to my mentors at Inria, Ludovic and Julien, for their invaluable guidance and insights during my secondment. A big thanks to Betty for being an incredible manager during my time at Amazon, and to my mentors, Mohammad and Christopher, for their trust and belief they placed in me during our research work work together. I would like to thank my examiners Rachel and Michael for their constructive feedback.

I also want to thank all the people who made this journey an unforgettable experience. Thanks to all the members of T-motion (Goksu, Jackey, Seth, Xiaohan); it has always been fun discussing research in our group meetings. A special thanks to Donal and Theo for being such a big support over the last few years. Also, thanks to other members of StackB (Sam, Priyansh, Yuan, Mauricio, and Darragh). Thanks to Gareth for being a big support and some fun EP memories. And thanks to Stephen for all the help and for organising those writing sessions. To the CLIPE family (Lisa, Eduardo, Marilena, Nefeli, Tairan, Klara, Ariel, Radek, Kiran, Haoran) I will always cherish the amazing workshops and fun (stressful) presentations we had. I am also grateful to the wonderful people I met in Inria and Rennes. Especially Thomas, Louise, and Elodie, who patiently spent their time with me when I couldn't speak a single sentence in French. Also, to my fellow interns at Amazon (Laia, Itzik, William), thank you for making my internship such an enjoyable experience.

I feel blessed to have the constant support of my friends and family. To my friends back home (Sky, Madaan, Kaushal, Prakhar, Tarun, Vibhu, Nirmit, Lakshay, and Prateek), thank you for being there for me, even from a distance. I also want to thank Zuilmah and her family for their emotional support during the tough time of lockdown days. Thanks to Lujain for being an incredible friend and flatmate. A

special thanks to Mary for constant support and always being there for me. And finally, a big thanks to - Maa, Paa and Divya. Your constant love and support mean the world to me, and I couldn't have done this without you. Also grateful to Bade papa, Jiji mummy, and all my extended family for their endless support.

Apologies if I have missed anyone's name, but I genuinely appreciate all the support and love I have received over these past years. Thank you, everyone!

Contents

1 Introduction **17**
1.1 Context . . . 19
1.2 Our Approach . . . 22
1.3 Contributions . . . 23
1.4 Overview . . . 26
1.5 Relevant Publications . . . 27

2 Background Literature **28**
2.1 Human Locomotion . . . 28
2.1.1 Gait & Centre of Mass . . . 29
2.1.2 Analysis of Locomotion . . . 31
2.2 Motion Synthesis & Animation . . . 35
2.3 Body Shape of Virtual Avatars . . . 37
2.4 Physics-based Character Animation . . . 41
2.5 Applied Perception of Motion . . . 46
2.6 Chapter Conclusion . . . 49

3 Body Shape & Motion **50**
3.1 Introduction . . . 50
3.2 Body & Motion Consistency . . . 51
3.3 Experiment . . . 52
3.3.1 Stimuli . . . 54
3.4 Perceptual Experiment . . . 55
3.4.1 Participants . . . 55
3.4.2 Procedure . . . 55
3.5 Results . . . 57
3.6 Discussion . . . 60
3.7 Chapter Conclusion . . . 62

4 Case Study I: Effect of Self Representation on Virtual Presence 64
4.1 Introduction . . . 64
4.2 Presence in VR . . . 65
4.3 Methods . . . 69
4.3.1 Equipment . . . 69
4.3.2 Avatar Conditions Stimuli . . . 70
4.3.3 Full Body Tracking in VR . . . 72
4.4 Experiment . . . 73
4.4.1 Tasks in VR . . . 73
4.4.2 Participants . . . 75
4.5 Procedure . . . 75
4.6 Results . . . 77
4.6.1 Body Ownership . . . 78
4.6.2 Agency . . . 79
4.6.3 Self-Location . . . 80
4.7 Discussion . . . 80

5 Body Parameterised Motion Controller 85
5.1 DRL & Physics based characters . . . 87
5.2 Reinforcement Learning Framework . . . 88
5.2.1 Network Architecture and Training Setup . . . 94
5.3 Results . . . 96
5.4 Discussion . . . 98

6 Balance Point Controller 100
6.1 Physics-based Character . . . 101
6.2 Balance Point Control . . . 104
6.3 Testing and Validation . . . 105
6.4 Experiment 1: Evaluation with Quasi-average Motion . . . 108
6.4.1 Method . . . 109
6.4.2 Results . . . 110
6.5 Experiment 2: Evaluation with Overall Average Motion . . . 112
6.5.1 Method . . . 113
6.5.2 Results . . . 116
6.6 Discussion . . . 117

7 Case Study II: How Body Shapes Change the Perception of Small-Scale Crowd Motions 121
7.1 Body Shape and Crowd . . . 122

7.2 Experiment . . . 123
7.2.1 Physics-Based Model . . . 124
7.2.2 Stimuli . . . 125
7.3 Participants . . . 127
7.4 Procedure . . . 127
7.5 Results . . . 129
7.6 Discussion . . . 131

8 Conclusion **134**
8.1 Avatar Body & Motion . . . 134
8.2 Motion Controller . . . 135
8.3 Perception of Simulated Motion . . . 136
8.4 Limitations . . . 137
8.5 Future Work . . . 139
8.6 Final Remarks . . . 140

Bibliography **163**

List of Figures

1.1 Some examples of applications using virtual humans developed in the context of the CLIPE project. The example (a) shows a virtual avatar partially controlled by a physics-based controller [ARC22]). Example (b) shows an approach where a single user in virtual scenarios successively acts as each crowd member for crowd data capture [YHC+22]. Example (c) shows a crowd simulation inside a museum scenario to analyse the crowd behaviour [PKL+22]. 18

1.2 In our approach: We combine motion perception, body shape and a physics-based framework for virtual avatars 23

2.1 Comparison of walking and running gait cycles showing key phases: double support, single support, and swing phases during walking (top), and single support, float, and swing phases during running (bottom).[1] 30

2.2 Vertical displacement of the center of mass (COM) during a complete gait cycle. The graph shows a sinusoidal pattern of vertical COM motion relative to gait percentage, with key walking postures illustrated above. Adapted from Lobet et al. [LDMH13]. 31

2.3 Visualisation of the three anatomical planes: the sagittal plane (red), coronal plane (green), and transverse plane (purple). Significant movements occur in all three planes during a gait cycle. 32

2.4 The graph illustrates knee flexion during the gait cycle. The joint angle (in degrees) is plotted against the percentage of the gait cycle. The image on the right shows the range of knee flexion from 0° (full extension) to 140° (maximum flexion). The graph is adapted from Whittle's introductory book on Gait Analsyis [Whi14]. 33

2.5 Plot illustrating the knee adduction moment (Nm/Kg) during the stance phase of the gait cycle. The moment is plotted against the percentage of the stance phase, emphasizing variations throughout key events such as heel strike, midstance, and toe-off. The diagram on the right shows the knee joint, lever arm, and the ground reaction force (GRF) vector directed towards the center of mass (COM). The plot is adapted from Foroughi [FSV09]. 34

2.6 Example of a motion and motion interactions generated with a deep learning based model, from Starke et al. [SZKZ20]. 36

2.7 The body shape model introduced by Anguelov et al. [ASK+05]. The model can be used for shape completion - generating a complete surface mesh given a limited set of markers. 39

2.8 The SMPL model presented by [LMR+15a], is a realistic learned model of human body shape. The model is based on a mesh composed of 6,890 vertices and 23 joints. 40

2.9 Controller-based method. A physics-based skeleton performing fall, rolling and getting up motions. Faloutsos et al. [FVdPT01] created composable controllers by defining a set of preconditions to identify the current state and evaluate future possible states. 43

2.10 Deep Reinforcement Learning. Peng et al. [PALvdP18] proposed a method to mimic the animation from motion capture data. In future works, as in [PMA+21], they incremented such techniques to obtain added behaviour to control simulated motions. 45

2.11 Distinctiveness & Attractiveness of human motions were explored by Hoyet et al. [HRZ+13]. They found that distinctiveness in one motion type does not transfer to another. 48

3.1 Male and female avatars used as stimuli, ordered from higher to lower BMI. Avatars represent a range of body types for both genders. . . . 53

3.2 Experiment Stimulus Generation Process: shape (β) and pose (θ) parameters are extracted from marker data using MoSh++ (green: consistent motion; red: inconsistent motion). 54

3.3 Study 1 example stimulus: Body shape and motion were varied across trials to investigate the effect of inconsistencies. The full set of stimulus videos can be accessed via the QR code or at: https://youtu.be/YhKqrFWRZTY. 56

3.4 Experiment Results: Mean rating scores for varying BMIs. Higher scores indicate more favourable ratings. Consistent motions, i.e., where the original body shape and motion match, are marked with a star. 59

4.1 Illustration of the VR setup and motion capture system. The participant is equipped with an HMD with controllers and motion sensors that capture body movements. The motion capture data is updated based on the HMD position and used to generate an avatar representation with IK for hand movements. 70

4.2 Visualisation of different avatar conditions and control types. Participants interacted with either controllers or finger-tracking, depending on the condition. 72

4.3 Virtual Scene with a workstation: (1) Large box; (2) Small box; (3) High shelf; (4) Open-top container; (5) Low shelf; (6) Starting or Origin Position; (7) Monitor screen 74

4.4 Mean scores for each condition with 95% confidence intervals. A p-value $< .05$ was considered statistically significant. 78

4.5 Mean scores for individual categories measuring overall virtual presence across different avatar self-representation conditions, with 95% confidence intervals. 79

4.6 Ratings for individual Body Ownership (BO) items. Error bars represent standard deviation. Asterisks (*) denote statistically significant main effects. 80

4.7 Ratings for individual Agency (AG) items. Error bars represent standard deviation. Asterisks (*) denote statistically significant main effects. 81

4.8 Ratings for individual Self-Location (SL) items. Error bars represent standard deviation. No significant main effects were observed. 82

5.1 The framework for each episode of training: (1) The character is generated using the SMPL body approximation. (2) The parameter θ is updated to calculate the total rewards. (3) Based on the rewards, an action is selected for the next possible state. (4) The state is updated based on the next action, completing the loop. 90

5.2 The base character model obtained from approximation of SMPL body (both $\beta(0)$ and $\beta(1)$ equals to zero). 92

5.3 Biomechanical metrics for the average, small and large bodies, for one complete gait cycle: normalised root COM displacement (left); and normalised values of knee flexion angle (right). 96

5.4 Human body representations with motion variations generated by the DRL framework, adapted to different body shape parameters. The full set of motion results can be accessed via the QR code or at: https://youtu.be/CnrKNBhSEuc. 98

6.1 Ragdoll model with colliders, rigid bodies, and configurable joints in body segments. The red points depict the default COM position for the respective rigid body. 101

6.2 Effect of Balance Control factor on the posture. The green character represents the kinematic skeleton acting as the target posture, while the red character shows the ragdoll skeleton attempting to mimic the target. The posture of the ragdoll is influenced by the balance control factor. 106

6.3 Average (A) adduction/abduction, (B) flexion/extension, (C) internal/external rotation angles, (D) adduction/abduction, (E) flexion/extension, (F) internal/external rotation moment waveforms for both participant groups at the matched speed condition. Adapted from [MCM16] . . . 107

6.4 Comparison of biomechanical parameters across characters with different BMIs for one complete gait cycle – i) Root COM Position: normalised position for the level of displacement; ii) Knee Moment: normalised values of the Knee extension moment; iii) Knee Flexion: Normalised values of knee flexion angle. 108

6.5 Experiment 1 example stimulus: Quasi-average and controller motion were presented on the same body shape and participants selected the one they found more natural. The full set of stimulus videos can be accessed via the QR code or at: https://youtu.be/9JexbzPHh3o. . . 109

6.6 Experiment 1 Results: Percentage of time each motion was chosen as more natural. 111

6.7 Experiment 1 Results: Each column heading shows, for the given motion, the percentage of times it was chosen for each of the row headings. 112

6.8 Experiment 2 Stimuli Generation Process: The process involves using average motion, where the *No COM* approach applies average motion directly to the meshed body. In the *COM* approach, a ragdoll skeleton mimics the average motion through retargeting, facilitated by the Balance Point Controller. 113
6.9 Experiment 2 example stimulus: Average and controller motion were presented on the same body shape and participants selected the one they found more natural. The full set of stimulus videos can be accessed via the QR code or at: https://youtu.be/YhKqrFWRZTY. 114
6.10 Experiment 2 Results: Percentage of time the No COM and COM motions were chosen. 116
6.11 Factors mentioned in the comments and their frequencies. 119

7.1 Ragdoll representation of an avatar with rigid bodies. This diagram shows the structure of a ragdoll created using primitive-shaped rigid bodies and configurable joints. 124
7.2 Stimuli Creation for the Experiment: **Motion Clone** (left): Featuring characters with two distinct motions. One motion is circled in green, and the other motion is in red (out-of-step motions). Each character has one of six different body shapes. **Baseline** (right): All characters have unique body shapes and are animated with unique motions. 126
7.3 Example Stimulus for Experiment: Participants were asked to indicate their confidence level regarding their decision on which side contains avatars with different motions. The full set of stimulus videos used in the experiment can be accessed via the QR code or at: https://youtu.be/dhj2ZqC7cTY 128
7.4 Mean Ratings from Participants: The boxplots represent the participants' mean ratings (with standard deviations) for their ability to detect stimuli with different motions under varied conditions of motion (1, 2, 3, and 6 motions) and body shapes (1, 3, 6, and 12 shapes), separated by sex (M = Male, F = Female). Higher ratings indicate lower motion clone detection accuracy, while lower ratings correspond to better performance in identifying distinct motions. 130
7.5 Distribution of key factors considered by participants when distinguishing between crowd motions in the experiment. 133

List of Tables

3.1 Attributes of female and male actors used to create stimuli. 53
3.2 Experiment ANOVA: Significant effects for Gait: Walk (W), Jog (J); Body (H1b, H2b, L1b, L2b); Motion (H1m, H2m, L1m, L2m) and Sex (M, F), with effect size (partial η^2), non-centrality (nc) and observed power ($\alpha = 0.05$). 57

4.1 Demographic attributes of study participants including mean age range, body height, weight, and standard deviations. 75
4.2 Embodiment Questionnaire used in the study, featuring questions under the categories of Agency, Body Ownership, and Self-Location. These questions assess participants' subjective experience of embodiment with the virtual avatar across these key dimensions. 76

5.1 Attributes of actors used to create body models (letters refer to Figure 5.4) 96

6.1 Distribution of mass across body segments and the COM displacement along the longitudinal axis. The data are derived from the anthropometric studies conducted by McConville et al. and Young et al. [MCC+80, YCS+83]. 104
6.2 Experiment 1 ANOVA: Significant effects with effect size (partial η^2), non-centrality (nc) and observed power ($\alpha = 0.05$). 112
6.3 Experiment 2 ANOVA: significant effects for Method: NoCOM (NC), COM (C); Gait: Walk (W), Jog (J); and Sex (M, F), with effect size (partial η^2), non-centrality (nc) and observed power ($\alpha = 0.05$). . . . 117

7.1 ANOVA results for the experiment showing interaction effects of MOTION, BODY, and SEX on motion clone detection. The table presents degrees of freedom (dof), mean squares (MS), F-test values, effect sizes (η^2), and p-values for one-way, two-way, and three-way interaction effects. 131

Nomenclature

Abbreviations

BMI Body Mass Index

COM Center of Mass

DoF Degree of Freedom

GRF Ground Reaction Force

HMD Head Mounted Display

IK Inverse Kinematics

IMU Inertial Measurement Unit

KAM Knee Adduction Moment

PD Proportional-Derivative

VR Virtual Reality

Chapter 1

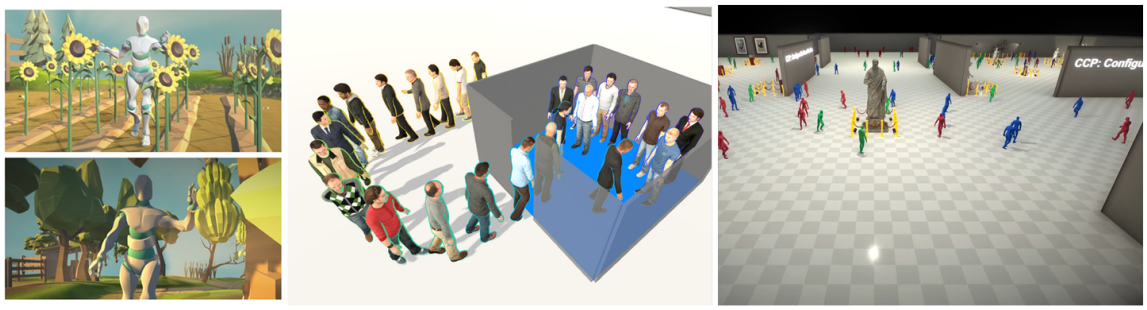
Introduction

The digital innovation era is transforming how we live, work, and interact with our surroundings. This is clear when it comes to human-looking virtual avatars - reflections of ourselves in the digital world that are now an integral part of our everyday lives. We see them in video games [SKNG23], virtual assistants [KBH+18], and virtual reality experiences [LBS13]. Beyond entertainment, they are also well-used in diverse research fields, e.g., in sports [MFC20], neuroscience [MTM+18], art exhibitions, museums [SKGD20], and other innovative applications. With this increased demand for avatars, bridging the gap between how they look and their plausibility in the real world is more important than ever. Because what happens when these versions of ourselves do not align with the reality we see in our real lives?

Since the technology is developing rapidly, end users are demanding more realistic and immersive interactions with virtual environments. However, achieving realism in virtual humanoid avatars remains a true challenge. Creating a visually convincing avatar that behaves and moves naturally depends on several factors, like their appearance, motion, and behaviour. At the same time, the interactions with their surroundings must also feel natural and believable. For this reason, avatars that follow the natural laws of physics are important. If these laws control the movement of avatars, their motion will look more realistic and match the real world.

The research work described in this thesis was developed as part of the Marie Skłodowska-Curie Actions research project CLIPE [1]. This project aimed to populate the environment with lively virtual avatars for varied scenarios; some of the examples are shown in Figure 1.1. However, the primary focus of this thesis was to investigate the physiological features of virtual avatars that affect the perception of their motion. It also presents some methods to generate plausible and natural-looking motions for these characters, especially the motions we see in our daily routines, like walking.

[1] https://www.clipe-itn.eu/



a b c

Figure 1.1: Some examples of applications using virtual humans developed in the context of the CLIPE project. The example (a) shows a virtual avatar partially controlled by a physics-based controller [ARC22]). Example (b) shows an approach where a single user in virtual scenarios successively acts as each crowd member for crowd data capture [YHC+22]. Example (c) shows a crowd simulation inside a museum scenario to analyse the crowd behaviour [PKL+22].

Every day, we encounter people with a wide range of personalities, body types, and other physiological characteristics which affect how they move and perform actions. There are also various physiological aspects that affect the human movements pattern, such as height, bone structure, fat distribution, and body weight [WS12, SGR+03]. Unfortunately, most of the virtual avatars created today are unable to adequately represent this diversity. Including such elements in the creation of virtual avatars is difficult but necessary to create a more realistic and relatable appearance. This becomes even more crucial as they have applications in user-facing interfaces across numerous industries. Thus, it is critical to find some mechanism to utilise these physiological aspects which will control the movements of these avatars. Traditionally, there are two common approaches to introducing this diversity into virtual environments. One involves capturing motion from a wide range of human subjects, which can then be used directly in animation pipelines. The other approach is a more labour-intensive process of keyframed hand animation, which requires expertise in both animation and human anatomy.

Capturing more comprehensive, high-quality motion data is one way to achieve diversity in virtual avatars. However, this approach has its own challenges. Especially with vast amounts of motion data already available, it is important to explore methods to modify this data to reflect the variety of human movements observed in real life. To better portray motion variation in virtual avatars, we investigated ways to use and modify currently available data. By doing this, we aimed to close the gap between the rich diversity of the actual world and the restricted portrayal of motion in digital contexts. However, motion perception was another challenge. For

instance, even if we successfully create a variety of avatars that accurately emulates the physiological traits of actual humans, it is essential that observers perceive these motions as natural. Since humans are highly sensitive to the subtleties of movement, even minor differences in an avatar's movements can cause the illusion of realism to be broken. Thus, it is not enough to just provide a variety of motions; we also need to ensure that people perceive these motions to be realistic and plausible. In that case, perceptual validation is crucial since it ensures that our methods produce avatars that move accurately and seem natural to the users.

In the next chapters, we will look at the complexities of creating lifelike avatars, including the technological and perceptual challenges of allowing virtual humans to move as naturally as their real-world counterparts. This thesis aims to contribute to the growing field of virtual human development by investigating physiological diversity and motion generation, providing insights that can improve the realism of digital avatars and how we interact with them.

1.1 Context

Human motion is a complex process that involves different body systems working together, such as the musculoskeletal, nervous, and cardiovascular systems [LC12]. Even though walking or running might seem simple, they are actually adaptive actions that depend on constant feedback from our senses and motor control [Cla15]. At a basic level, locomotion, like walking, is the repeated and alternating movement of the arms and legs. This helps move the body forward while keeping balance.

However, walking is not just a repetitive action. It indeed shows the human body's amazing ability to adapt to different environments and conditions. Walking reflects a fine balance of stability, flexibility, and energy efficiency, which is a key feature of human evolution [HS07]. Given the complexity of such a routine action, simulating this motion in a virtual avatar is a big challenge. Multiple factors influence human motion, all of which need to be carefully considered for accurate simulation. For instance, Ayyappa [Ayy97] states that the shape, position and function of musculoskeletal structure can influence the gait characteristics. Various parameters of gait, like the centre of mass position, knee flexion, and knee moment, constantly change to adapt to different terrains or crowded spaces, demonstrating the dynamic nature of human locomotion.

Role of Body Shape in Human Motion

Human body shape directly correlates to height, weight, and limb proportions. These parameters further influence how mass is distributed throughout the body, which shapes the mechanics of movement. Taller people, for example, may take longer strides because their limbs are longer, whereas people with more body mass (whether muscle or fat) may have different gait mechanics, such as shortened swing time and increased swing velocity [BPCS+11]. This variability demonstrates how physical characteristics are intricately linked to movement patterns, resulting in diversity in human motion.

The field of biomechanics, which studies forces acting on the body and their effects on movement, has done a lot of research on the connection between body shape and motion. The body's centre of mass, joint angles, and muscle strength are all important factors in determining how efficiently a person moves. For example, differences in body weight and height partly affect muscle strength and functional mobility [SMC+00]. Hence, we were interested in investigating if inertial parameters can be used to simulate these postural changes in virtual avatars. This was another reason we included simulations from physics-based avatars.

We argue that this diversity of body shape is an important consideration when replicating human movement in virtual environments. The challenge of developing virtual avatars that accurately reflect real-world variations extends beyond simply changing their visual features. These avatars' movement mechanics must also reflect their physical characteristics, which complicates the task of creating lifelike avatars.

Motion Perception and Its Social Significance

In addition to the mechanics of locomotion, how humans perceive motion is important, particularly when modelling virtual scenarios. Humans are naturally adept at interpreting others' movements and applying these cues to predict intentions, emotions, and actions [MST19, VPCN17]. People can also learn to associate the apparent kinematic pattern of action with the latent forces that would have been required to produce the observed motion [RF83]. Therefore, it is crucial to understand the key factors in motion perception.

Previous research has demonstrated that humans are extremely sensitive to the nuances of biological motion. Johansson [Joh73] demonstrated that people could perceive human movement from only a few point-light sources placed on a person's joints. The ability to recognise movement patterns with few visual cues demonstrates the sophistication of the human perceptual system. Our sensitivity to motion is so strong that even minor inaccuracies in the movement of virtual avatars can ruin the

illusion of immersion, making them appear less real.

There is ample evidence to suggest that motion frequently conveys social information. For example, in 1872, Darwin [Dar72] identified these cues in his research on emotional expression through physical movement, which shows that such cues are deeply rooted in our evolutionary history of social interaction. Therefore, it is important to perceptually validate that replicated motion in virtual environments appears natural and is consistent with real-world counterparts.

Research Questions

In this thesis, we address the following key research questions:

- **Q1:** How do physiological factors, such as body shape, influence the motion perception of virtual avatars?
- **Q2:** Can physics-based simulations be used to generate variations in motion based on differences in body shape?
- **Q2:** How are the resulting variations in human motion perceived by users in virtual environments?

These questions served as the motivation to explore the intersection between body shape, movement mechanics, and motion perception, particularly in the context of virtual avatars.

Body Shape, Perception, and Virtual Avatars

Question **Q1** is particularly interesting as it lies at the intersection of avatar body shape and motion perception. Observers often form opinions about a person's attitude, lifestyle, and even emotional state based on how they walk or move. These perceptions are based not only on the observer's previous experiences but also on cultural context, which forms expectations about how different body types should move. However, it is unclear whether the same perceptual theory applies when they look at virtual avatars.

Avatars often need to reflect both the physical and behavioural aspects of users. However, it can negatively impact user experience over time if this representation is inaccurate. For example, in a video game or virtual reality experience, avatars that move unnaturally or do not reflect the physical characteristics of the real world can disrupt the user's sense of presence and immersion. Addressing this challenge requires more than just making visual changes to an avatar's appearance. It is also important to investigate the specific features that people focus on when assessing motion variations in relation to an avatar's body shape.

Challenges in Virtual Representation

As previously stated, replicating the diversity of human body shapes and movement patterns in virtual environments presents a significant challenge. Even if we only consider common types of locomotion such as walking and running, it is extremely difficult to handle all of the gait parameters and muscle activations required to simulate avatar motion. In another solution, the motion of various humans could be captured, but motion capture is a time-consuming process that is not accessible to all. Furthermore, a large amount of previously captured data is already freely available. It is, therefore, more practical to investigate novel methodologies to change or adapt these existing motions.

In this thesis, we explored several methods to incorporate the body shape diversity into physics-based avatar simulation. By focusing on these physics driven avatars, we aimed to ensure that motion is physically accurate and some inertial parameters can be modified based on the principles of biomechanics. We also investigated whether these approaches enhance realism and capture the nuances of human motion diversity. This allowed us to assess the way users perceive and interact with these avatars in virtual environments.

1.2 Our Approach

Our approach addressed the challenge of creating and modifying human locomotion for virtual avatars by combining motion analysis, motion synthesis, and perceptual validation. Initially, we investigated whether the body shape of a virtual avatar influences the perception of motion. This was achieved by investigating the consistency of avatar locomotion animations based on their key anthropometric variables such as height, weight, and BMI. In addition, we explored the role of the avatar body on embodiment and virtual presence in immersive VR systems. The insights from these studies guided our future directions and helped us to explore different approaches.

We explored two approaches for creating controllers and modifying the motion of physics-based avatars. We first investigated an unsupervised learning approach for animating avatars based on their body parameters. Then, we developed another physics-based controller that can accurately adapt walking animations by adjusting some inertial parameters. These methods ensured that avatars of varying body shapes moved in ways that reflected their unique physical characteristics. We compared the motion results from these simulations to real-world biomechanical data to ensure that the variations in gait and movement were correct.

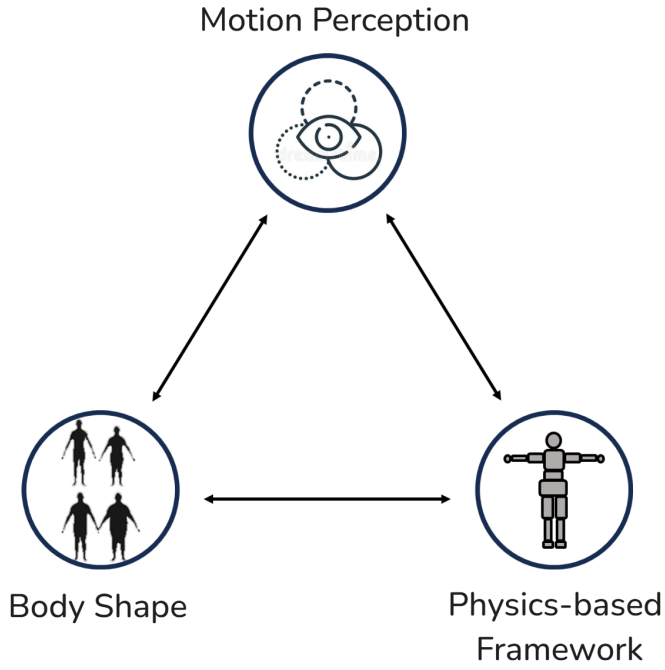


Figure 1.2: In our approach: We combine motion perception, body shape and a physics-based framework for virtual avatars

To ensure the perceptual realism of these animations, we employed user studies to evaluate the naturalness of avatar movements. In these studies, we aimed to understand how variables such as body shape influence the perceived believability of avatar motion.

1.3 Contributions

This thesis focuses on the intersection between body shape, motion consistency, and the perception of realistic motion. It presents four main contributions to the field of virtual avatars and character animation. These contributions collectively help us understand motion perception in virtual avatars and the techniques required to create them with more natural and convincing movements. The key contributions are as follows:

Contribution 1: *Study on Body Shape and Motion Consistency*

Chapter 3 discusses the first major contribution of this work. It is a study that investigated the relationship between body shape, specifically body mass index (BMI), and how people perceived motion in virtual avatars performing walking and jogging motions. In this study, we aimed to explore how consistency between motion and body shape affected user ratings.

We created two sets of stimuli: consistent stimuli in which the body shape matched the motion and inconsistent stimuli in which the body shape did not match the motion. The body shapes were categorised based on their BMI. Participants were then asked to rate consistency, indicating whether they thought the motion was consistent with the avatars' body shapes, or not. The findings revealed that users generally gave higher consistency ratings to consistent motion among avatars within the same BMI category. This means that participants could intuitively match motion to body shape, thereby reinforcing the notion that body morphology influences how motion is perceived and whether it appears plausible.

Contribution 2: *Deep Reinforcement Learning (DRL)-Based Framework with Body Shape Parameters*

In the second contribution of this work, in Chapter 5, we leveraged a deep reinforcement learning technique to generate physics-based characters with diverse body shapes. The objective was to create a population of characters capable of replicating a reference motion, such as walking or jogging, while accounting for individual variations in body shape parameters and mass.

During the training process, the characters were trained to imitate the reference motion based on their specific body shape parameters. The reference motion served as a target template, providing a desired motion outcome for the generated characters. To introduce variation in the generated motions, the influence of mass was taken into account by considering variation in shape parameters. When a character's shape parameters deviated significantly from those of the reference motion actor, the effect of mass was amplified, resulting in more noticeable differences in movement. This approach enabled the generation of animations that capture the intricate details and physical attributes of human motion, leading to realistic and diverse character performances.

Contribution 3: *A Simple Physics-Based Controller for Avatar Motion*
The third major contribution of this thesis is explored in Chapter 6. We created a simple physics-based controller that generates avatar motions based on physical parameters such as the centre of mass, body weight, and height (essentially BMI). The controller adjusted these parameters to produce individualised movements that correspond to the avatar's body size, causing larger, heavier avatars to move differently than smaller, lighter ones.

The physics-based controller used fundamental physics principles to control the avatar's motions. It modified the COM position of individual limbs depending on the avatar's weight and height. This, in turn, changed the COM of the overall skeletal body. The biomechanically plausible controller resulted in natural-looking animations. This contribution is especially useful in situations requiring large-scale simulations or real-time applications, such as crowd simulations, video games, and virtual training environments. Such a controller design allows for computational efficiency while still providing realistic and diverse motions for avatars.

Contribution 4: *Perceptual Study Mechanism for Evaluating Simulated Virtual Avatars*
The fourth and final contribution is discussed in Chapter 6. It explores the creation of a perceptual study mechanism that can be used to assess the simulated virtual avatars using user feedback. This mechanism offers a structured approach to determining how users perceive the realism and believability of simulated avatar motion in comparison to the baseline motions.

This perceptual study methodology is intended to be adaptable and applicable to a wide range of scenarios, allowing researchers and developers to evaluate the effectiveness of various animation techniques. By collecting perceptual data, such as how well users can detect inconsistencies between an avatar's body shape and motion, we can gain a better understanding of how to improve the realism of virtual characters. This contribution not only provides a tool for validating the thesis's findings but also outlines a method for future character animation research.

Furthermore, we used this mechanism to evaluate the controller mentioned in Chapter 6. We also showed that this perceptual study mechanism is scalable, which means it can be used for both small and large-scale avatar crowd evaluation, as discussed in Chapter 7. This makes it an important tool for future work in virtual avatar animation, ensuring that new techniques are thoroughly validated from a user standpoint before being used in real-world applications.

1.4 Overview

This thesis is structured as follows:

- **Chapter 2** discusses the background literature and the state of the art in areas relevant to this research. It discusses human locomotion and its underlying parameters, applied perception in virtual avatars, body shape models for virtual characters, motion editing techniques, and physics-based character animation. This chapter provides the theoretical foundation and contextual background required to understand the research contributions presented in later chapters.

- **Chapter 3** presents the study on the perception of body shape and motion consistency. It discusses the development of stimuli, the experimental methodology, and the findings, which provide insights into how body shape influences motion perception and consistency.

- **Chapter 4** presents another study exploring the impact of the self-avatar body on virtual presence and embodiment within the immersive virtual environment.

- **Chapter 5** investigates the incorporation of body shape parameters into a deep reinforcement learning framework. We enhanced an existing state-of-the-art framework by introducing shape parameters that simulate biomechanical properties, resulting in more diverse motion variations. This chapter provides a detailed explanation of the framework, the changes made, and the results of these improvements in producing more lifelike motions for avatars.

- **Chapter 6** introduces a framework that customises motions, based on user-defined weight and height, by manipulating physical parameters, specifically the centre of mass. In this chapter, we show how adjusting these physical attributes results in more realistic and varied motions for avatars. An evaluation methodology is also presented, which includes a perceptual study to assess the efficacy of this approach.

– **Chapter 7** presents a case study on the use of physics-based characters in small-scale crowds. This chapter also contains a perceptual study that investigates how variations in body shape affect the ability to detect motion clones in a crowd. The case study focused on the practical applications and challenges of implementing diverse virtual characters in interactive environments.

– **Chapter 8** concludes the thesis by summarising key research contributions, drawing general conclusions, and suggesting potential future directions for expanding on the presented work. This final chapter also discusses some limitations of our work and suggests areas for future research, particularly in the development of more advanced and convincing virtual avatars.

1.5 Relevant Publications

The following publications took place during the preparation of this thesis:

- **Vyas, B.**, Hoyet, L. and O'Sullivan, C., 2023, December. Exploring the Perception of Center of Mass changes for VR Avatars. In ICAT-EGVE 2023-International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments (pp. 1-12). *Best Paper Award*

- **Vyas, B.**, O'Sullivan, C. (2024). ShapeVerse: Physics-based Characters with Varied Body Shapes. In Eurographics 2024 - Posters. The Eurographics Association.

- **Vyas, B.**, O'Sullivan, C. (2024). Shape Shifters: How Body Shapes Change the Perception of Small-Scale Crowd Motions. In 2025 IEEE Virtual Humans and Crowds for Immersive Environments (VHCIE)

Chapter 2

Background Literature

In this section, we will provide an overview of the key background topics and relevant literature that inform our work. The primary focus will be on the main topics which form the basis of our work, namely - human locomotion, motion synthesis & animation, body shape, and applied perception in the context of virtual avatars.

2.1 Human Locomotion

Human locomotion, as described by Inman et al. [Inm66], is a learned process and purely human achievement, particularly the development of bipedal plantigrade progression, i.e., walking on the sole with the heel touching the ground. Locomotion can be broadly defined as the ability to move from one point to another. From this definition, human locomotion can occur through various means: some are driven by human capability, such as walking and running, while it could also involve the use of vehicles, such as cycling or transport. However for the context of this thesis, we will focus mainly on walking and running motions.

Human locomotion has long been an active topic of research studies across diverse fields, including biomechanics [SM03], robotics [LCC19], and ergonomics [AMY08]. These studies provide a rich knowledge of *normal* locomotion patterns and try to utilise that knowledge to replicate the motion for various applications. They also suggest that each individual exhibits personal peculiarities layered over the fundamental pattern of bipedal movement. However, as Andriacchi et al. reviewed [AA00], methods of studying locomotion have evolved over time, and future methods are likely to shift as new questions arise. Despite these changes, several aspects of locomotion are important in understanding the subject. We are mainly interested in exploring features and parameters of locomotion that would help us simulate and validate the motion of virtual avatars.

2.1.1 Gait & Centre of Mass

Locomotion is inherently cyclic and is often characterised by two primary gait modalities: walking and running. A gait cycle is defined as the time interval between two identical events in the motion cycle [Whi14]. This cycle begins when one foot makes contact with the ground and ends when the same foot touches the ground again, referred to as *initial contact*. The *stance phase* concludes when the foot leaves the ground, while *toe-off* marks the beginning of the swing phase of the gait cycle, as shown in Figure 2.1.

Previous research demonstrates that the body's centre of mass (COM) dynamics is essential for understanding locomotion [LF98, DL04, OSK+04]. Saunders et al. [SIE53] even define locomotion as the movement of the COM through space in such a way that energy expenditure is minimised. The COM is an abstract point where the entire mass of the body is assumed to be concentrated. Only the external forces acting on the COM of the distribution system can be enough to determine linear acceleration, and the effects of internal forces could be neglected [Gol02]. In general, this point is at the geometric centre of the body, though it may be necessary to obtain it by direct measurements for irregular and changing shapes like the human body.

Over the years, various methods have been explored to calculate the position and movement of the COM in the human body [Dur08]. The majority of early works relied on cadaver-based data to approximate the position of the COM for individual body segments [CMCY69, Dem55]. However, other techniques, such as gamma-ray scanning [VZ83], later refined by [dL96], have provided alternative methods for determining the position of the COM. Typically, when a person stands upright, their COM is located around the pelvis. However, the COM can shift outside the body in certain postures. For example, when a person bends forward to touch their toes, the COM moves outside the physical body. This movement of the COM is closely related to the body's stability, as it reflects how balanced a person is [Win95]. The COM dynamically shifts in response to bodily movements, such as moving forward or in other directions. As illustrated in Figure 2.1, walking and running consist of distinct phases, each occurring within a single cycle. As a consequence, the trajectory of the COM also significantly differs between walking and running gaits [CTZ76, MAS94].

Walking is characterised by slower body movement and is the most common gait modality [Sch03]. The walking gait cycle consists of two primary phases [Ayy97]:

-> *Stance phase* – the portion of the cycle during which the foot remains in contact with the ground, comprising approximately 62% of the cycle.

-> *Swing phase* – the portion of the cycle where the foot is in the air, accounting for 38% of the cycle.

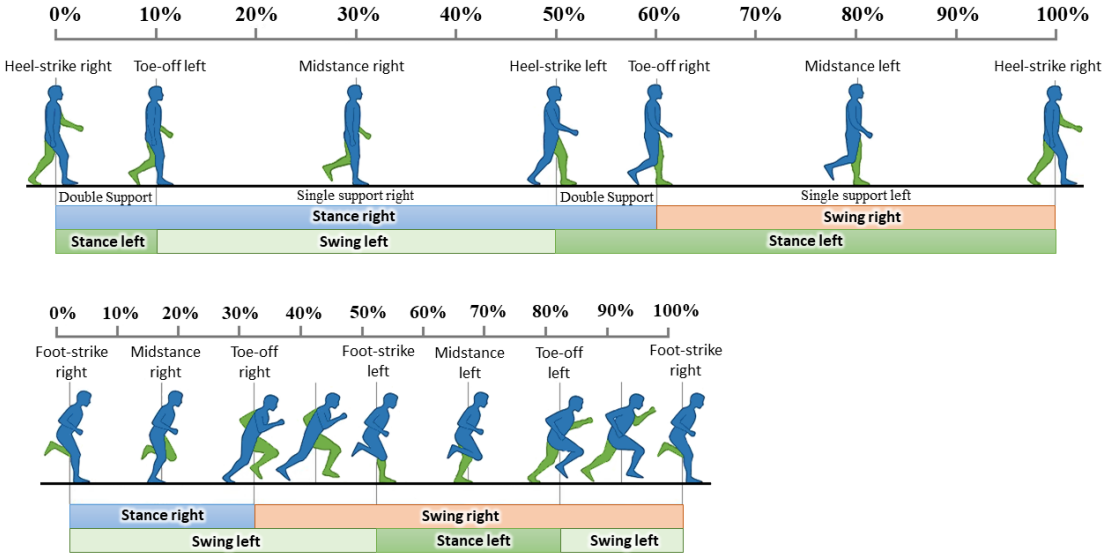


Figure 2.1: Comparison of walking and running gait cycles showing key phases: double support, single support, and swing phases during walking (top), and single support, float, and swing phases during running (bottom).[1]

During walking, the COM vaults over the extended lower limbs in an arc [LDMH13], and there is a notable vertical displacement of the COM as shown in Figure 2.2. This vertical displacement causes the COM to reach its lowest point at certain points during the walking cycle.

Running, on the other hand, occurs at higher speeds and follows a different pattern from walking. Running is often modelled as a spring-mass system [MC90], where energy efficiency is achieved through a bouncing motion. Unlike walking, the running gait features a more flexed posture in the lower limb joints (hip and knee) and lacks a double support phase, meaning both feet never touch the ground simultaneously. Instead, there is a distinct *flight phase*, during which neither foot is in contact with the ground. Lee and Farley [LF98] investigated the distance between the point of ground contact and the COM. They found that this distance decreases more during the walking stance phase compared to running. As a result, the COM moves downward in the first half of the stance phase for running, reflecting the distinct motion mechanics between the two gaits.

[1]Image adapted from BoH, CC BY-SA 4.0, via Wikimedia Commons. License: `https://creativecommons.org/licenses/by-sa/4.0`

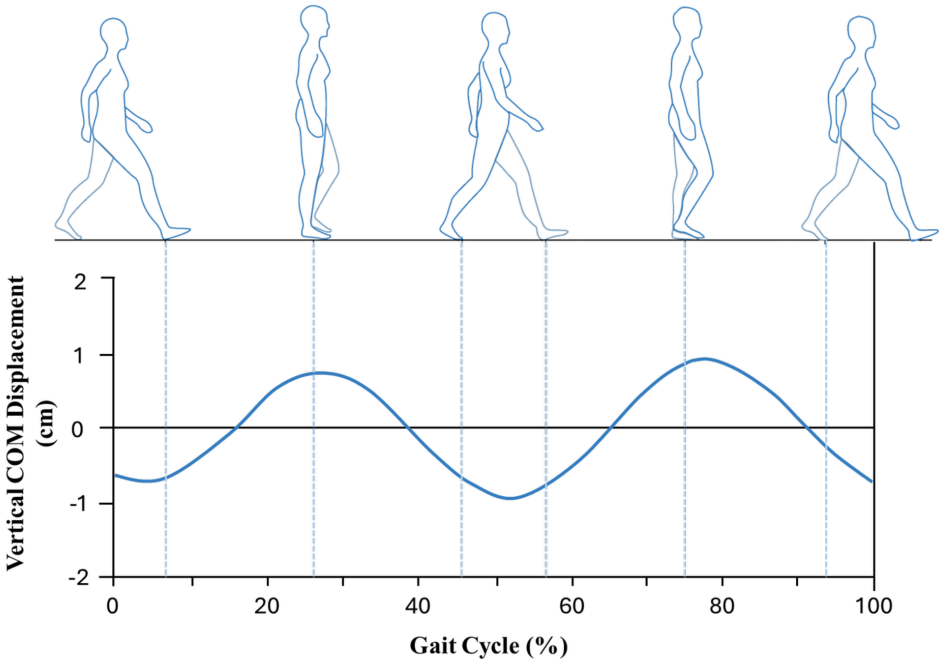


Figure 2.2: Vertical displacement of the center of mass (COM) during a complete gait cycle. The graph shows a sinusoidal pattern of vertical COM motion relative to gait percentage, with key walking postures illustrated above. Adapted from Lobet et al. [LDMH13].

2.1.2 Analysis of Locomotion

We have defined locomotion and some of its characteristics, but it is important to understand how human motion is analysed beyond the basic principles. Motion analysis plays an important role in understanding various parameters and invariants involved in locomotion [Whi14]. While we primarily focus on specific kinetic parameters in this thesis, we briefly introduce other relevant parameters for context. In the review of gait analysis, Roberts et al. [RMP17] identified walking speed, cadence, and step/stride length as the most commonly measured spatio-temporal parameters in locomotion studies. Similarly, Das and McCollum [DM88] identified that certain variables like speed, direction and shape of the support surface are important in particular instances of locomotion. This suggests that some variables rely on underlying body mechanisms that remain consistent during the motion. In addition, it allows minimal variations at the intra-individual level. The spatio-temporal parameters are also widely used in detecting gait abnormalities. For example, Andriacchi et al. [AOG77] used walking speed as a basis for assessing gait abnormalities. Such analyses are commonly applied to compare the gait characteristics of pathological locomotion with normal gait, including in Stroke patients [GME96], individuals with

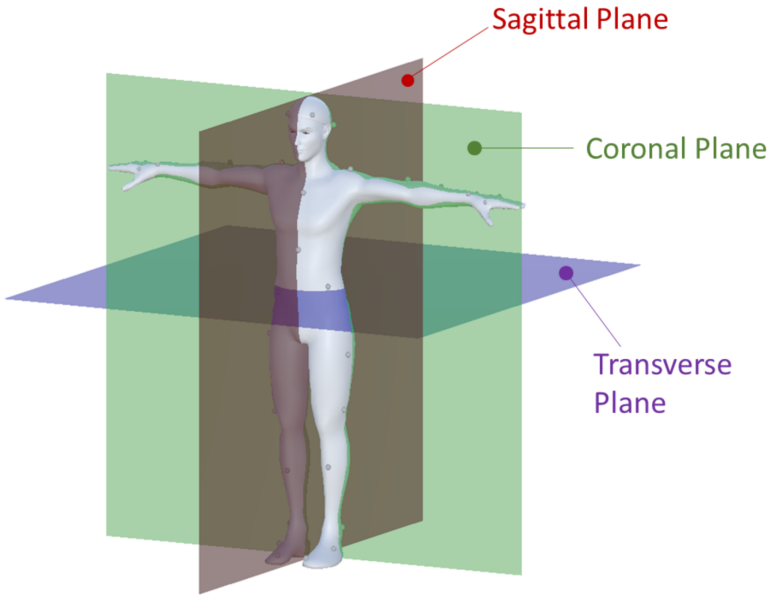


Figure 2.3: Visualisation of the three anatomical planes: the sagittal plane (red), coronal plane (green), and transverse plane (purple). Significant movements occur in all three planes during a gait cycle.

cognitive impairment [EKL+15], multiple sclerosis [GZA09], and idiopathic scoliosis [KdQMS+04].

Beyond spatio-temporal parameters, kinetic analysis computes the forces and energy associated with movement. Since Saunders et al. [SIE53] defined locomotion as the translation of COM, they also identified six key determinants of energy expenditure during locomotion: pelvic rotation, pelvic tilt, knee flexion at midstance, foot and ankle motion, knee motion, and lateral pelvic displacement. These determinants significantly influence the trajectory of the COM during gait, and explain both the vertical and horizontal displacements of the COM [Inm66]. Saini et al. [SKTDR98] also highlighted this behaviour of vertical COM movement. They even further compared four different measurement methods of vertical displacement of COM during walking. Inman's work [IE+53] demonstrated that during walking and jogging, the COM follows a sinusoidal trajectory vertically and horizontally, which adds to our further understanding of gait kinetics. Kinetics is basically a branch of mechanics that deals with motion in relation to forces and torques. It plays a crucial role in understanding the joint forces and moments and, as a result, helps in the mathematical modelling of gait. The lower body joints, such as the hip, knee, and ankle joints, provide valuable kinetic information due to their significant role in forward motion. Common kinetic measures include GRFs, net muscle moments,

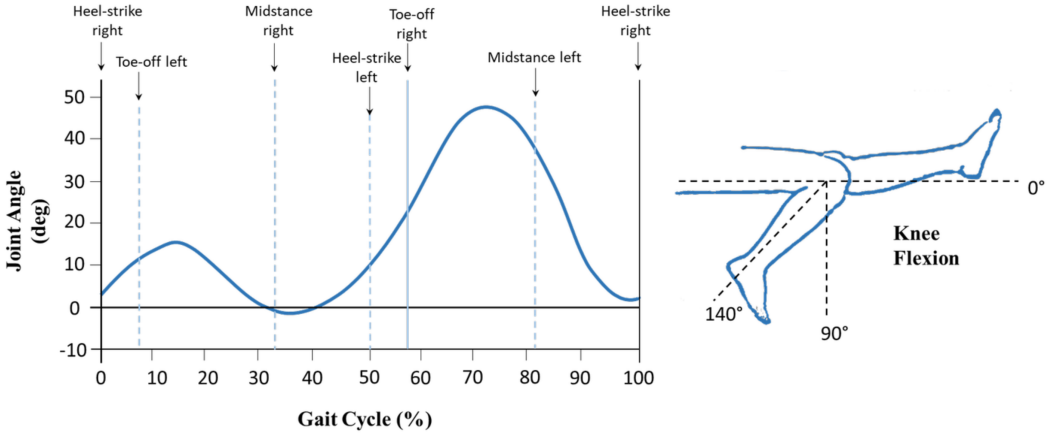


Figure 2.4: The graph illustrates knee flexion during the gait cycle. The joint angle (in degrees) is plotted against the percentage of the gait cycle. The image on the right shows the range of knee flexion from 0° (full extension) to 140° (maximum flexion). The graph is adapted from Whittle's introductory book on Gait Analsyis [Whi14].

joint reaction forces (calculated via inverse dynamics), and joint work and power.

During gait, significant movement occurs across all three planes: sagittal, frontal, and transverse (see Figure 2.3). However, most motion occurs in the sagittal plane, which will be the focus here. Accurate kinetic analysis relies heavily on understanding joint angle variations throughout the gait cycle. For example, the knee joint experiences two *flexion* and two *extension* peaks in each gait cycle. The flexion is the bending of a joint and occurs when the joint angle decreases, as shown in Figure 2.4. While the extension is the straightening of a joint and occurs when the joint angle increases [Vau92]. On the other hand, ground reaction forces (GRFs) are crucial to calculate joint forces and moments, as demonstrated by Bresler & Frankel [BF50] and Cavanagh [CL80]. These forces are particularly relevant for analysing lower extremities, including the ankle, knee, and hip joints, as these joints bear most of the load during locomotion. For example, the knee adduction moment (KAM) [KKC+19] measures the magnitude of the GRF passing medial to the centre of the knee joint, as shown in Figure 2.5.

Recent advancements in motion analysis have facilitated the use of wearable sensors and machine learning to estimate kinetic parameters such as knee flexion and adduction moments. Stetter et al. [SKR+20] used two inertial measurement unit (IMU) sensors placed on the upper and lower knee to collect data, which was then used to train an artificial neural network to estimate these values. Similarly, Giarmatzis et al. [GZM20] showed that joint forces can be predicted using machine learning algorithms, including artificial neural networks and regression-based sup-

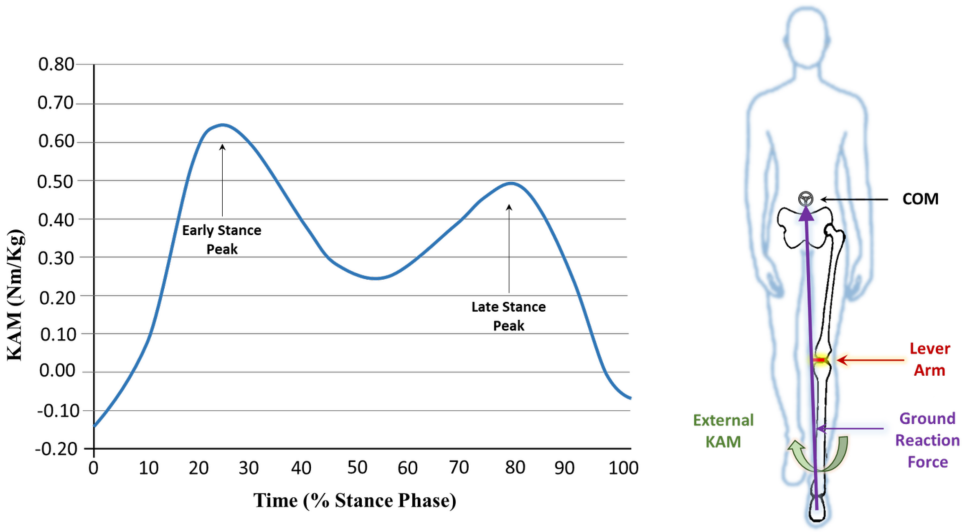


Figure 2.5: Plot illustrating the knee adduction moment (Nm/Kg) during the stance phase of the gait cycle. The moment is plotted against the percentage of the stance phase, emphasizing variations throughout key events such as heel strike, midstance, and toe-off. The diagram on the right shows the knee joint, lever arm, and the ground reaction force (GRF) vector directed towards the center of mass (COM). The plot is adapted from Foroughi [FSV09].

port vector machines, based on joint angle data. However, in our work, we used data from previous literature to compare the simulated results with real human body data.

Anthropometric & Demographic Factors

Anthropometric and Demographic factors such as body mass index (BMI), sex, and body shape could significantly influence gait mechanics. Chehab et al. [EC17] showed that increased BMI is associated with a more pronounced forward pelvic tilt and greater hip flexion, even when walking speed is consistent. These changes are likely due to shifts in the COM as a result of increased body mass. Delabastita et al. [DPPSRJ12] found that obesity alters locomotion by increasing mass and inertia across all body segments, leading to changes in movement patterns and responses to force. Additionally, Browning et al. [BK07] demonstrated that ground reaction forces during walking are higher in individuals with obesity, resulting in greater muscle moments at the hip, knee, and ankle compared to normal-weight adults.

Other factors, such as sex, naturally influence gait mechanics due to differences in body shape. In line with this, Troje [Tro02] investigated how biological motion patterns encode information, developing a framework that applies linear statistical

methods and pattern recognition to analyse these movements. Their research revealed that dynamic aspects of motion, such as gait, convey more information about gender than structural cues alone. These findings underline the importance of body shape and sex in determining locomotion patterns, which is one of the major focuses of this thesis.

In summary, this section presented the characteristics of human locomotion, its influential factors and parameters in motion analysis. The next section will explore how motion is synthesised to create animated virtual avatars.

2.2 Motion Synthesis & Animation

For artistic virtual characters, a variety of creative principles can be applied based on the fundamentals of animation [TJ95]. However, in the case of virtual humans, the goal is to replicate the persona and movement of real humans to enhance realism and immersion in the virtual environment. One of the most common and effective ways to achieve this is through motion capture, which records the movements of real actors and transfers them to virtual avatars. Depending on the type of sensors, motion capture can be distinguished into two categories: highly accurate but intrusive marker-based systems (mechanical, magnetic, optical) and less intrusive, vision-based markerless systems [SVKS19]. While motion capture is effective, it has notable limitations, including high costs for equipment, storage, and data processing. Furthermore, a large amount of motion capture data would be needed to animate multiple characters with unique motions. However, various methods exist to modify and reuse motions from captured databases [BRRP97].

Motion synthesis is a widely used technique in character animation, particularly for introducing variations in repetitive movements such as locomotion. It allows animators to take existing motion data and modify it to suit specific needs. It could be adjusting a character's gait, altering its posture, or adapting movements for different environments. The ability to efficiently edit and customise motion data is crucial in industries such as gaming, film, and virtual reality, where characters need to perform a wide range of actions without requiring entirely new motion capture sessions for every scenario.

Le Callennec and Boulic [LCB06] developed a constraint-based interactive tool for motion editing. This tool offered animators an intuitive, user-friendly interface through which they could manipulate and refine motion data according to the requirements of their project. Inverse Kinematics (IK) is a common technique used in constraint-based methods for controlling character interaction with the environment. It computes the joint angles based on the desired target end-effector position.



Figure 2.6: Example of a motion and motion interactions generated with a deep learning based model, from Starke et al. [SZKZ20].

A constraint-based controller can be implemented using either simple equations [KMA05, Kal08] or some optimisation equations [HP97]. However, this technique suffers challenges such as limited capability to generalise across different motions and failure to handle multiple tasks simultaneously. The different types of motions often demand different equations to be solved, which could be tricky at times. To solve similar problems, Lyard et al. [LMT08] tackled this issue by proposing a method for editing and retargeting existing motion to fit the body shape of a target character. Their approach used a space-time optimisation technique to prevent self-penetration, a common issue in character animation where different parts of the character's body intersect unnaturally. Additionally, they employed radial basis functions for interpolation, enabling their system to work in real-time. However, achieving high-quality motion output still remained a challenge. Multon et al. [MKHK09] presented a framework that synthesises new motions with morphological adaptations. This system adjusts motions based on the physical characteristics of different characters, such as height or limb length. It also applies kinematic and physically-based corrections to ensure the generated motions are realistic and physically plausible.

Various other approaches have been explored for efficient motion generation and editing. Ahmed et al. [AMH01] employed techniques such as time warping and Principal Component Analysis (PCA) to generate new animations from pre-captured motion data. Time warping allows the synchronisation of motions that occur at different speeds, while PCA helps reduce the dimensionality of motion data, making it easier to manipulate. Glardon et al. [GBT04] further applied PCA to create lower-dimensional representations of walking motions, allowing them to generate new walking patterns with less computational complexity. These approaches demonstrate the versatility of PCA in motion editing, as it can be used to simplify and generalise motion data while maintaining the essential features of human locomotion.

Motion Blending is another technique which blends together two similar motions by applying weight adjustments, resulting in specific motion characteristics

[FHKS12]. Park et al. [PSS02] proposed an approach to generate locomotion in real time based on parameters like speed, turning angle, and style. Meanwhile, Huang et al. [HK10] proposed inverse motion blending for other motion tasks, such as the reaching task. However, motion blending sometimes produces unnatural motions and may require different techniques based on the task. To overcome this problem, Kovar et al. [KGP02] proposed the concept of *Motion Graphs*, which computes the similarity between two poses based on the distance between joints for a smooth transition. Later, other works on motion graphs compared joint angle and velocities instead of distance [LCR+02, WB03]. Although motion graphs require a large dataset, some methods employ statistical models to learn from these datasets and generate task-specific behaviours. For instance, Brand and Hertzmann [BH00] developed a model that learns motion styles from a large database, allowing for variations in action generation through interpolation or extrapolation of the learned styles.

As reviewed by Mourot et al. [MHLC+22], the recent advancements in deep learning have further expanded the approaches of motion synthesis. Holden et al. [HSK16] introduced a technique that utilises learned motion manifolds to generate human motions. In particular, the technique uses the hidden units of a convolutional autoencoder to create high-level motion representations. This method facilitates seamless blending and editing of different motion types. Later, Holden et al. [HKPP20] proposed a model for *Motion Matching*, which uses neural networks to find the most context-appropriate motion at each step, thereby producing more natural results. Holden et al. [HKS17] later introduced the motion phase to depict the progression in time of synthesised animations. Later, the use of the mixture-of-experts (MoE) architecture also became popular in other works [MSK22, SMK22]. However, despite these advancements in motion editing techniques, several challenges persist. Many approaches still rely on large, high-quality motion datasets or require manual adjustments to constraints. This limits the efficiency and adaptability of these techniques, particularly for complex or highly specific motion tasks.

2.3 Body Shape of Virtual Avatars

Humans exhibit wide diversity in both physical appearance and behaviour. It is crucial to understand how body shape influences the perception of motion, especially when modelling virtual avatars. Exploring body shape in virtual bodies has been a longstanding area of interest in computer vision, with significant research focused on fitting 3D virtual bodies to images [SBB07, PZZD18] and videos [HBL+17, KAB20]. However, in the context of virtual avatars, body shape can influence not only the

self-avatars but also the perception of others within immersive virtual environments. Latoschik et al. [LRG+17] found that the realism of others' avatars can impact a user's sense of immersion, suggesting that both self-representation and the representation of others contribute to the overall virtual experience. Additionally, body shape can influence spatial perception in virtual reality, with research showing that the presence of a realistic avatar can improve one's accuracy in perceiving spatial layouts [MCRTB10, RIKA08].

The demand for virtual characters with diverse body shapes has grown over the years due to their wide range of applications and the increasing need for more personalised digital representations. Lee & Magnenat-Thalmann [LMT01] designed a rapid 3D morphing system, using interpolation and 2D-image metamorphosis, to create avatars with varying body shapes. Similarly, Aubel and Thalmann [AT00] introduced a toolkit that generates versatile animated figures using a multi-layer model consisting of skeleton, muscle, fat, and skin layers. However, their pipeline focused primarily on the visual appearance of these characters rather than biomechanical accuracy.

The representation of biomechanically accurate bodies in virtual characters is not only a technical challenge but also raises cultural and psychological considerations. For example, Rowe [Row19] examined how animated female characters influence young girls' body image, classifying 239 characters into four common body types: Hourglass, Pear, Rectangle, and Inverted Triangle. The study highlighted the impact these characters can have on viewers, particularly when they are of similar age to the characters. This suggests that body diversity in virtual avatars can influence self-perception in multiple ways.

Early research in the context of body shapes explored how viewers perceive gender and attractiveness in virtual characters. McDonnell et al. [MJH+07] conducted a study to assess the impact of body shape (categorised as male, female, neutral, or point-light) and the motion of characters on how participants identified their sex. The findings demonstrated that both the physical shape of the body and the motion dynamics played critical roles in influencing perceptions. Similarly, Johnson and Tassinary [JT07] reached a comparable conclusion in their investigation of attractiveness judgements. Their research further confirmed that both the appearance and the movement of virtual bodies significantly affect how individuals perceive attractiveness. These findings demonstrate that accurate representation of both body shape and motion is essential in creating virtual humans that are perceived as realistic and relatable.

In recent years, advanced 3D models have been developed to better capture and represent a wide variety of human body shapes. These models account for variations



Figure 2.7: The body shape model introduced by Anguelov et al. [ASK+05]. The model can be used for shape completion - generating a complete surface mesh given a limited set of markers.

across gender, age, and other factors, aiming to generate realistic human postures and movements. Such models are crucial for enhancing the realism and utility of digital humans across industries like gaming, virtual environments, and simulations. Furthermore, these models are finding applications in ergonomics, healthcare, and even virtual therapy, where accurate human representations can assist in everything from physical rehabilitation to mental health treatments. We will discuss two of the most popular body modelling systems.

SCAPE: Shape Completion and Animation for People

SCAPE (Shape Completion and Animation for PEople) is a particularly significant system from the early research, introduced by Anguelov et al. [ASK+05]. SCAPE used a large set of 3D range scans from various human subjects to build its models. These range scans include individuals in different poses, allowing the system to capture a broad array of body shapes and movements. To register these scans, SCAPE uses a semi-automatic marker placement method, which enables it to account for individual differences in body shapes and poses across subjects. SCAPE relies on data-driven approaches to model shape variation. It utilises the captured scans to represent both the subject's shape and skeletal pose. These variables are incorporated as functions of shape variation, allowing the system to recreate a full 3D body mesh from a limited number of markers. The system efficiently searches for optimal model parameters in both shape and pose space, ensuring that the resulting mesh is highly realistic and anatomically consistent. Figure 2.7 illustrates the output of the SCAPE system, which shows its ability to accurately recover human shapes and poses from minimal input data.



Figure 2.8: The SMPL model presented by [LMR+15a], is a realistic learned model of human body shape. The model is based on a mesh composed of 6,890 vertices and 23 joints.

SMPL: Skinned Multi-Person Linear Model

Another highly sophisticated and popular approach to body modelling is the Skinned Multi-Person Linear (SMPL) model, presented by Loper et al. [LMR+15b]. The SMPL model is known for its ability to represent a wide variety of body shapes in natural human poses with a high degree of realism. It has become one of the most popular models used in computer graphics for human shape representation. As shown in Figure 2.8, the SMPL model is based on a mesh composed of 6,890 vertices and 23 joints. It is trained using a large dataset of human body scans, specifically the CAESAR dataset [RDP99], which captures a wide range of body types and sizes. The SMPL model integrates several key parameters to achieve its realistic representations. These include a rest pose template, blend weights, pose-dependent blend shapes, identity-dependent blend shapes, and a regressor for determining joint locations based on vertex positions. A major advantage of the SMPL model is its compatibility with existing rendering engines and graphics pipelines, making it widely applicable in various industries. The model's reliance on a linear blend shape framework helps avoid common artefacts in areas of complex joint movement, such as the elbows, shoulders, and knees. By maintaining smooth transitions in these regions, SMPL ensures that the resulting animations appear lifelike and free of distortions.

Motion Capture and Body Shape Estimation

To create accurate body shape models, data-driven methods often utilise extensive datasets of human motion and shape information. Motion capture technology is a common tool used in this process, enabling the capture of detailed movement data from real humans. Several publicly available datasets, such as those from

CMU Motion Capture Database [CMU], SFU Motion Dataset [YY], and the KIT Whole-Body Human Motion Database [MTD+15], provide hours of recorded human motion sequences. These datasets offer valuable information regarding 3D skeletons and joint orientations, which can be used to animate virtual characters. However, one limitation of these datasets is that they often lack detailed information about the body dimensions and shapes of the subjects involved.

Loper et al. [LMB14b] addressed this limitation by developing techniques to estimate body shape and pose from sparse marker data. A related method, SOMA [GB21], automates the process of labelling motion capture point clouds by associating captured points with corresponding markers on the subject's body. This autolabeling process greatly simplifies the creation of realistic body models by facilitating extra marker points in the point cloud. This helps fit the 3D mesh tightly to the point cloud. Their work facilitated the accurate capture of both motion and body shape for the motion capture datasets. As a result, the AMASS dataset [MGT+19a] was created with more than 40 hours of motion data, spanning over 300 subjects and more than 11,000 motions. The dataset contains parameters for a realistic 3D human mesh, represented by an SMPL body model.

2.4 Physics-based Character Animation

Motion capture and motion synthesis approaches require predefined movements for every scenario, making it difficult to account for unpredictable interactions like collisions or external forces. Physics-based methods offer a solution by simulating motion based on mechanical laws. These approaches compute the character's movement by deriving accelerations from forces and calculating joint rotations from torques, allowing characters to respond dynamically to their surroundings. For example, if a character is pushed, a physics-based system can naturally simulate the resulting stumble or fall. Armstrong and Green [AG85] proposed a method for simulating the motion of virtual characters by representing them as connected segments and joints. Each segment has its own mass, and equations of motion are formulated for each one based on its connection to adjacent joints. These equations describe how forces, torques, and accelerations influence the segment's motion. This allows the character's movement to be generated by solving these equations over time. In general, forward dynamics allows motion to be calculated from known forces, while inverse dynamics calculates the necessary forces based on desired movement patterns. This framework remains central to physics-based animation today, with real-time simulators solving these equations to generate realistic motion.

As reviewed by Geijtenbeek and Pronost [GP12], numerous approaches exist

for animating virtual humans using physics-based techniques. Broadly, these approaches can be classified in several ways, depending on their implementation. Forward simulation techniques apply control inputs in real time to produce motions subjected to the laws of physics. In contrast, optimization-based approaches compute entire motion sequences by solving for trajectories that satisfy physical constraints. However, in this thesis, we focus on specific approaches that are more relevant to our objectives. În this section, we first discuss the controller-based methods. These methods apply control in real time as part of a forward simulation framework. They are typically task-specific and can be relatively inflexible in practice. Specifically, the controller restricts the forces and torques at each joint to generate a desired motion. Following this, we discuss the reinforcement learning approach. This approach automates the control process so that the forces and torques are handled through some learned policy.

Controller based Methods

Most physics-based systems use simplified models of the characters to simulate motions. Since physics simulation is computationally expensive, the use of simplified models allows for the exploration of different control systems techniques. One of the earliest examples is the work by Witkin and Kass [WK88] who introduced spacetime constraints. They formulated the desired animation as an optimization problem over both space and time, constrained by the laws of physics. Their approach allowed characters such as the iconic Luxo lamp to perform coordinated and physically plausible actions. Building on the idea of modular control, Faloutsos et al. [FVdPT01] developed composable controllers to generate actions such as balance, arm reactions, and recovery from external disturbances. They used a simple rigid-body skeleton, yet their system could produce complex and realistic character motion. Later, Liu et al. [LHP05] focused on learning control parameters directly from data. They introduced a non-linear inverse optimization technique that extracts the physical style and control structure from motion capture sequences. This enabled the retargeting or editing of motion by adjusting the underlying physical properties of the character, while preserving stylistic elements.

In the work by Martinez et al. [MNMO08], particle filters constrained by human physiology were used for tracking different parts of the human body. By tracking and replicating the motion, a simulated character could take the same pose and perform similar actions as their human counterpart. These simulated characters can also be made interactive to their surroundings. In a hybrid approach, Alvarado et al. [ARC22] combined kinematic constraints and lightweight physics, thereby

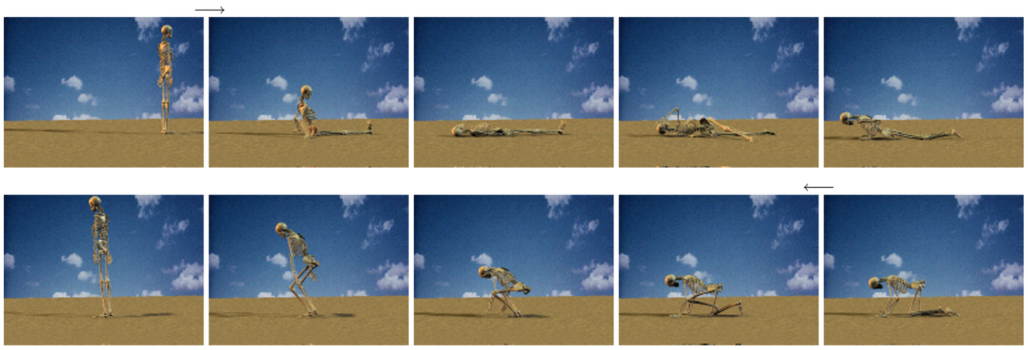


Figure 2.9: Controller-based method. A physics-based skeleton performing fall, rolling and getting up motions. Faloutsos et al. [FVdPT01] created composable controllers by defining a set of preconditions to identify the current state and evaluate future possible states.

making the character interact with objects in its surroundings. Earlier, Liu and Hodgins [LH17] employed a deep Q-learning algorithm to learn schedulers capable of designing control systems for dynamic behaviours. This algorithmic approach enabled the generation of characters that could handle perturbations and be reactive.

Motion adaptation based on character body type has also been explored. Lyard and Magnenat-Thalmann [LMT08] proposed a space-time optimisation approach and a variant of the Normalised Radial Basis Function (NRBF) to fit motion to different body types at runtime. On the other hand, Al-Asqhar et al. [AAKC13] presented a framework for interactive motion adaptation that could handle significant deviations in mesh structure and character morphology. This framework is particularly useful for artists as they can do motion retargeting, interactive character control and deformation transfer for scenes that demand close interactions. A novel space-time optimisation-based physical controller developed by Al Borno et al. [ABRB$^+$18] learned how to adapt to different body shapes and constraints. They showed that the simulated motions look appropriate for the character's body shape, and the motions are also robust to perturbations.

However, as discussed by Geijtenbeek et al. [GP12], physics-based animation can be achieved through various other techniques. Since these characters are in a simulation, there are various tricks to achieve similar results. For example, Levine and Popović [LP12] introduced a quasi-physical simulation method that applies non-physical forces to the character's root. This optimizes the character's equations of motion for the control system for animation. Simulating major muscles and tissues in the human body is another common practice adopted in these controllers. Geijtenbeek et al. [GVDPVDS13] introduced muscle routing optimisation, which enabled the design of muscle-based control strategies that are adaptable to various

character morphologies. As our primary interest lies in the perception of motion based on body shape rather than in generating various stylised movements, a basic feedback control system is used in our work, which we will discuss in Section 6.

Deep Reinforcement Learning

Physics-based character animation typically operates in a high-dimensional space, due to the complexity of human motion and the interactions between joints, muscles, and external forces. This presents a challenge in crafting animations that are both realistic and responsive to dynamic environments. As a result, unsupervised learning techniques are particularly well-suited for addressing these problems. They allow models to automatically learn patterns and structures from data without the need for explicit labels or supervised signals. Unsupervised learning is especially valuable in scenarios where the goal is to produce characters that can react naturally to unseen environments or unpredictable situations. Among these approaches, Deep Reinforcement Learning (DRL) has shown considerable promise in recent years, advancing the field significantly.

As reviewed by Kwiatkowski et al. [KAK+22], using reinforcement learning methods has become a popular approach for controlling physics-based characters. Early work by Liu et al. [LYVdP+10] introduced an open-loop trajectory optimisation method. This method was capable of imitating the motion capture data for highly dynamic movements like flips. Subsequently, Liu and Hodgins [LH17] proposed controllers trained using deep Q-learning, where a short-horizon linear feedback controller acted as a single action unit. This was later adapted to learn basketball playing skills [LH18]. Peng et al. [PBYVDP17] were the first to explore policy gradient-based RL methods. This enabled characters to imitate motion capture data used as reference clips. Their approach also trained a low-level policy to make characters reach target locations while adhering to the imitation of reference motion. A major advancement came with *DeepMimic* by Peng et al. [PALvdP18], which involved training controllers for diverse motions, including complex actions like backflips. DeepMimic uses imitation learning, where control policies are learned by defining rewards for closely mimicking a reference motion, typically consisting of short motion capture clips. Key elements in this framework are reference state initialisation and early termination, making it an effective approach to imitation learning. The method employs Proximal Policy Optimization (PPO) [SWD+17], a technique widely used in contemporary DRL due to its performance and ease of implementation. Further expanding on these ideas, Peng et al. [PMA+21] explored adversarial imitation learning with unstructured motion data. By combining a high-

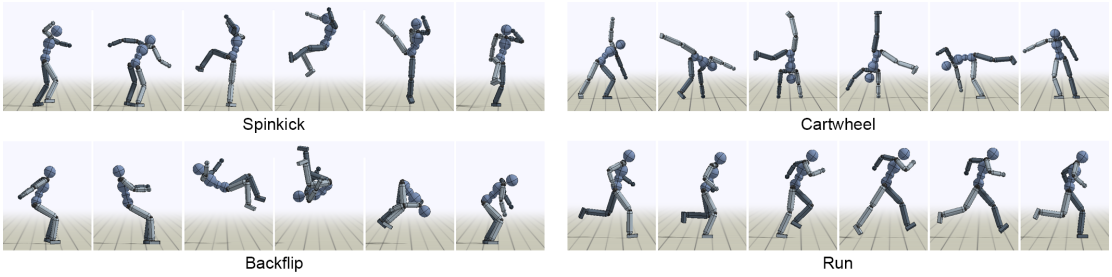


Figure 2.10: Deep Reinforcement Learning. Peng et al. [PALvdP18] proposed a method to mimic the animation from motion capture data. In future works, as in [PMA+21], they incremented such techniques to obtain added behaviour to control simulated motions.

level reward function for behaviour control with low-level control, this approach produces high-quality animations for both humanoid and non-humanoid characters.

Beyond reference motion clips, researchers have also explored the use of computer vision techniques for pose estimation, which serves as input for imitation learning methods. For instance, Peng et al. [PKM+18] used videos as the input source instead of motion clips. Later, Yuan et al. [YWS+21] introduced a framework which trains policies to control physics-based characters while conditioning on a monocular video. Some recent work has developed systems that allow users to interact with characters directly. Bergamin et al. [BCHF19] proposed a two-step approach to simulate a controlled character using unstructured motion capture data, using a kinematic controller for motion matching, followed by reinforcement learning training. This resulted in responsive characters that react to user input. Other works, such as Chentanez et al. [CMM+18], emphasised the importance of scalability in their frameworks. Won et al. [WGH20] achieved both scalability and variety by creating clusters of similar motions for which dynamic controllers, called *experts*, were trained. A single dynamic controller was then learned by combining these experts, producing a large variety of motions. Further scalability was achieved by focusing on motion style. For example, Ma et al. [MYT+21] used loose space-time constraints (spacetime bounds) in combination with heuristic style rewards to generate diverse motion styles.

In addition to reinforcement learning approaches, generative models have also been developed to construct *manifolds for physics-based characters*. Won et al. [WGH22] developed a controller using conditional variational autoencoders (VAEs), conditioning controllers on a random vector sampled from a standard Gaussian distribution. They pre-trained controllers by learning DRL policies for downstream tasks. Earlier, Merel et al. [MHG+18] proposed a motor architecture using an encoder-decoder policy using expert trajectories generated by the pre-trained imi-

tation policies. Similarly, Won et al. [WGH21] used a task encoder and a mixture-of-experts motor decoder to train policies for multi-agent tasks. Their two-way approach first learns an imitation policy and then transfers it into a task-based policy. However, since we are particularly interested in how body shape affects motion variety in virtual avatars, the method that inspired our approach was proposed by Won and Lee [WL19]. They developed parametric controllers for body shape variation, where the state space consists of the body shape state, dynamic state, and additional sensory information. Our framework, which is detailed further in Chapter 5, builds upon this approach.

2.5 Applied Perception of Motion

Over the years, extensive research has been conducted on how humans perceive biological motion, particularly focusing on how subtle differences in motion can convey critical information about identity, gender, age, and other traits. One of the pioneering studies in this field was conducted by Johansson [Joh73], who introduced the concept of biological motion using an innovative display of point-lights placed on the joints of a human body moving against a high-contrast background. Despite the minimal visual information, observers were able to identify human figures and interpret the motion. This demonstrates the human brain's remarkable ability to reconstruct complex motion from limited visual cues. Further studies, such as those by Cutting and Kozlowski [CK77], revealed that even familiar motions, such as a friend's walk, could be recognised from these point-light displays, thereby highlighting the human capacity for detailed motion perception.

Troje [Tro02] extended this foundational work by developing a framework to extract specific information from biological motion patterns, such as gender. Through their research, they showed that observers could reliably classify the gender of individuals based on subtle differences in walking styles, underscoring how motion carries rich social and biological information. Patla's research [Pat97] emphasised the critical role of the visual system, not only in regulating our own locomotion but also in how observing others' movements influences our posture and gait. His work revealed how individuals adjust their own movements in response to observing the motion of others, thereby showcasing the interactive nature of motion perception.

The perception of biological motion extends beyond recognising individual identities or gender. Jokisch and Troje [JT03] demonstrated that biological motion also serves as a cue for estimating the size of walking creatures. They found that observers could infer the size of a walking figure based on the dynamics of its movement, thus adding another dimension to how motion is processed by the brain. Jain

et al. [JAA+16] explored a similar concept by showing that viewers could accurately distinguish between the motions of children and adults performing the same action. These findings illustrate the vast amount of information embedded in motion, from age and size to identity and physical characteristics. This further highlights the importance of motion style in shaping perception.

Niay et al. [NOZ+20] conducted a study focused on the perception of walking animations, revealing that observers could select their preferred walk ratio for virtual characters within a range established by existing literature. They defined walk ratio as the ratio of step length and step frequency, which is an invariant of human walk. This study reinforced the idea that humans are highly sensitive to subtle variations in movement, especially in walking, which is the most frequently studied motion in animation and motion perception. In contrast, Kenny et al. [KMH+19] found that inconsistencies between body shape and motion were less noticeable to observers, particularly in interactions with objects. However, they found an exception in cases involving walking. This suggests that while certain inconsistencies in animation can go unnoticed, body shape remains a key factor that influences how we perceive locomotion in virtual characters. One limitation of their work was that the manipulated objects were not shown in the animations. This may have made it easier for viewers to explain the motion by making assumptions about the object. For example, if a character applied a lot of force, viewers might assume the object was heavy, even though it wasn't shown. More recently, Yamac et al. [YON24] studied how people notice errors in the physical behaviour of animated scenes. They found that the size of an object being lifted had a strong effect on how heavy it seemed to viewers, showing how visual cues about objects influence our understanding of motion.

Another intriguing dimension of motion perception is its relationship to perceived attractiveness. Klüver et al. [KHT16] investigated how manipulating internal consistencies between anthropometric and kinematic cues affected perceived sexual attractiveness. In other words, they investigated whether consistency (i.e., the match between the original body shape and motion of an actor) affects attractiveness. Their findings showed a significant correlation between internal consistency and perceived attractiveness. This suggests that observers favour movements that align with the physical traits of the individual performing them. Similarly, Hoyet et al. [HRZ+13] explored how distinctiveness and attractiveness were perceived while observing human motion, and found that non-distinctive motions were rated highly, whereas the opposite was true for distinctive motions. These studies demonstrate that human observers are sensitive to identity-related cues in human motion, as well as to how well the motion matches the displayed performer's physical appearance.



Figure 2.11: Distinctiveness & Attractiveness of human motions were explored by Hoyet et al. [HRZ+13]. They found that distinctiveness in one motion type does not transfer to another.

Recently, Russell et al. [RSM+22] investigated whether observers could detect inconsistencies between shape and motion in virtual characters. Their research revealed that except in cases involving walking, detection rates for these inconsistencies were generally low. This suggests that viewers are highly sensitive to inconsistencies in certain motions, such as walking. This is likely due to their familiarity with this motion in everyday life. In contrast, other forms of movement may be less familiar and, therefore, appear more acceptable, despite the inconsistencies present. This observation, therefore, allows for greater flexibility in animation for these motions without compromising believability. In our work, we used a motion database featuring multiple male and female actors performing locomotion at the same speed. We also explored how additional physical parameters, such as body size and shape, influence motion perception in this context.

While we have, up to now, discussed studies that are focused primarily on individual virtual humans, other works have also examined motion perception in crowd settings. In virtual crowd perception, McDonnell et al. [MLD+08] explored how variety in appearance and motion affects how we perceive crowds. They created stimuli that depicted crowds of 20 characters, where the number of replications of the appearance or motion of the characters was varied. Participants were asked to detect *clones* as quickly as possible. They found that the detection of motion clones was more challenging than for appearance clones, especially when visual distractions such as colour variations or random orientations were introduced. This research highlights how variations in motion can be effectively used to create the perception of diversity in crowds, even when the same motion is repeated across multiple characters. In a follow-up study, McDonnell et al. [MLH+09] used eye-tracking technology to investigate which body parts participants focused on when identifying appearance clones. They found that the head and upper torso were the most fixated areas, indicating that these regions play a critical role in how we perceive and differentiate virtual characters. Adili et al. [ANZ+21] explored the effect

of motion cloning in virtual crowds, showing that even with two or more repeated motions, the crowd was perceived as having sufficient variety. They randomised the visual appearance of the characters, drawing from a broad pool of available character designs. This approach demonstrates that even with limited motion variety, a randomised visual appearance can create the perception of diversity. Similarly, Prazak and O'Sullivan [PO11] developed a framework that allowed motion from one character to be applied to several others, in order to investigate how noticeable motion clones in a crowd are. Their findings suggest that using just three desynchronised motions may be as effective as having entirely different motions for every character.

Shi et al. [SOWO17] examined the role of body shape in creating perceptually varied crowds. They demonstrated that by incorporating highly distinctive shapes, such as extreme body sizes or proportions, greater perceived variety could be achieved. However, they also noted that extreme shapes made it easier for observers to spot clones, while more average shapes reduced the overall variety but made clones less noticeable. Our research builds on this concept by investigating whether the choice of distinctive body shapes can also help mask motion clones in virtual crowds. Unlike Shi et al.'s work, which focused on static models, our work also explored how body shape affects the perception of motion clones in dynamic, moving crowds.

2.6 Chapter Conclusion

In this chapter, we briefly explored the characteristics of human locomotion, emphasising the importance of understanding the relevant factors and parameters. While recent years have seen extensive research into techniques for creating virtual avatars, our primary focus is on body shape and motion. Furthermore, although there are numerous ways to animate these avatars, we have primarily explored physics-based character simulation approaches. We also discussed how the perception of motion and other characteristics of virtual avatars are interdependent, thus highlighting the significance of the questions this thesis aimed to answer.

Chapter 3

Body Shape & Motion

3.1 Introduction

As discussed in Section 2.3, body shape and the overall outlook of virtual avatars play a significant role in determining how users perceive them. This suggests that the visual fidelity of a virtual human can influence the sense of immersion, and factors such as body shape, facial expressions, and clothing can drastically affect how realistic the virtual human feels to the viewer. This chapter presents a study that explored our research question **Q1** (in Section 1.1), the relationship between avatar body and motion within virtual scenarios.

We investigated the influence of an avatar's body shape on the perception of motion. Specifically, we examined how motions captured from actors with varying BMI are perceived when applied to avatars of different body shapes. BMI (Body Mass Index) is a widely used anthropometric measure that approximates body fat based on a person's mass and height. It is calculated using the formula:

$$\text{BMI} = \frac{\text{mass (kg)}}{\text{height (m)}^2} \tag{3.1}$$

This metric has been extensively adopted in health and biomechanical studies due to its simplicity and ease of use. According to the World Health Organization (WHO), BMI categories are defined as underweight (BMI < 18.5), normal weight (18.5–24.9), overweight (25.0–29.9), and obese (30.0 and above) [Org95]. In this chapter, we aimed to determine whether users would rate perceived consistency higher when the avatar's body shape aligns with that of the actor from whom the motion was captured. The results showed that BMI influences how people perceive the motion of avatars with different body shapes, highlighting the relationship between body shape and motion perception.

The following sections present the experimental setup, stimuli, results, and key findings of the study, shedding light on how avatar body shape influences perceived motion consistency.

3.2 Body & Motion Consistency

Populating virtual environments with animated virtual characters often involves animating 3D body models with realistic-looking motion. Motion capture technology has been widely adopted to achieve such realistic motions for virtual humans. This technology captures the precise movements of human actors and then transfers or *retargets* onto virtual characters. However, this process becomes more complex when the body measurements of the virtual character differ from those of the original actor who performed the motion. For example, the motion captured from an actor with a slender frame might look unnatural when applied to a bulkier character, as the physical differences between the two body types would imply different biomechanical constraints and motions. Troje [Tro08] have suggested that humans are susceptible to subtle discrepancies in both movement and the visual appearance of familiar patterns, such as walking. Even smaller differences in how a virtual human moves or is shaped can catch the viewer's attention. As a result, this could potentially disrupt the perception of realism.

A study by Russell et al. [RSM+22] examined this topic by conducting some perceptual experiments. They aimed to explore the relationship between body shape and motion in virtual humans. Specifically, they determined whether participants could detect inconsistencies between the body shape of a virtual character and the motion being performed. In this context, *consistency* refers to a scenario where both the body shape and motion correspond to the same original actor. This means the body shape belongs to the same individual from whom the motion is recorded. *Inconsistency*, on the other hand, refers to a situation where the body shape and motion do not match. In such a scenario, an actor's motion is applied to a character with a body shape different from the original actor. The results of the study revealed that while participants struggled to detect inconsistencies in most types of motion, they were more noticeable in walking animations. Since people are highly familiar with motions like walking, it is very likely that even small differences from natural locomotion become more apparent to the viewer.

Building on these findings, this study aimed to investigate further how body shape affects the perception of virtual character motion. Specifically, we focused on walking and jogging motions, which are common and known by most people. For this study, we examined how individuals perceive virtual characters whose move-

ments were captured from actors with different body mass indices (BMIs) and then retargeted to characters with varying BMIs. BMI is a widely used metric to categorise body weight, and according to the World Health Organization (WHO), it is classified into three categories: underweight (BMI < 18.5), normal weight (BMI 18.5–24.9), and overweight or obese (BMI > 25.0) [O$^+$95]. Here, we used BMI as a means of variation in body shape. To generate realistic body shapes and motions for our virtual characters, we used the MoSH++ algorithm [LMB14a] combined with the SMPL-X statistical body model [LMR$^+$15a]. MoSH++ is a powerful tool for recovering human motion from sparse sensor data, making it an ideal choice for our motion capture data. Then, this recovered motion data was rendered using SMPL-X, which provides a comprehensive and flexible model for creating detailed and accurate virtual body shapes. This pipeline allowed us to create virtual humans with a wide variety of body shapes and realistic movements for our experiment.

The experiment was specifically focused on examining how participants rated the consistency of motion based on the body shape of the virtual characters. The key question was whether participants would perceive any difference in motion depending on whether the body shape and motion of the virtual human were consistent or inconsistent. Participants were presented with animations of virtual characters performing walking and jogging motions. The body shapes and motions were either matched (consistent) or mismatched (inconsistent), and participants were asked to rate the perceived consistency of the animations using a 5-point Likert scale. The Likert scale allowed us to easily quantify the degree to which participants felt the motion aligned with the character's body shape, with higher ratings indicating a stronger sense of consistency.

We hypothesised that participants would rate consistent motions more highly than inconsistent ones. In other words, the virtual characters would be perceived as more realistic when the body shape and motion were well-matched. We also predicted that ratings would be more similar for motions within the same BMI group, as characters with similar body shapes might have similar motion patterns, making any differences less noticeable. For example, a character in the normal BMI range might be able to convincingly perform the walking motion of another character within the same BMI range. However, applying the motion of a lower BMI actor to a high BMI character might result in lower consistency ratings.

3.3 Experiment

For the experiment, we required motion capture data from different subjects with varying BMIs. We found the motion and marker dataset by Hoyet et al. [HRZ$^+$13]

as the most suitable dataset for our purpose. The dataset contains 15 male and 15 female subjects of known and varying BMI. The full dataset includes a variety of motions from actors with a range of BMI values, which was well suited for the experiment. Furthermore, a metronome was used to practice each motion before it was captured (but turned off during capture), to harmonise the speed and frequency of movement across all actors for each gait. Therefore, the motion data vary solely based on the unique motion style of each actor. The walking motion was recorded at the frequency of ≈ 112 steps per minute, while jogging motion was recorded at $\approx$ 138 steps per minute. For the experiment, four actors were selected for each male and female category, as shown in Figure 3.1. They were further evenly distributed between low and high BMI groups. We tried to select actors in such a way that the actors' body weight difference was maximised. However, the heights were maintained at similar values to avoid any retargeting artefacts in the stimuli. The details of the selected actors used to create stimuli are in Table 3.1.

Table 3.1: Attributes of female and male actors used to create stimuli.

	ID	**BMI category**	**Height (m)**	**BMI**
Female Actors	Hi1	Higher	1.60	28.1
	Hi2	Higher	1.64	25.3
	Lo1	Lower	1.60	24.8
	Lo2	Lower	1.60	19.5
Male Actors	Hi1	Higher	1.79	32.1
	Hi2	Higher	1.83	35.8
	Lo1	Lower	1.81	22.7
	Lo2	Lower	1.82	21.7

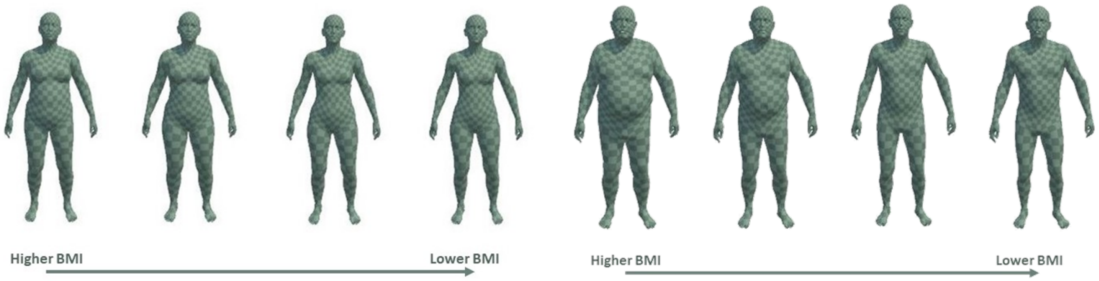


Figure 3.1: Male and female avatars used as stimuli, ordered from higher to lower BMI. Avatars represent a range of body types for both genders.

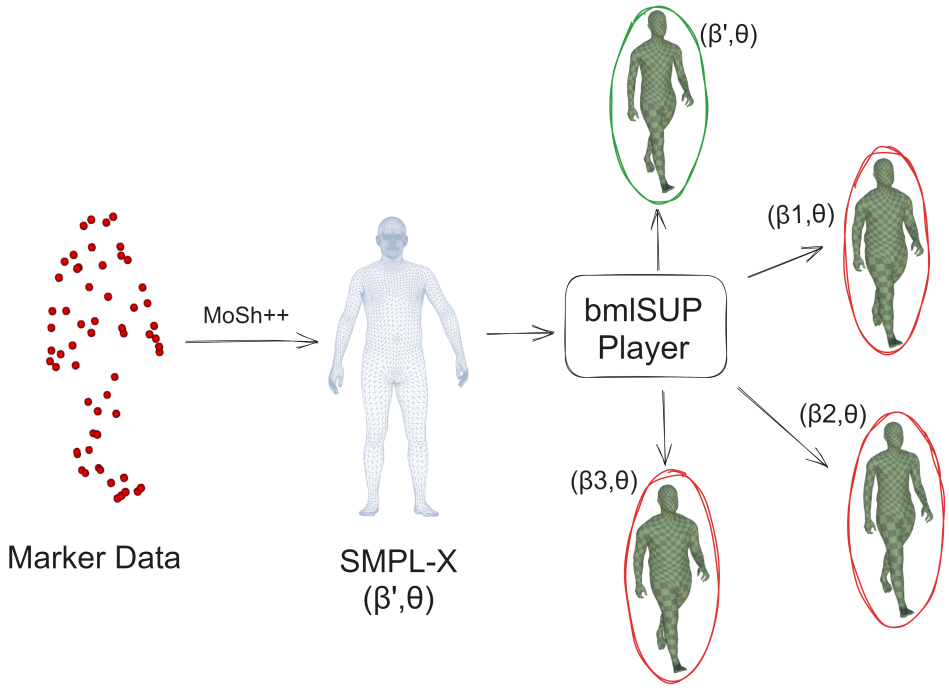


Figure 3.2: Experiment Stimulus Generation Process: shape (β) and pose (θ) parameters are extracted from marker data using MoSh++ (green: consistent motion; red: inconsistent motion).

3.3.1 Stimuli

The stimuli were created in a similar fashion to previous studies [KMH+19, RSM+22]. We used the MoSh++ method [MGT+19b] and statistical body model SMPL-X[PCG+19, RTB17] to generate the body shape and motion for each actor. Using a data-driven probabilistic model, MoSh++ generates a 3D surface by creating a triangulated mesh of 10,777 triangles. This model accounts for inter-individual variance in body shape as well as pose-dependent deformations. The body surface gets reconstructed with a high level of accuracy, with a mean distance between the true and estimated surfaces of less than 1cm. The outputs from MoSh++ were .npz files which were further converted to c# readable format using the tool provided in the bmlSUP Unity Player [BTT21], which we also used to play animations in Unity™. Each file contains the body shape parameters and animated body poses of the actors.

The *consistent* stimuli were created directly using the original file with the body shape and pose of the actor, as shown in Figure 3.2 (green boundary). The *inconsistent* stimuli were created by exchanging the shape values of one actor with the shape values of another actor. Thus, the motion was kept the same for inconsistent stimuli while changing the body shape used in the animation, as shown in Figure

3.2 (red boundary). The animation clips were rendered in the Unity™ game engine and exported as video clips. The scene consisted of a treadmill model on the floor of an empty environment, with a simple checkerboard texture applied to the character's mesh. As our main aim was to generate and analyse realistic body shapes and motions, we wished to avoid any potential uncanny valley effect that could arise from using highly detailed skin/body textures [SWH18, MMK12].

3.4 Perceptual Experiment

3.4.1 Participants

Sixteen volunteers from a variety of backgrounds participated in the online survey. Of these, 8 identified as male, 7 as female, and 1 preferred not to disclose their gender. Participants were grouped into age categories: 7 were aged 18–29, 8 were aged 30–45, and 1 chose not to disclose their age. In terms of relevant experience with computer graphics systems, 8 participants reported having expert-level experience, 5 indicated having some experience (e.g., as users such as gamers), and 2 reported little or no prior experience. None of the participants had previously taken part in the motion capture sessions used to create the stimuli. Recruitment was carried out through email lists and social media. The experiment was conducted using Qualtrics, an online platform that provides fully anonymised results. Prior to the start of the experiment, participants provided informed consent by clicking a button and could also input optional demographic details.

3.4.2 Procedure

The study employed a mixed design to investigate the effects of the between-subjects factor *SEX* and the within-subjects independent variables GAIT, BODY, and MOTION on consistency ratings. Participants were assigned in counterbalanced order to either the male or female groups to ensure that an equal number of responses for each sex was recorded. Within each sex group, participants viewed animations depicting combinations of 2 GAIT (Walk or Jog) X 4 BODY (high BMI: Hi1, Hi2, low BMI: Lo1, Lo2) X 4 MOTION (high BMI: H1, H2, low BMI: Lo1, Lo2). Within each sex group, participants viewed two blocks of animations separated by Gait (walk or jog), the order of which was randomised across participants. In other words, approximately half of the participants viewed walking animations first, followed by jogging animations, and vice versa for the other half. Within each block, all of the videos were viewed in random order.



Figure 3.3: Study 1 example stimulus: Body shape and motion were varied across trials to investigate the effect of inconsistencies. The full set of stimulus videos can be accessed via the QR code or at: https://youtu.be/YhKqrFWRZTY.

At the start of each Gait block, examples of consistent and inconsistent animations were shown, and participants were allowed to practice answering. These video clips featured actors with different body mass indices (BMI) walking or jogging, but neither their bodies nor motions were used in the actual experiment. Each Gait block contained 16 distinct videos: 4 BODY x 4 MOTION (1 consistent, 3 inconsistent), each of which was repeated twice. Therefore, each Gait block consisted of 32 questions in total, and each video lasted 4 seconds. The total duration of the experiment was approximately 20 minutes on average.

During the experiment, participants watched video clips and rated the consistency of the motion with the body shape. They entered their ratings on a 5-point Likert scale, ranging from *Very Inconsistent* to *Very Consistent* (see Figure 3.3), then pressed the "Next" button to proceed to the next video clip. The "Next" button only appeared after participants had viewed the clip at least once.

3.5 Results

We conducted a mixed analysis of variance (ANOVA) to investigate the effects of the between-subject factor SEX and the within-subject factors GAIT, BODY, and MOTION on participants' ratings of motion consistency. This analysis helped to explore how these variables interacted and influenced the perceived consistency of virtual human animations. Specifically, it looked at how the different combinations of gait (walking or jogging), body shape (varying BMIs), and motion type affected participants' ratings. Additionally, we look to determine whether male and female stimuli were perceived differently from each other. To further investigate any main effects, we conducted post-hoc analyses using Tukey's tests. This analysis examines the significant differences between all possible pairs of means. Only significant effects with a p-value less than 0.05 are reported here.

The results of the mixed ANOVA, presented in Table 3.2 and Figure 3.4, revealed several significant main effects and interaction effects between the factors GAIT, BODY, MOTION, and SEX.

Table 3.2: Experiment ANOVA: Significant effects for Gait: Walk (W), Jog (J); Body (H1b, H2b, L1b, L2b); Motion (H1m, H2m, L1m, L2m) and Sex (M, F), with effect size (partial η^2), non-centrality (nc) and observed power ($\alpha = 0.05$).

Effect	**SS**	**dof**	**MS**	**F-Test**	p	η^2	**nc**	**power**	**post-hoc (Tukey),** $p < .05$
Main Effects									
GAIT	3.78	1.00	3.78	8.11	<0.01	0.37	8.11	0.75	W > J
BODY	21.35	3.00	7.12	2.89	<0.05	0.17	8.68	0.65	L2b > H2b
Two-way Interaction Effects									
GAIT*SEX	2.82	1.00	2.82	6.05	<0.05	0.30	6.05	0.63	F/W > F/J
BODY*SEX	87.39	3.00	29.13	11.84	< 0.001	0.46	35.52	1.00	F/L1b > F/H1b, M/H2b, M/L1b F/L2b > F/H1b, M/H2b M/H1b > F/H1b, M/H2b, M/L1b
BODY*MOTION	21.48	9.00	2.39	2.59	<0.01	0.16	23.29	0.93	H1b/H1m > H2b/L2m L1b/L1m > H2b/L1m, H2b/L2m L1b/L2m > H2b/L2m L2b/H1m > H2b/L2m L2b/H2m > H2b/L2m L2b/L1m > H2b/L2m L2b/L2m > H2b/L1m, H2b/L2m
Three-way Interaction Effects									
GAIT*BODY*SEX	14.52	3.00	4.84	4.99	< 0.001	0.26	14.98	0.89	see discussion
BODY*MOTION*SEX	53.22	9.00	5.91	6.41	< 0.001	0.31	57.69	1.00	and Figure 3.4

A significant two-way interaction effect between GAIT and SEX was observed, $F(1, 8) = 6.05$, $p = 0.03$, $\eta^2 = 0.30$. This interaction indicates that the relationship between gait type and perceived motion consistency was influenced by the sex of the virtual character. Post-hoc tests revealed that for female virtual characters, walking motions were rated significantly more consistent than jogging motions. However,

this effect was not present for male characters, where walking and jogging motions received similar consistency ratings. These findings suggest that participants were more sensitive to inconsistencies in jogging animations when evaluating female characters.

Given this interaction, the main effect of GAIT must be interpreted with caution. Nevertheless, a significant main effect of GAIT was found, $F(1, 8) = 8.11$, $p < 0.01$, $\eta^2 = 0.37$. On average, participants rated walking motions as more consistent than jogging motions. The effect size ($\eta^2 = 0.37$) indicates a moderate to large effect. However, as noted above, this effect appears to be primarily driven by the ratings for female characters, rather than being consistent across all conditions.

In addition to the GAIT*SEX interaction, a significant three-way interaction between GAIT, BODY, and SEX was also found, $F(3, 8) = 4.99$, $p < 0.001$, $\eta^2 = 0.26$. This suggests that the effects of gait and body shape on perceived motion consistency were further modulated by the sex of the virtual character. Post-hoc analysis indicated that female characters with low BMI body types received higher consistency ratings across both walking and jogging motions. In contrast, body shape had less influence on consistency ratings for male characters. These results further emphasize that participants paid closer attention to the combination of gait and body appearance in female virtual characters.

The analysis also revealed a significant main effect of BODY ($F(3, 24) = 2.89, p = 0.05, \eta^2 = 0.17$), demonstrating that participants' ratings of motion consistency were influenced by the body shape of the virtual character. Specifically, participants rated animations of the low BMI body (Lo1) significantly higher in terms of motion consistency than the high BMI body (Hi1), as confirmed by post-hoc analysis. This finding indicates a preference for slimmer body types in terms of perceived consistency, possibly because slimmer bodies are associated with more fluid and lighter movement, whereas heavier body types may be expected to move differently. The effect size ($\eta^2 = 0.17$) suggests that the influence of body shape on perceived motion consistency, while significant, was smaller than the effects of gait or the interactions with sex.

A two-way interaction effect between BODY and SEX was also observed ($F(3, 8) = 11.84, p < 0.001, \eta^2 = 0.46$), which indicates that the influence of body shape on perceived consistency varied depending on the virtual character's sex. Post-hoc analysis revealed that while female virtual characters with low BMI bodies were consistently rated higher, male virtual characters with high BMI bodies (Hi1) received the highest ratings overall. This suggests that participants perceived male and female virtual characters differently in terms of their body shape. And there are some underlying factors associated with their relationship.

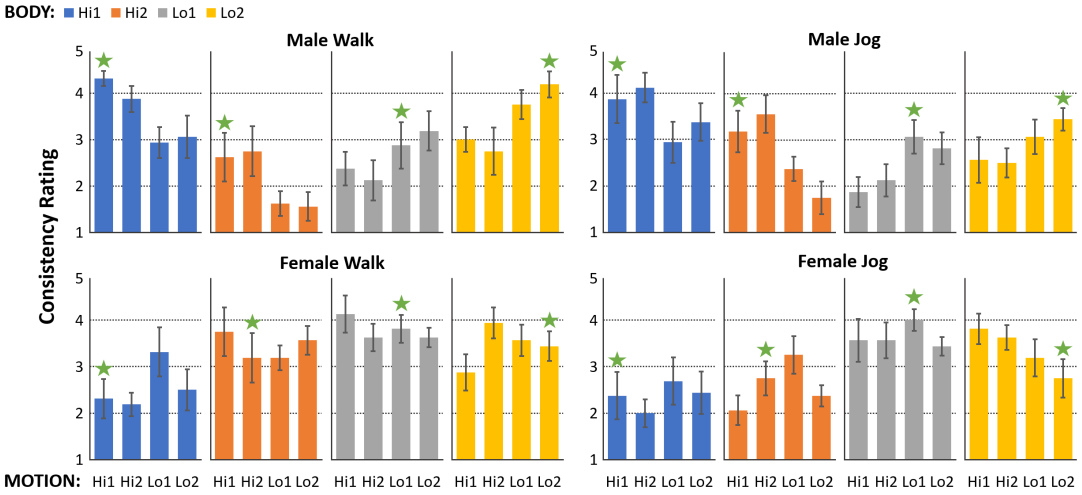


Figure 3.4: Experiment Results: Mean rating scores for varying BMIs. Higher scores indicate more favourable ratings. Consistent motions, i.e., where the original body shape and motion match, are marked with a star.

Additionally, a significant interaction effect of BODY and MOTION ($F(9, 24) = 2.59, p = 0.01, \eta^2 = 0.16$) was found, indicating that participants' ratings were influenced by the interaction between the body shape and the specific motion being performed. In general, participants preferred when the body shape and motion were similar in terms of BMI, with higher consistency ratings for combinations of similar BMI bodies and motions. However, an exception was found for the low BMI female body (Lo2), which received higher ratings for motions associated with all bodies, including the high BMI (Hi1) motion. This finding may suggest that participants found certain low BMI bodies to be more adaptable to a range of motion styles, or it could reflect a bias toward thinner body types.

Finally, the analysis revealed a three-way interaction effect between BODY, MOTION, and SEX ($F(9, 24) = 6.41, p < 0.001, \eta^2 = 0.31$). This interaction effect suggests that the influence of body shape and motion on consistency ratings was further modulated by the sex of the virtual character. Post-hoc analysis indicated that female bodies with low BMIs were rated higher regardless of the motion being performed, while male bodies showed a similar but weaker trend, with higher ratings observed primarily for high BMI bodies. These findings highlight the complex ways in which body shape, motion, and sex interact to influence participants' perceptions of motion consistency.

Overall, the results of the mixed ANOVA provide valuable insights into the factors that affect the perception of motion consistency in virtual characters. The significant effects of GAIT, BODY, MOTION, and SEX, along with their interactions, demonstrate the importance of considering multiple factors when designing

virtual humans for realistic animations. This is further validated by the moderate to large effect sizes (η^2) across the various analyses.

3.6 Discussion

Based on previous research by Russell et al. [RSM+22], we initially hypothesised that participants would give higher ratings to animations where the original body shape and motion were consistent. We expected that this consistency would be more natural-looking or realistic to participants, leading to higher ratings. However, the results from the experiment did not fully support this hypothesis. As shown in Figure 3.4, only a minority of participants rated these consistent, or matched, motions higher than those where the body shape and motion were mismatched. Nonetheless, the two-way interaction between BODY and MOTION indicated that ratings were influenced by both body shape (low or high BMI) and the type of motion (consistent or inconsistent). The post-hoc analysis revealed that participants tended to favour body shapes within the same BMI group as the actor performing the motion. In other words, if a consistent motion was performed by an actor with a low BMI, participants were more likely to rate low BMI body shapes more highly than high BMI body shapes. This finding underscores the importance of consistency between body shape and motion. When the body shape aligned with the motion, participants seemed to find the animation more believable, suggesting that consistency between physical appearance and movement enhances the perception of coherence or plausibility.

These results also align with previous work on motion perception. Toje [Tro02] has shown that human observers are highly receptive to even subtle inconsistencies in movement, particularly when it comes to familiar actions like walking. It could be concluded from the results that when there is a mismatch between the expected body shape and the motion associated with it, this inconsistency may become more apparent and divert from the overall perception of consistency.

The main effect of GAIT (walking or jogging) revealed that participants gave different consistency ratings to walking and jogging motions. Specifically, participants rated walking motions as more consistent than jogging motions, regardless of body shape or motion consistency. This finding suggests that walking, as a more common movement, is perceived differently than jogging. This also suggests that jogging, which involves more complex dynamics, is harder to judge and form abrupt perceptions. This also reflects the varying levels of expectations of participants from different kinds of locomotion. Additionally, the results revealed a significant effect of SEX (male or female) on participant ratings. Participants appeared to easily

distinguish between different sexes, and their ratings reflected varying perceptions associated with the category. Another possible explanation is the difference in motion style found in different sexes due to their body dynamics. These results are consistent with research by Troje [Tro02], and Niay et al. [NOZ+20], both of which emphasise the significant role that dynamic information, such as motion style and gait, plays in forming human perception.

The influence of the factor SEX on participant ratings was also apparent, particularly in the case of female body shapes with low BMIs (Lo1 and Lo2), which consistently received higher consistency scores. This suggests that certain physical characteristics—particularly in female virtual characters—may play a role in shaping how natural or coherent their motion is perceived to be. While our study did not directly assess perceived attractiveness, previous research by Thaler et al.[TBM+20] and Troje[Tro03] has demonstrated that sexually dimorphic walking patterns (e.g., hip sway in females) contribute to impressions of attractiveness in biological motion. Our findings may align with this line of research, suggesting that physical attributes commonly associated with feminine walking patterns could enhance the perceived naturalness or coherence of the animation, even when participants were not explicitly evaluating attractiveness.

Our findings also suggest that different aspects of motion perception may be influenced by distinct factors, some of which we have only begun to explore in this study. It is important to note that the physical characteristics represented in our virtual characters, such as body mass and height, are closely linked to fundamental physical properties like the effects of gravity, velocity, and momentum. Heavier characters, for example, may move with more force and momentum, while lighter characters might exhibit more fluid and agile movements. These physical characteristics, in turn, affect participants' perceptions of the motion's consistency and believability. These properties influence how virtual humans move and interact with their environments in ways that are rooted in the laws of physics. For instance, while we focused primarily on body shape as the defining characteristic of BMI, there are numerous other biological and physical factors that likely play a role in shaping how motion is perceived. These could include muscle activation, joint flexibility, and the distribution of body fat. All of these factors influence how real human bodies move but are difficult to accurately simulate in virtual characters. Future research could explore these additional factors to gain a more nuanced understanding of how various elements of physicality impact the perception of virtual human motion.

Our study's results suggest that consistency between body shape and motion influences user ratings. At the same time, other factors also play an important role, like gait and the physical attributes of the virtual character, especially for female

characters. While our study did not assess perceptions of attractiveness directly, the findings point to the influence of both biomechanical and perceptual factors in shaping judgments of virtual human motion. Future research should explore a broader range of contributing variables. Incorporating more detailed physical and biological modeling could help clarify how virtual characters can be designed to move in ways that are perceived as both physically plausible and visually coherent.

While our findings did not fully support our first hypothesis, they supported our argument that consistency between original motion and body shape within the same BMI group has been rated more favourably. They also provide valuable insights into the ways that body shape, motion, and sex interact to affect participants' perceptions of virtual human animations.

3.7 Chapter Conclusion

This chapter investigated how the visual appearance of avatars—specifically body shape defined by BMI—affects user perception of motion in non-immersive virtual scenarios. The study was designed to explore whether users perceive motion as more consistent or plausible when there is alignment between an avatar's body shape and the motion it performs, particularly when the motion originates from an actor with a similar BMI.

The results showed that perceived consistency was influenced by the interaction between motion type, body shape, gait (walking vs. jogging), and sex of the avatar. Although only a minority of participants consistently rated shape-motion matched stimuli as more natural, the data revealed that motions were rated more favourably when applied to avatars with body shapes within the same BMI category as the original motion source. This indicates that physical plausibility—in this case, the visual alignment between motion dynamics and avatar form—can play a subtle yet measurable role in shaping motion perception.

Female avatars with low BMI received higher ratings overall, which may reflect both biomechanical coherence and implicit perceptual biases related to body shape and sex. These findings are consistent with prior research suggesting that sexually dimorphic gait patterns influence perception, and that even small deviations in expected body-motion pairings can affect user judgments of naturalness.

Gait type also influenced ratings, with walking generally perceived as more consistent than jogging, possibly due to its familiarity and lower dynamic complexity. These results highlight the role of motion type in shaping user expectations and perceptions.

Overall, this study suggests that body shape and motion are not perceived in

isolation—users form judgments based on how well these elements align with each other and with their own expectations of physical plausibility. While our results support the idea that shape-motion consistency enhances perception, they also underscore that such perception is shaped by multiple interacting factors, including sex, gait type, and possibly perceived attractiveness.

It is worth noting, however, that these findings are based solely on observer judgments of motion consistency and not from immersive VR contexts. In the following chapter, the case study on self representation in VR will address this exploration.

Chapter 4

Case Study I: Effect of Self Representation on Virtual Presence

4.1 Introduction

In the previous chapter, we explored how avatar body shape and motion consistency influence user perception in third-person, observational settings. This case study shifts focus from third-person perception to first-person experience, investigating how the representation of one's own avatar affects presence and embodiment in immersive virtual reality.

The use of virtual reality has expanded beyond its initial applications and is now widely used in various industries. However, the principles of realism and naturalness, which work for traditional viewers, do not always apply in immersive VR environments. The effectiveness of VR depends mainly on the users' sense of virtual presence, a complex concept that includes multiple elements. Virtual presence, the feeling of *"being there"* in a virtual environment, is a key psychological phenomenon that researchers have commonly studied in relation to VR technology [BAP+17, HKM+94, SUC95]. This concept has been widely explored in fields such as education [RMFW20], medicine [YTK+21], tourism [FJD22], and rehabilitation [RNC18].

In this study, our objective was to understand the impact of self-avatar body representation on the feeling of presence in VR. Specifically, we aimed to explore how the appearance of self-avatar influences the feeling of embodiment, which is an aspect of virtual presence. The concept of virtual presence refers to the sensation of "being there" in a virtual environment, while embodiment encompasses the user's sense of control and ownership over their avatar. Prior research indicates that enhanced presence and embodiment lead to greater user engagement and satisfaction

in VR [SSV16]. In this pilot study, we explored five different levels of avatar representation - *Full body Avatar* (FA), *Hands Only* (HO), *Arms Only* (AO), *Finger tracked Hands* (FH), and *No Avatar*(NA). We hypothesised that more complete representations of the self would yield higher levels of perceived embodiment, aligning with the notion that visual fidelity in avatars creates a deeper connection to the virtual environment. We assessed the participants' experiences through self-reported measures. This provided insights into how different self-avatar designs can affect key psychological aspects.

Similar to our previous study on body shape and motion perception, this study further explored how avatar design affects user experience in VR. The findings from this study helped in the broader understanding of effective avatar representation, emphasising the importance of thoughtful design in creating engaging virtual environments. In this chapter, we will first examine the concepts of presence and embodiment in VR and how they can be measured. Next, we will outline the experiment setup and procedure. Finally, we will discuss the results of the experiments and present our conclusions.

4.2 Presence in VR

Presence in the context of virtual reality refers to the subjective feeling of "being there" or existing within a computer-generated environment as if the virtual space is real. The sensation is often referred to as *spatial presence* or *place illusion*, with Slater et al. [SSV16] defining it as the feeling of being physically located in the virtual environment. For this work, we adopted a similar definition proposed by Slater [Sla03], which states that presence is when people behave in a virtual environment in a way that is similar to what their behaviour would have been in a similar real-life situation. There are various definitions and theoretical models of presence in virtual environments. Lombard and Jones [LJ15] described presence as the perceptual illusion of non-mediation, where users feel directly engaged with the virtual world rather than with the medium itself. Sheridan [S$^+$92] emphasized sensory fidelity and real-time interaction as core contributors to presence, focusing on technological immersion. Schubert et al. [Sch09] proposed a multidimensional framework that separates presence into spatial presence, involvement, and realness. Similarly, Wirth et al. [WHB$^+$07] introduced a process model that considers user expectations and cognitive processing as central to the experience of presence. These perspectives reflect both technological and psychological interpretations, suggesting that presence is not a singular construct but a complex interplay of immersion, attention, and user expectations.

The concept of embodiment is often discussed in relation to presence in virtual environments [Bio97]. Since we are more interested in the avatar body and its self-representation, it is more rational to lean towards the notion of embodiment. Embodiment, as explored by Kilteni et al. [KGS12], refers to the feeling of being inside, controlling, and having ownership of a virtual body. Understanding embodiment was essential in this work, as it directly relates to how users experience presence in VR environments, especially when considering self-representation through avatars [SSC10]. The sense of embodiment is broken down into three core components: self-location, agency, and body ownership [KGS12]. Each of these aspects plays a critical role in determining how immersive and natural the virtual experience feels.

Components of Embodiment

1. Sense of Self-Location: The sense of self-location refers to the subjective feeling of being physically situated within a body, whether that body is biological or a virtual avatar. In VR, users often experience a shift in self-location when they feel immersed within an avatar's body. Maselli and Slater [MS13] showed that embodiment illusion could be produced without full-body motion tracking by simply collocating a virtual avatar with the participant's perspective. Nakul et al. [NODLL20] conducted an experiment where they applied tactile stimulation to participants' backs and synchronised this stimulation with the visual display of an avatar's back. Their results indicated that participants were more likely to identify with the avatar when the stimulation was synchronised, highlighting the importance of multisensory integration in achieving a strong sense of self-location.

2. Sense of Agency: Agency in VR refers to the sense of control a user feels over their actions within the virtual environment. In this context, the agency is experienced when users feel that their movements in the real world are mirrored by their virtual avatars, creating a sense of intentional action within the virtual space [BM09]. Studies have shown that the synchronisation of movement between a user's real body and their virtual avatar can enhance the sense of agency [KS14]. However, discrepancies between visual feedback and actual movement can negatively impact this sense of control [BWF02, FFG+01, SY05]. For instance, when the virtual body's movements lag or fail to accurately replicate the user's real-world actions, the feeling of agency diminishes, reducing the overall sense of presence.

3. Sense of Body Ownership: Body ownership refers to the sensation that the virtual body or avatar is one's own [Gal00]. This sense of ownership is largely

driven by multisensory integration, where visual, tactile, and proprioceptive stimuli are aligned to create the illusion that an artificial body part belongs to the user. Tsakiris [Tsa10] discussed how the combination of bottom-up sensory inputs and top-down cognitive processes can induce a strong sense of body ownership. When the virtual body behaves as expected, and users receive appropriate sensory feedback, they are more likely to feel that the avatar is an extension of their own body, thereby enhancing the sense of embodiment.

Self-Representation in VR

The appearance and level of representation of avatars in VR can significantly affect the user's sense of presence and embodiment. Several studies have examined how different aspects of self-representation influence user experiences in virtual environments. For example, Wirth et al. [WGP+21] analysed how avatar appearance, environment awareness, and camera perspective impacted gait parameters and found that while avatar appearance affected the perception of presence, it did not significantly alter gait performance. This suggests that visual representation can shape how users feel about being in a virtual environment without necessarily influencing their physical actions.

Bailenson et al. [BBG08] explored how participants interacted with virtual representations of themselves and others, finding that users were more consistent in their behaviours when represented by avatars that resembled their real selves. Gonçalves et al. [GMB+22] found that partial body representations (e.g., hands and feet without a full-body avatar) negatively impacted the sense of embodiment. Zhang et al. [ZHT+22] also investigated how the realism of virtual hands affected user perception and task performance, highlighting the critical role of realistic self-representation in forming a sense of presence and embodiment. Lin and Jörg have made significant contributions to this line of research by studying how visual and sensory characteristics of avatar hands affect embodiment. In one study, they examined how model realism, rendering style, and individual sensitivity to pain impacted the virtual hand illusion [LJ16]. They found that most participants experienced some level of embodiment regardless of model realism, but anthropomorphic hands elicited the strongest effects, whereas blocky, non-human shapes were least effective. In another study, they explored how mismatches between a user's actual hand size and their avatar's hand size influenced embodiment and user experience [LNA+19]. While hand size variation did not significantly alter the sense of agency or illusion intensity, participants reported stronger feelings of ownership and realism when using a glove-like avatar hand.

Overall, these studies suggest that avatar appearance and the extent of body representation play a vital role in shaping user experiences in VR. However, there is still a need for further exploration of how different levels of self-representation affect embodiment and presence in task-specific contexts.

Subjective Measures for Presence

The measurement of virtual presence is complex as it is an internal and inherently subjective experience. As Slater [Sla09] noted, the experience of presence is not directly measurable, and it cannot be confirmed that all people experience it the same way. As reviewed by Felton & Jackson [FJ22], there are various measurement categories for presence, and subjective measurement is one of them. The subjective measure of virtual presence relies on conscious feedback explicitly requested by the user. This could involve post-immersion questionnaires and self-reported continuous ratings during the virtual presence experience.

Witmer and Singer [WS98] developed a 32-item presence questionnaire with three subscales: involvement/control, natural, and interface quality. In another approach, the Slater-Usoh-Steed (SUS) questionnaire [SSMM98, UCAS00], focuses on themes such as the sense of being in the virtual environment, the extent to which it becomes the dominant reality, and how well it's remembered as a *place*.

Schubert et al. [SFR01] in IPQ (Igroup Presence Questionnaire) describe Spatial Presence as a special mental mode of virtual space, and their questionnaire includes 14 items to assess this aspect. Engagement, another crucial factor, can be measured using the Game Engagement Questionnaire (GEQ) by Brockmyer et al. [BFC+09]. Hartmann et al. [HWS+15] introduced the Spatial Presence Experience Scale (SPES), a shorter version of the combined previous presence questionnaire [EMFM21]. These questionnaires consider a broader range of feelings related to immersion.

It's worth noting that most surveys in VR are typically conducted after participants leave the VR environment. Schwind et al. [SKHH19] explored the effect of completing presence questionnaires directly in VR. While they found no significant difference in mean scores between real-world and VR questionnaires, VR increased the consistency of variance.

Since we mentioned the sense of embodiment aspect of Virtual presence aligns more with our research question, we explore some widely used embodiment questionnaires. Gonzalez-Franco & Peck [GFP18] reviewed more than 30 embodiment research studies that use questionnaires for embodiment measurement. In the review, they categorized the studies since 1998 when the rubber hand illusion study

was published [BC98]. Recently, Roth & Latoschik [RL20] presented a Virtual Embodiment Questionnaire (VEQ) which is used in most embodiment studies now. Similar to Eubanks et al. [EMFM21], we used a combination of two questionnaires to measure virtual presence in our specific task. We used agency and ownership questions from VEQ [RL20] and self-location questions from SPES Questionnaire [HWS+15].

Our study aimed to assess how self-representation impacts users' virtual reality (VR) presence within virtual workstations. With VR increasingly used in ergonomic risk assessment and occupational health and safety, especially in tasks like manual material handling, we focused on the arms and hands. We examined five levels of self-representation, from no avatar and just visual representation of the physical VR controllers to finger-tracked hands mimicking the user's hands directly, to a full-body avatar using robust motion capture tracking. Our research question was: To what extent does virtual presence, as quantified through a user's experience of body ownership, agency, and self-location vary across different levels of self-representation in virtual workstations? We hypothesized significant differences in virtual presence across these levels of self-representation (H1), while the null hypothesis posited no significant differences (H0).

The novelty of our study lies in exploring the effects of five levels of self-representation on presence and embodiment during manual material handling tasks in virtual workstations, a topic not previously investigated to our knowledge. By analyzing subjective data from participants, we gained valuable understandings into the influence of self-representation on virtual presence. Additionally, we aimed to provide recommendations for optimizing VR applications, particularly focusing on avatar design to enhance presence within virtual workstations.

4.3 Methods

In this study, we simulated common material handling tasks in virtual reality (VR), and participants performed these tasks under various self-representation conditions. We employed a combination of motion capture and inverse kinematics techniques to visually replicate human movements within the VR environment.

4.3.1 Equipment

The virtual environment for our study was developed using the Unity 3D game engine. We utilized a Meta Quest 2 headset for the VR head-mounted display (HMD), which was tethered to a computer. The Quest 2 employs inside-out cameras to of-

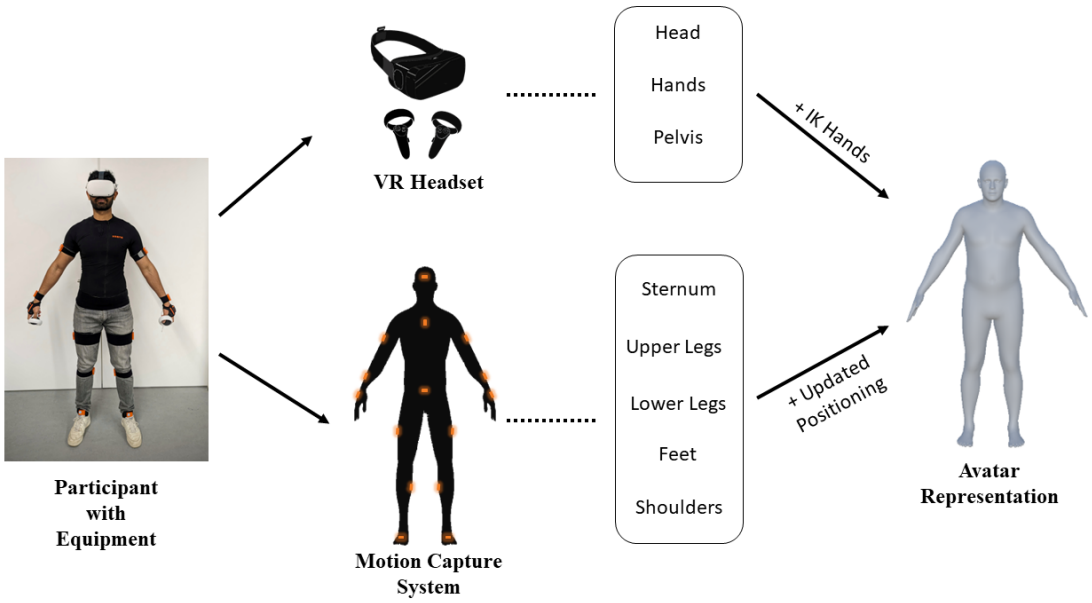


Figure 4.1: Illustration of the VR setup and motion capture system. The participant is equipped with an HMD with controllers and motion sensors that capture body movements. The motion capture data is updated based on the HMD position and used to generate an avatar representation with IK for hand movements.

fer 6 degrees of freedom for head mounted display tracking. Participants could use either tracked hand controllers with physical buttons or hand tracking and gesture detection for user input. This technology ensured precise and responsive tracking of participants' head and hand movements within the virtual environment. To maintain a high frame rate and prevent motion sickness, we utilized a VR capable gaming PC (Intel Core i7-7700K CPU, NVIDIA GeForce GTX 2080 GPU, and 32 GB of RAM). For motion capture, we employed the MVN Awinda system by Xsens. This system comprises 17 wireless IMU (Inertial Measurement Unit) units placed on different body segments to measure orientation in real-time. The Figure 4.1 illustrates the VR setup and motion capture system.

4.3.2 Avatar Conditions Stimuli

To ensure consistency throughout the experiment and with different avatar conditions, we used the same virtual scene for all conditions. We used an SMPL-X Neutral body model [PCG+19] to represent a full-body virtual avatar of the participants. We measured the body weight and height of each participant and used the blender plugin for SMPL-X to estimate the body shape parameters of the neutral body model. Since each avatar condition was distinct, we generated stimuli for each

condition in a customised manner to match the specific requirements of that condition, as can be seen in Figure 4.2.

No Avatar (NA)
In the *No Avatar* condition, participants could only see the two virtual handheld controllers without any avatar representation. The default 3D models of the Meta Quest 2 controllers were used in this condition, as shown in Figure 4.2.

Hands Only (HO)
In this condition, participants could only see their animated hands. We applied a transparent texture to the entire body except for the hands, which were rendered in grey, as shown in Figure 4.2. The hand movements were animated using the controller buttons with the animations triggered by button presses.

Arms Only (AO)
In this condition, participants could see their full arms. The body, except for the arms, was rendered with a transparent texture, while the arms were textured in grey, as shown in Figure 4.2. The hand animations were the same as in the previous condition, controlled by button presses on the controller. The arms were animated using an IK solver, with the hand/controller position acting as the IK target.

Full body Avatar (FA)
In this condition, the full body of the participant was visible. A consistent grey texture was applied to the entire SMPL body [PCG+19]. The hands were animated and controlled via button presses, similar to the previous conditions.

Finger-tracked Hands (FH)
In this condition, only the participant's hands were visible, but they were tracked using a finger-tracking feature. Instead of using controllers, we applied the OpenXR interaction toolkit, allowing participants to use hand gestures to interact with the virtual environment. The grab or pinch gestures triggered actions, much like pressing a button on the controller. Since participants already wore motion-tracking gloves, we provided blue gloves to wear over them. This ensured the uniform hand colour and improved camera tracking for both hands and fingers during interactions.

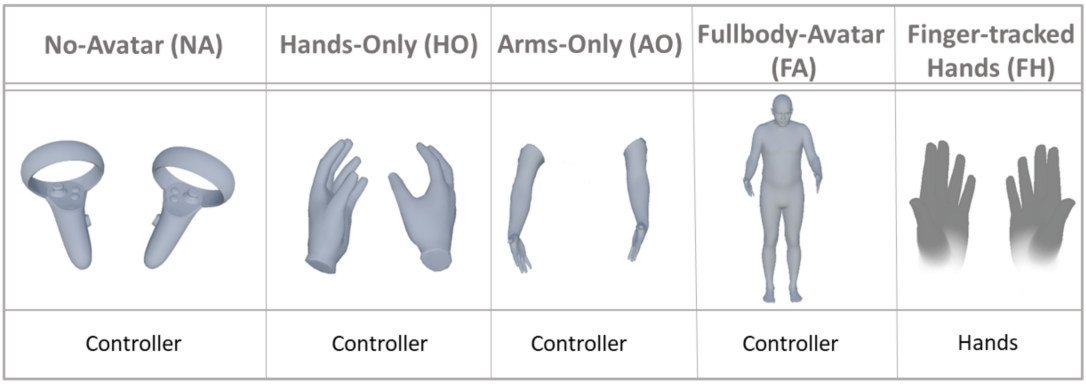


Figure 4.2: Visualisation of different avatar conditions and control types. Participants interacted with either controllers or finger-tracking, depending on the condition.

4.3.3 Full Body Tracking in VR

To maintain consistency across the experiment and across different avatar conditions, we utilized the same virtual scene depicted in Figure 4.3 for each trial. For avatar representation, we employed the SMPL-X Neutral body model [PCG+19] to create a full-body virtual avatar for each participant. The 3D avatar body models were rendered with a simple grey texture to prioritize consistency over detailed skin textures and to avoid the uncanny valley effect [MMK12, SWH18]. Additionally, we measured each participant's weight and height and utilized the Blender plugin for SMPL-X to estimate the body shape parameters of the neutral body model [PCG+19].

To enable full-body tracking of participants throughout the experiment, we integrated both the MVN Awinda motion capture system and the built-in positional tracking of the VR headset (see Figure 4.1). This system utilized orientation information from the 17 IMU sensors affixed to the body to map body motion to a skeleton in real-time, accommodating different body dimensions as input. Consequently, the resulting skeleton was unique to each individual, reflecting their specific body and limb sizes.

Utilizing the MVN Analyze software's data streaming protocol, the motion data of the entire skeleton was streamed into the Unity game engine for integration into our virtual scene. The integration operated at a frame rate of 60 fps, which is typically sufficient for VR applications. One limitation of IMU-based motion capture is the lack of an absolute positioning reference frame, leading to potential drift in the system's position over time. To address this, we leveraged the head-mounted

display's (HMD) accurate tracking of head movements with 6 degrees of freedom (DOF), capturing both rotation and translation. We used these values as the ground truth to correct the position of the motion capture data.

Alignment of the motion-captured skeleton with the HMD position was achieved by matching the center of the eyes from the skeleton with the virtual camera position in Unity. Although a small amount of latency between the MVN data and the VR headset was unnoticeable for the headset's position, it resulted in noticeable errors in rotation, particularly in the pitch and roll axes of head rotation. Consequently, the rotation of the skeleton was aligned with the VR headset by matching the yaw of the head data from MVN Analyze with the forward direction of the VR headset.

Similarly, the absolute positions of the hands from the motion capture system did not perfectly align with the known positions of the hands obtained from the VR controllers. We utilized the position of the VR controllers as the ground truth for hand position and implemented a remapping process to align the hand positions and orientations from MVN Awinda with those of the controllers. Arm positions were then corrected using an inverse kinematics (IK) solution to determine the likely position and rotation of the forearms and elbows. The Two-Bone IK Constraint component from Unity's Animation Rigging package was employed for this purpose. This solution utilized shoulder position data from the MVN motion capture, aligned to the headset as described earlier, and hand position data from the motion controllers as constraints, with additional data of the approximate elbow position from motion capture as inputs to the Two Bone IK component. For the three conditions that employed physical controllers for input but visualized hand models, animations were triggered using controller buttons, with default animations from the OpenXR interaction package.

4.4 Experiment

4.4.1 Tasks in VR

The study comprised a series of manual material handling tasks intended to simulate a typical scenario in a workstation environment. These tasks, resembling activities of daily living (ADLs), hold potential for generalization to other applications. The virtual workstation featured a table with two sizes of boxes – large ($\sim$ 46 cm x 36 cm x 15 cm) and small ($\sim$ 18 cm x 13 cm x 5 cm). These boxes could be positioned in three different locations within the scene: in an open-top container on the table, on a higher shelf to the right, or on a lower shelf to the left, as illustrated in Figure 4.3. To provide participants with a sense of relative position, the starting position was

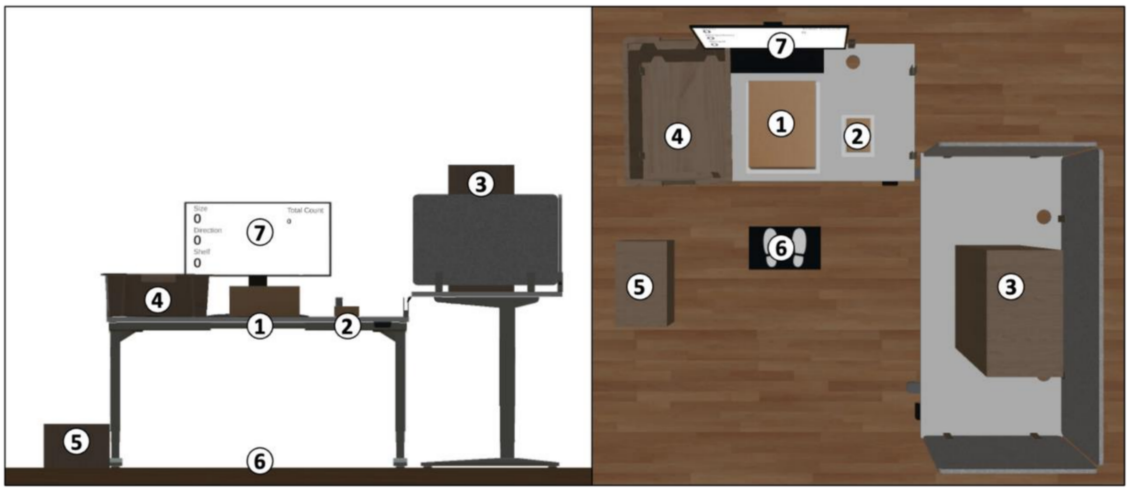


Figure 4.3: Virtual Scene with a workstation: (1) Large box; (2) Small box; (3) High shelf; (4) Open-top container; (5) Low shelf; (6) Starting or Origin Position; (7) Monitor screen

marked on the ground. Additionally, a virtual screen was incorporated into the scene to offer instructions to the user. The virtual layout of the space was meticulously crafted to emulate an office environment, aligning with the physical setting where the study was conducted.

The manual handling tasks in our study encompassed several actions: transferring a box from a table to an open-top container, retrieving a box from the container and placing it back on the table, moving a box from the table to a low shelf, returning a box from the low shelf to the table, placing a box from the table onto a high shelf, and returning a box from the high shelf to the table.

Participants executed these tasks based on instructions displayed on a monitor screen, as depicted in Figure 4.3. The monitor screen presented three parameters simultaneously: size (either large or small box), direction ('to' or 'from'), and shelf location (one of three possibilities: container, low shelf, or high shelf). Initially, participants were provided with boxes positioned on a table in front of them, representing the original positions of the boxes. When the direction indicated 'to', participants were instructed to move the boxes to the designated shelf. Conversely, when the direction was 'from', participants had to retrieve the boxes from the indicated shelf and return them to their original positions on the table. Before the task commenced, participants underwent a practice period to familiarize themselves with the sequential flow and requirements of moving boxes according to the rotating on-screen instructions.

4.4.2 Participants

Nineteen volunteers (5 female, 14 male, aged 18-45) participated in the study. Some demographic attributes of the participants are detailed in Table 4.1. Recruitment was conducted via email lists and social media platforms. Prior to the experiment, all participants reviewed and signed the consent form. Participation in the study was voluntary, with participants retaining the right to withdraw at any time without obligation. Additionally, participants were given the option to provide demographic information. They were also given the opportunity to seek clarification and address any questions or concerns both before and during the experiment. Individuals with recent musculoskeletal injuries (within the past few weeks) and those with a history of motion sickness or sensitivity to virtual reality were excluded from the study. Participants were asked to self-report their prior experience with virtual reality. The majority indicated limited exposure: 9 participants reported having never used virtual reality, and 6 stated they had used it rarely. A smaller number of participants reported moderate to frequent use, with 2 indicating occasional use, 1 reporting frequent use, and 1 reporting very frequent use.

Table 4.1: Demographic attributes of study participants including mean age range, body height, weight, and standard deviations.

Participants	**Count**	**Age Range (Mean)**	**Body Height (cm)**		**Body Weight (kg)**	
			Mean	**Std**	**Mean**	**Std**
Total	19	26–35	173.66	10.51	75.00	15.32
Male	14	26–35	177.38	8.54	81.07	12.77
Female	5	26–35	164.00	9.40	58.00	5.95

4.5 Procedure

The study featured a simple task designed to simulate a common scenario to pick and place objects in a virtual environment. This task movement can also be considered a part of activities of daily living (ADLs). Such movements help adapt and transition to everyday living and allow for better evaluation of the capabilities of the virtual reality tool [LSJC14]. The following were the possible tasks during the experiment:

- Take the box from a table and place it into an open-top container.
- Remove the box from the box and place it back on the table.
- Take the box from the table and place it on a low shelf.
- Remove the box from the low shelf and place it on the table.

- Take the box from the table and place it on a high shelf.
- Remove the box from the high shelf and place it on the table.

The study employed a repeated measures (or within-subjects) design, in which all participants performed the tasks under various avatar representation conditions. The same virtual environment was used for each condition, and each participant repeated the task three times for each of the five avatar representation conditions. The order of tasks was randomised for each participant to mitigate any potential learning effects.

Therefore, each participant performed: (For each Avatar condition) 6 task types x 2 Box types x 3 repetitions of each task/representation = 36 trials for each representation condition. Before each representation condition, participants explored the virtual environment and practised moving boxes. During this time, they were (voice) prompted to perform a series of body motions in order to test the embodiment methods. The same (or very similar) prompts were used for all users and for all conditions. Example prompts included statements such as *"lift your arms straight"*, *"move your arms up and down"*, *"lift your feet one at a time"*, *"lift and move your leg"* or *"look slightly down and rotate your leg"*.

Table 4.2: Embodiment Questionnaire used in the study, featuring questions under the categories of Agency, Body Ownership, and Self-Location. These questions assess participants' subjective experience of embodiment with the virtual avatar across these key dimensions.

ID	Question
SL1	Overall, I felt as if my body/body part was located where I saw the virtual body/body part to be.
SL2	I felt like my body/body part was actually there in the environment.
SL3	I felt like my body/body part appeared in the environment.
AG1	The movements of the virtual body/body part were caused by my movements.
AG2	I felt like my bodily movements occurred in the environment.
AG3	I felt like my body/body part affected the environment.
OW1	Overall, I felt that the virtual body/body part was my own body.
OW2	It seemed as if I might have more than one body.
OW3	I felt like the environment affected my body.

The participants were asked to fill out the questionnaire after each avatar condition. The questionnaire consisted of 9 questions (see Table 4.2). Similar to Eubanks et al. [EMFM21], we combined two questionnaires to measure virtual presence in the specific context of our task. We modelled agency and ownership questions from VEQ [RL20] and self-location questions from SPES Questionnaire [HWS+15].

Questions SL1, SL2 and SL3 were taken from SPES Questionnaire [HWS+15] and were intended to capture spatial presence in the virtual environment. The remaining 6 questions were taken from VEQ [RL20] for agency and ownership. Participants rated each question on a 5-point Likert scale, ranging from "Strongly Disagree" (1) to "Strongly Agree" (5). After each avatar condition, participants were asked for general feedback about their overall experience during an avatar representation.

To complete the questionnaire, participants removed the HMD and filled out the responses on a laptop. Afterwards, they put the HMD back on to proceed with the following condition. Each representation condition took approximately four minutes to complete. Since there were five avatar conditions, we chose a random order of conditions for each participant. In other words, a subset of all possible orders of conditions was used using the Partially balanced Latin Squares method. In total, participants completed 180 trials: 36 trials $\times$ 5 avatar conditions.

4.6 Results

We conducted a repeated measures analysis of variance (ANOVA) to investigate the impact of avatar representation condition on participants' responses. The first independent variable, *avatar representation condition*, included: *No Avatar*, *Controller Hand*, *Controller Full Arm*, *Controller Full Body*, and *Finger Tracked Hand*. Prior to analysis, normality assumptions for each questionnaire item were assessed using the Shapiro–Wilk test. Additionally, we examined the influence of *question category* (Body Ownership, Agency, and Self-Location) as the second independent variable on participants' average ratings. For items where normality assumptions were violated, a non-parametric Friedman test was conducted. Post-hoc analyses for all main and interaction effects were performed using Tukey's test to identify significant differences between all possible pairs of means.

We calculated the overall average score of all questionnaire items for each condition and conducted pairwise comparisons. A significant effect on overall presence was found between the *No Avatar* condition and the *Controller Hand* ($p = .011$), *Controller Full Body* ($p = .035$), and *Finger Tracked Hand* ($p = .005$) conditions. However, the comparison between *No Avatar* and *Controller Full Arm* was not significant ($p = .085$), as shown in Figure 4.4.

To further understand each component of presence, we calculated the average ratings for each question category and conducted a repeated measures ANOVA to evaluate the main effects of condition and category. The analysis revealed a significant effect of condition ($F = 6.80$, $p = .002$, $\eta^2 = .086$), indicating that different avatar conditions significantly influenced participants' overall virtual presence experiences.

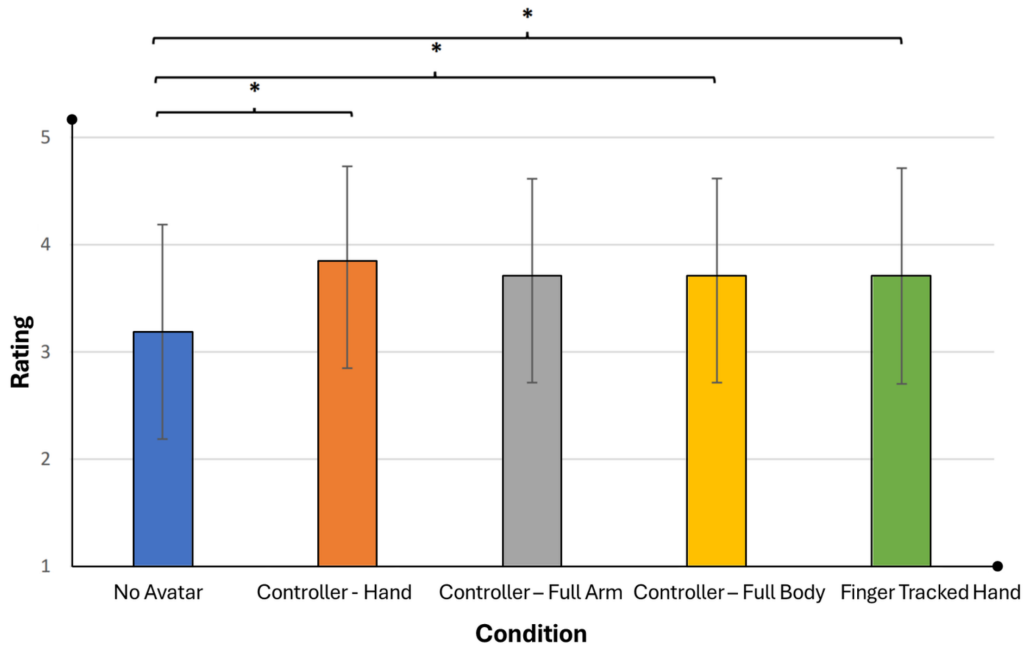


Figure 4.4: Mean scores for each condition with 95% confidence intervals. A p-value $< .05$ was considered statistically significant.

A significant effect of question category was also observed ($F = 64.14$, $p < .001$, $\eta^2 = .349$), suggesting that the question categories elicited varying responses and influenced perceptions of virtual presence. Figure 4.5 presents an overview of the ratings for each category.

4.6.1 Body Ownership

We computed the average scores for the three Body Ownership items (BO1, BO2, BO3) using the formula: ([BO1] + [BO2] + [BO3]) / 3, following the approach by Roth and Latoschik [RL20], and conducted an ANOVA on the resulting scores. The analysis revealed a significant effect of condition on Body Ownership ($F = 2.96$, $p = .04$, $\eta^2 = .082$), indicating that different avatar conditions significantly affected participants' reported sense of ownership. Participants reported the lowest sense of ownership in the *No Avatar* condition (M = 3.37, SD = 1.03), and the highest in both the *Controller Hand* (M = 4.12, SD = 0.61) and *Finger Tracked Hand* (M = 4.12, SD = 0.95) conditions.

Further analysis of individual items within the Body Ownership category revealed a main effect of condition for BO2 ($F = 2.85$, $p < .05$, $\eta^2 = .095$) and BO3 ($F = 3.47$, $p = .021$, $\eta^2 = .116$). Post-hoc comparisons showed significantly lower ratings for the *No Avatar* condition compared to the *Finger Tracked Hand* condi-

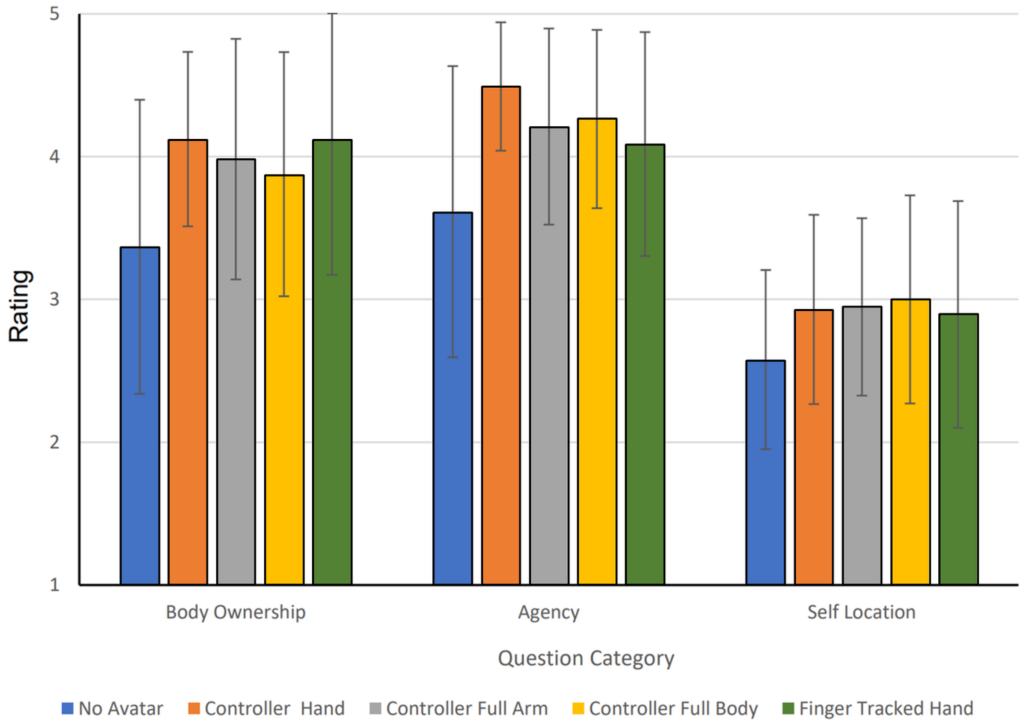


Figure 4.5: Mean scores for individual categories measuring overall virtual presence across different avatar self-representation conditions, with 95% confidence intervals.

tion. Although no main effect was found for BO1 ($F = 2.19$, $p = .091$, $\eta^2 = .084$), a similar trend was observed. These results are illustrated in Figure 4.6.

4.6.2 Agency

We also conducted a combined analysis of the Agency items, which revealed a main effect of condition ($F = 4.80$, $p = .007$, $\eta^2 = .11$). This suggests that participants experienced varying levels of agency depending on the avatar condition. Post-hoc tests indicated that agency ratings were significantly lower in the *No Avatar* condition (M = 3.61, SD = 1.02) than in the *Controller Hand* (M = 4.49, SD = 0.45) and *Controller Full Body* (M = 4.26, SD = 0.62) conditions.

Further item-level analysis showed significant main effects for all three agency items: AG1 ($F = 4.17$, $p = .009$, $\eta^2 = .125$), AG2 ($F = 3.07$, $p = .035$, $\eta^2 = .112$), and AG3 ($F = 3.88$, $p = .021$, $\eta^2 = .11$). Across all three items, the highest agency ratings were observed in the *Controller Hand* condition and the lowest in the *No Avatar* condition. Participants reported a greater sense of initiating movement when they could see only their hands, and less so when no avatar was present. Additionally, the *Controller Full Body* condition elicited higher agency ratings than

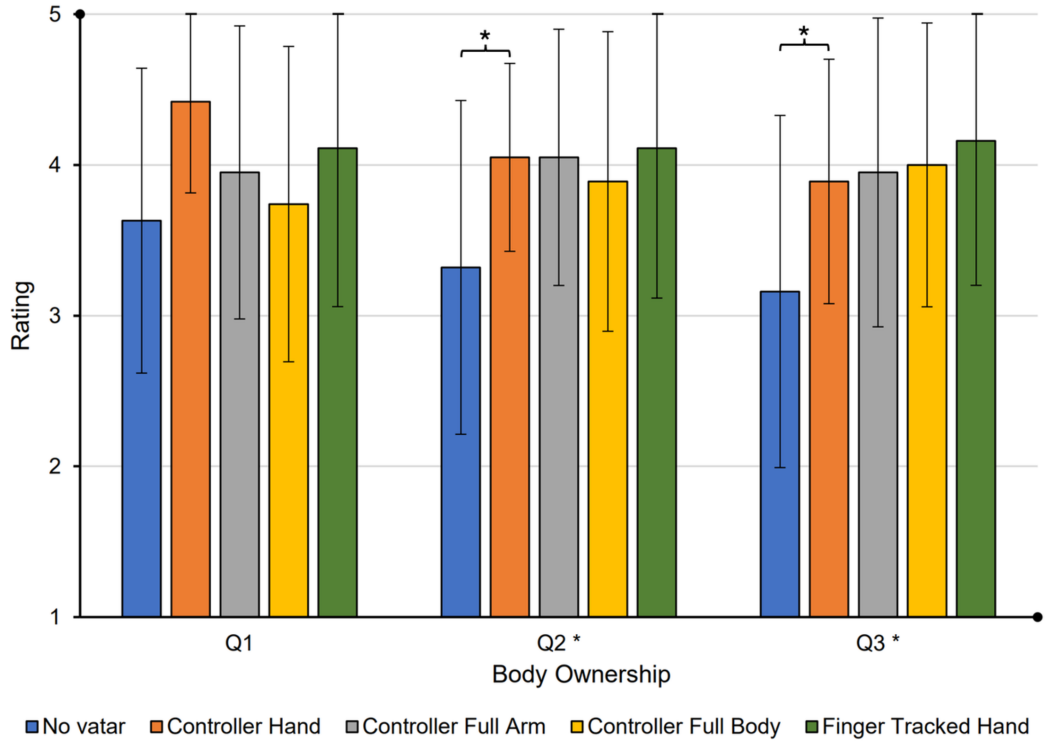


Figure 4.6: Ratings for individual Body Ownership (BO) items. Error bars represent standard deviation. Asterisks (*) denote statistically significant main effects.

the *Controller Full Arm* and *Finger Tracked Hand* conditions. These trends are shown in Figure 4.7.

4.6.3 Self-Location

A combined analysis of the Self-Location items revealed no significant main effect of condition, indicating that participants experienced similar levels of spatial presence across all avatar conditions. Item-level analyses for SL1, SL2, and SL3 also showed no significant effects. As illustrated in Figure 4.8, the average ratings for these items were generally lower compared to those in the Body Ownership and Agency categories. This suggests that participants did not strongly perceive their body as being located within the virtual environment.

4.7 Discussion

The aim of this study was to explore the influence of various avatar representations on participants' virtual presence experiences across multiple dimensions, including body ownership, agency, and self-location within a virtual workstation. Initially, we

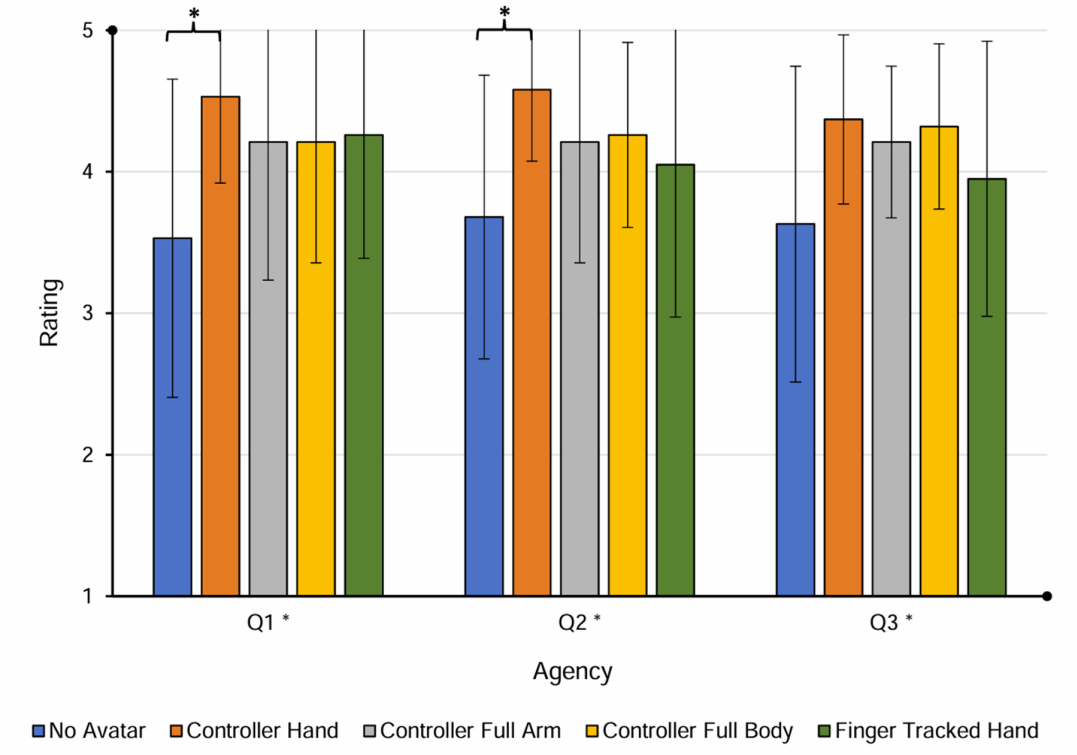


Figure 4.7: Ratings for individual Agency (AG) items. Error bars represent standard deviation. Asterisks (*) denote statistically significant main effects.

hypothesized that all avatar conditions would yield distinct levels of virtual presence. As illustrated in Figures 4.6 and 4.7, the results supported this hypothesis for two dimensions—body ownership and agency. However, the findings varied across avatar conditions, with some performing better for body ownership while others scored higher for agency. Nonetheless, the absence of any avatar representation in the *No Avatar* condition appeared to diminish the virtual presence experience across all measured dimensions (Figure 4.4).

The results revealed significant differences in body ownership ratings across various avatar conditions. Participants reported the least sense of body ownership in the *No Avatar* condition, indicating that the absence of an embodied representation in the virtual scene diminishes users' perception of ownership over their virtual bodies. Conversely, participants experienced the highest sense of body ownership in conditions where they could see their hands or have their fingers tracked, suggesting that visual feedback of one's own body parts enhances the feeling of embodiment. Specifically, participants reported a strong sense of control in the *Controller Hand* condition. The *Finger Tracked Hand* condition also enhanced the sense of ownership by conveying the impression of possessing multiple identifiable body parts in the virtual space.

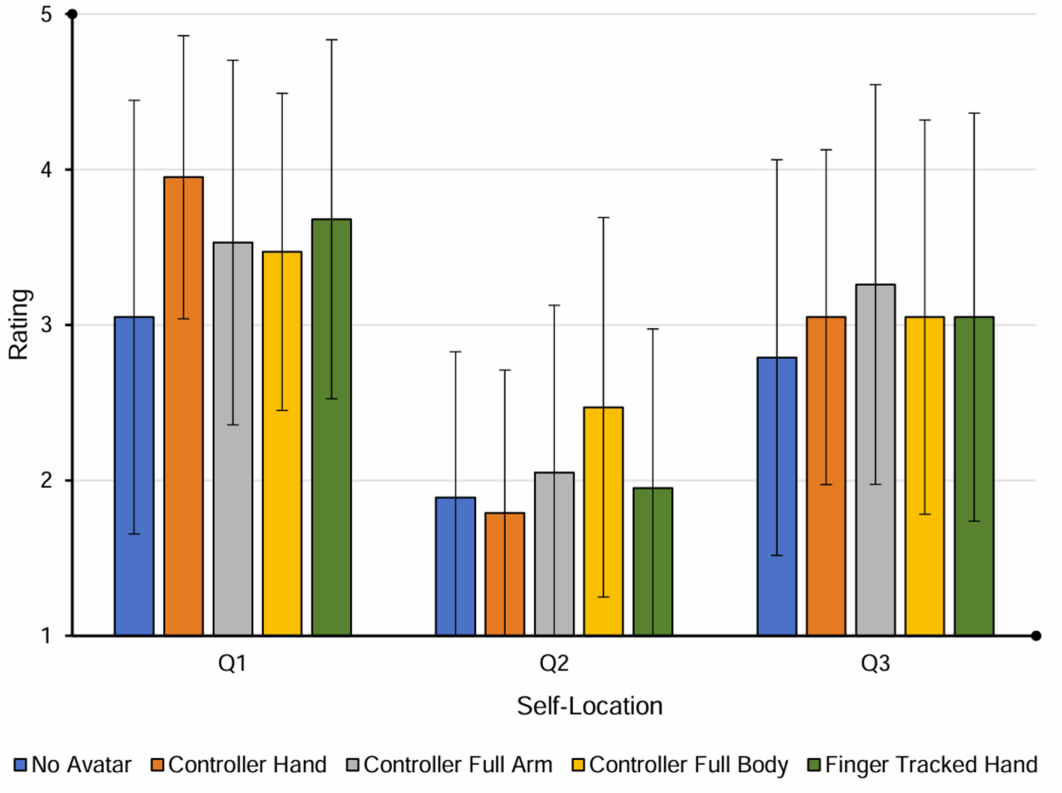


Figure 4.8: Ratings for individual Self-Location (SL) items. Error bars represent standard deviation. No significant main effects were observed.

Contrary to our expectations, the *Controller Full Body* condition did not receive the highest ratings among avatar representations. This discrepancy may be attributed to the partial visibility of certain body parts, which could have influenced participants' perceptions. These findings align with previous research by Brade et al. [BKFK20], which suggests that even partial body representations are adequate for eliciting virtual acceptance and presence.

Our analysis also revealed significant variation in participants' feelings of agency across different avatar conditions. The lowest agency ratings were reported in the *No Avatar* condition, indicating a diminished sense of control over virtual actions in the absence of visible embodiment. In contrast, conditions that included visual representation of hands or a full body showed higher levels of agency. Notably, the *Controller Hand* condition yielded the highest agency ratings, underscoring the importance of hand presence in reinforcing the user's sense of initiating and controlling actions in the virtual space.

Contrary to our hypothesis, no significant differences in self-location ratings were found across avatar conditions. Participants reported similar levels of spatial presence regardless of avatar representation. However, average self-location ratings were

generally lower than those for body ownership and agency. This suggests that participants may not have fully perceived their virtual bodies as occupying space in the virtual environment. These findings point to the need for further investigation into the factors that influence spatial presence and body localization in virtual contexts.

It is worth noting that most participants (15 out of 19) reported having never or rarely used virtual reality. While the range of VR experience—from novice to expert—may have influenced participants' sense of presence, all participants were given a familiarization period before data collection. They had time to acclimate to the virtual workstation and practice the manual handling tasks. There was no strict time limit, and each participant proceeded at their own pace. Furthermore, each participant completed the task three times under each of the five avatar conditions, resulting in 180 total trials per participant. This extensive exposure ensured that participants could provide informed responses across conditions.

Despite the valuable insights gained, this study has several limitations. One limitation concerns the simulation setup. The integration of full-body tracking using a head-mounted display (HMD) and motion capture sometimes produced unnatural motion, which could disrupt immersion and negatively affect ratings in the *Full Body* condition. Improved alignment between motion capture and HMD systems may help mitigate these issues. Similarly, the *Finger Tracked Hand* condition experienced occasional glitches due to the motion capture gloves, which sometimes produced distorted hand shapes and reduced tracking accuracy. These issues may be resolved through better integration of hand tracking technologies.

In addition, the use of inverse kinematics (IK) solvers to animate the arms may not have yielded the most realistic hand motions, potentially influencing participants' perceptions. A larger and more diverse participant sample would also help improve the generalizability of our findings.

Future research could explore multiple directions to expand on these findings. For instance, multimodal embodiment cues—such as body movement, facial expression, or environmental interaction—could be examined for their impact on virtual presence [OBPB22]. Sensory modalities like auditory and haptic feedback may also play significant roles in shaping users' immersive experiences. Longitudinal studies could provide insight into habituation, adaptation, and retention of presence-enhancing features [ONA+21].

Lastly, while our use of self-reported questionnaires aligns with standard practices in virtual presence research, future studies would benefit from integrating both subjective and objective measures. Physiological data (e.g., heart rate, skin conductance) or task performance metrics can complement subjective responses to provide a more comprehensive understanding of presence.

In conclusion, our findings suggest that some form of avatar representation is essential for users to experience virtual presence. Visual embodiment cues significantly influence perceptions of body ownership and agency. Notably, even partial representations such as hands were sufficient to improve these perceptions. Therefore, the additional complexity of implementing full-body or finger-tracked avatars may not always be necessary, particularly for applications such as ergonomic evaluation or safety training. Nonetheless, alternative representations should not be overlooked, as they may further enhance immersion and performance when optimized. Future work should continue to investigate the determinants of virtual presence and refine embodiment strategies for virtual workspaces.

Chapter 5

Body Parameterised Motion Controller

The results from our study discussed in Chapter 3, clearly demonstrated that viewers could detect consistent motion patterns for avatars within the same BMI group. This suggests that an avatar's body characteristics play a significant role in how its motion is perceived. Given this, we face two potential solutions: either we use only specific body shapes for the avatars or capture as many consistent motions as possible for various body types. Currently, the common approach to animating realistic virtual characters involves capturing human motion and retargeting it to fit the character's body model [Men00, GY03, Gle98]. However, these options present computational and financial challenges. An alternative approach is to explore methods that modify motions based on the character's physical properties. This would allow the motion to adapt naturally to the avatar's body shape, eliminating the need for repeatedly capturing motion for every new avatar.

In this chapter, we proposed a framework for generating physics-based characters (avatars) with varying body shapes that closely resemble real humans. This chapter explored the research question **Q2**; listed in Section 1.1. Our goal was to create a diverse population of avatars capable of imitating reference motions, with the focus on basic locomotion, such as walking and jogging. Typically, imitation learning policies work well for a single character [PALvdP18, PMA+21, BCHF19]. However, to incorporate the effects of individual body shape parameters and mass, we needed to simulate a variety of avatars. A key contribution of our framework is developing a policy that can adapt to a wide range of character body types. We referred to this as the Body Parameterised Controller, which adjusts motion based on the character's body parameters.

In this framework, characters were trained to imitate reference motions, with

each character's specific body shape parameters influencing how they mimic the motion. The reference motion acted as a sequence of target poses for the character to follow. We generated new avatars with different body shapes in each training episode to enable scalable training for diverse characters. We varied the body shape parameters during the training process, specifically the β parameters from the SMPL body model [LMR$^+$15a]. This ensured the generated characters had anatomically feasible limb proportions and body dimensions similar to real humans. We achieved this by fitting a set of capsules into the SMPL body model, allowing flexibility and customisation of the characters' physical properties.

Another important component of our framework was the reward system, which encouraged diversity in motion during training. We combined two types of rewards: imitation rewards and regularisation (energy) rewards. The imitation reward evaluated the difference between the character's generated motion and the reference motion. The specific task of this reward was to ensure that the character accurately replicates the target motion. On the other hand, the regularisation reward promoted energy efficiency by encouraging the character to use optimal joint torques, resulting in more physically plausible and natural movements. A balance parameter controlled the interaction between these two rewards, thereby allowing us to fine-tune the desired motion characteristics. Our results demonstrated that this approach generates motions that are faithful to human movement data while adapting to each character's unique body shape.

While our approach is inspired by prior work, such as DeepMimic [PALvdP18] and the parametric controllers introduced by Won and Lee [WL19], it differs in several important ways. DeepMimic focuses on training policies for individual characters with fixed morphologies, whereas our method introduces a single controller that generalises across a wide range of body shapes. Unlike Won and Lee, who incorporate body shape into the state representation, our framework additionally integrates shape-aware control through an energy-based reward function that dynamically balances imitation and efficiency based on shape variation. Moreover, our method employs SMPL-based capsule approximations to ensure anatomical plausibility during character generation. These distinctions allow our system to produce physically consistent and realistic motion for diverse avatars, without retraining for each variation.

In the following sections, we will first provide an overview of physics-based characters and discuss some state-of-the-art methods in the field. We will then discuss the specifics of our framework, examining the techniques used to create motion variations. Finally, we will present the results of our approach and examine the findings.

5.1 DRL & Physics based characters

Virtual avatar motion diversity has traditionally been achieved through techniques like motion retargeting and editing. These methods modify existing animations to suit different characters or environments. For instance, motion retargeting framework [Gle98], interactive editing systems [LCB06], neural motion retargeting models [VYCL18], and more recent contact-based character animation methods [BWBM20]. The introduction of diversity among characters has also been explored by the popular game Spore [HRE+08], which allowed users to create arbitrarily shaped characters and animate them in real-time, showcasing an early example of user-driven character diversity. Later, Ho et al. [HKT10] introduced the concept of interactive meshes, enabling the synthesis of new motions for different body shapes by representing spatial relationships between body parts and objects.

Although we previously discussed physics-based characters in detail in Section 2.4, it is useful to revisit some recent developments in this field. Physics-based animation is a powerful tool for generating realistic character movements, where motion data serves as a reference for simulated models. These models can track target poses while interacting with their environment, maintaining balance and reacting plausibly to changes. For instance, Zordan et al. [ZH02] demonstrated how physics-based controllers can be used to synthesise responsive and physically plausible character movements. Similarly, Levine et al. [LP12] simulated plausible character responses by streaming motion data and feeding it into a physics-based model, with the goal of maintaining realism in a physically simulated environment. Faloutsos et al. [FVdPT01] introduced composable controllers, where each controller operates based on a predefined set of preconditions for determining the current state of the character and predicting transitions between states. These controllers enabled the prediction of in-between states based on example clips, improving motion continuity and realism.

In recent years, Deep Reinforcement Learning (DRL) has gained popularity for controlling physically simulated characters. DRL offers the advantage of learning complex behaviours through trial and error, making it ideal for simulating dynamic characters. Liu et al. [LH17] developed controllers using deep Q-learning, where each action is represented by a short-horizon linear feedback controller. In their approach, control policies are learned by associating them with specific motion sequences. One of the most widely used DRL algorithms in this context is Proximal Policy Optimisation (PPO), developed by OpenAI [SWD+17]. PPO has been successfully applied in various works, including character control, due to its stability and ease of implementation. Based on this algorithm, Peng et al. introduced Deepmimic

[PALvdP18]. In this work, the control policies are trained to imitate reference motions, typically derived from motion capture clips. DeepMimic uses reference state initialisation and early termination techniques to ensure that the learned policies closely follow the target motion, producing high-quality and realistic imitations.

Later, Bergamin et al. [BCHF19] proposed a two-step approach that combines a kinematic controller for motion matching with a dynamic controller trained using reinforcement learning. This allows their system to track a large distribution of motion data while ensuring physical realism in dynamic situations. Won et al. [WGH20] addressed scalability by clustering similar motion styles and training separate dynamic controllers (referred to as "experts") for each cluster. These experts are then combined into a single, unified controller capable of handling a wide variety of motions, significantly expanding the range of animations that can be generated.

Further advancements have explored additional constraints and heuristics to refine the learning process. For instance, Ma et al. [MYT+21] integrated loose space-time constraints with style rewards to guide the learning process, enabling the exploration of different motion styles. Their approach allowed for the generation of motions that not only respect physical constraints but also exhibit stylistic variety.

The method that closely inspired our approach was developed by Won and Lee [WL19]. They introduced parametric controllers designed to accommodate variations in body shape. In their framework, the state space is composed of multiple elements, including the body shape, dynamic state, and additional sensory inputs. By incorporating body shape directly into the state space, their system is able to generalise across a wide range of body types, offering a flexible solution for handling character diversity in animation.

We built our work on these approaches by combining the strengths of DRL and physics-based simulation to generate motion for characters with different body shapes and sizes. We aimed to extend these concepts further by addressing challenges in character control when body shapes vary significantly, ensuring that our framework remains robust and scalable across diverse characters and environments.

5.2 Reinforcement Learning Framework

To formulate our problem as a Deep Reinforcement Learning (DRL) task, we drew on foundational work in motion imitation using RL techniques, particularly the approaches by Peng et al. [PALvdP18] and Won et al. [WL19]. These methods successfully combine motion imitation with reinforcement learning to simulate natural and physically plausible movements for characters in a virtual environment. In our framework, we aimed to imitate a reference motion represented as a sequence

of target poses, $\boldsymbol{q_t}$, where the objective of the policy was to replicate these desired motions using physics-based simulation.

State Space

The state space in DRL represents the configuration of the character in the environment. Our state $\boldsymbol{s}$ consists of joint angles $\boldsymbol{q}$, which capture the character's posture and their corresponding velocities $\dot{\boldsymbol{q}}$. These features define the movement and position of the character at any given moment, crucial for informing decisions within the RL framework.

Additionally, to better capture the character's physical structure, we introduced a body shape state, $\boldsymbol{s_b}$, which encodes information about the size and dimensions of the rigid bodies used to represent the character's body. This body shape state allowed the system to adapt motions based on different body types, making the framework work across a variety of characters. In contrast to DeepMimic, which does not incorporate body shape into the state vector, our formulation explicitly includes body shape parameters to inform the policy. This enables the controller to account for morphology-specific dynamics during training, improving generalisability.

Each of these features was computed relative to the character's local coordinate frame, rooted at the base of the character. This root aligned the character's x-axis with the direction of the root link. This enabled consistent state descriptions irrespective of the character's global orientation. To enhance the state space further, we introduced sensory information $\boldsymbol{s_s}$, following the work of Won and Lee [WL19]. This sensory state contained posture values at different future points in time (0.0, 0.5, and 1.0 seconds ahead), providing the model with foresight into the character's future configurations, improving stability and motion fidelity.

Action Space

The action $\boldsymbol{a}$ generated by the policy represents modification based on body shape to the character's current pose relative to the reference motion, denoted as $\Delta\boldsymbol{q}$. Rather than strictly enforcing exact pose replication, this formulation allows the policy to learn how to best match the intent of the reference motion while adapting it to the specific physical characteristics of each avatar. The PD controller then interprets $\Delta\boldsymbol{q}$ as the desired adjustment, which is applied in the context of the character's mass, shape, and joint dynamics. As a result, the system learns to imitate the reference in a way that is biomechanically plausible for each individual body shape, rather than enforcing rigid pose tracking.

Joints with multiple degrees of freedom (DOF), such as spherical joints, were

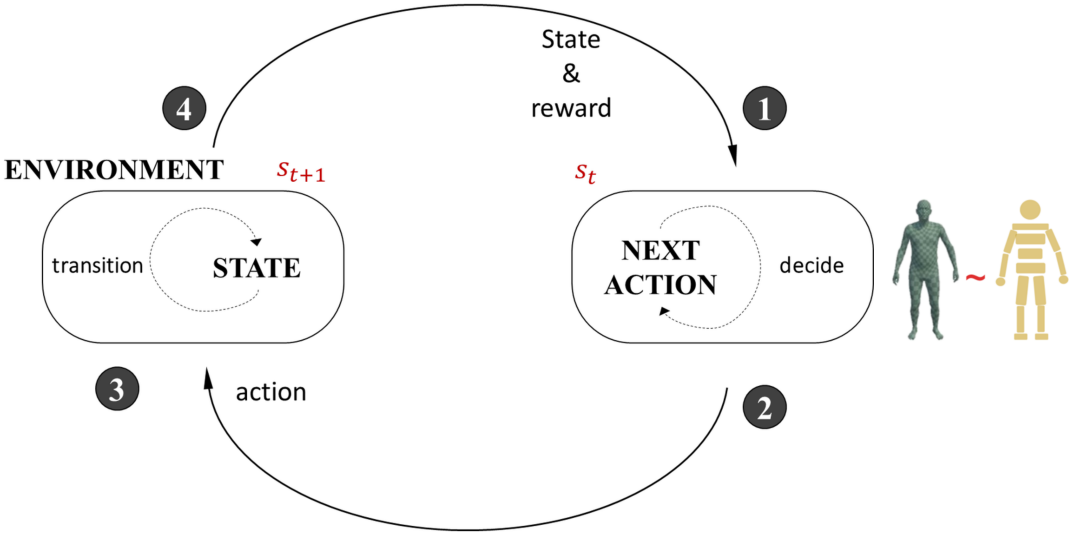


Figure 5.1: The framework for each episode of training: (1) The character is generated using the SMPL body approximation. (2) The parameter θ is updated to calculate the total rewards. (3) Based on the rewards, an action is selected for the next possible state. (4) The state is updated based on the next action, completing the loop.

represented using axis-angle notation, while joints with a single DOF, i.e., revolute joints, were represented as scalar values corresponding to the joint angles. This representation allowed the policy to efficiently control the character's complex joint structures and perform various tasks in the simulation.

In the context of biomechanical motion, the action space plays a vital role in ensuring that the character's movements are both physically plausible and dynamically responsive to environmental factors. Figure 5.1 illustrates the training loop, where actions are selected based on the current state, and the resulting state is updated accordingly.

Reward Function

In general, training the RL policy requires a carefully designed reward function. In our framework, we designed a reward function that encouraged both accurate imitations of reference motions and efficient use of energy. We adopted a reward structure inspired by prior works in imitation learning [PALvdP18, BCHF19, WL19]. Our method dynamically adjusts the trade-off between imitation and regularisation based on the difference in body shape. This body-aware weighting improves motion plausibility across diverse physiques. This reward function consisted of two key components: the imitation reward and the regularisation (energy) reward.

The imitation reward (R_I) measured how closely the character's current pose

matches the reference pose, encouraging accurate replication of the target motions. This reward was composed of several terms:

$$R_I = w^p r_p + w^v r_v + w^e r_e + w^c r_c$$

- The pose reward r_p ensured alignment of joint orientations between the character and the reference motion.

$$r_p = \exp\left(-\sum_j \left\| q^j \ominus \bar{q}^j \right\|^2\right)$$

- The velocity reward r_v encouraged matching joint velocities, promoting smoother transitions between movements.

$$r_v = \exp\left(-\sum_j \left\| \dot{q}^j - \dot{\bar{q}}^j \right\|^2\right)$$

- The end-effector reward r_e measured the positional accuracy of the character's hands and feet compared to the reference motion, ensuring the limbs track the desired paths.

$$r_e = \exp\left(-\sum_i \left\| e^i - \bar{e}^i \right\|^2\right)$$

- The centre of mass (COM) reward r_c encouraged the character to follow the same trajectory as the reference by minimising deviation in the character's COM displacement.

$$r_c = \exp\left(-\left\| v_c - \bar{v}_c \right\|^2\right)$$

The regularisation reward (R_E) promoted energy efficiency by minimising the torques applied to the character's joints, thereby discouraging excessive or unrealistic forces. The regularisation reward was defined as:

$$R_E = w^E \exp\left(-\sum \left\| \boldsymbol{m}_J \tau_i \right\|^2\right)$$

Here, $\boldsymbol{m_j}$ represents the total mass of the rigid body connected to the joint j, and τ_j is the torque applied to the j_{th} joint. By including this term, we ensured that the character used only the necessary amount of force to perform the desired actions, leading to more natural and energy-efficient movements.

During training, the total reward was a weighted combination of the imitation reward and the regularisation reward, with the parameter θ controlling the balance between the two:

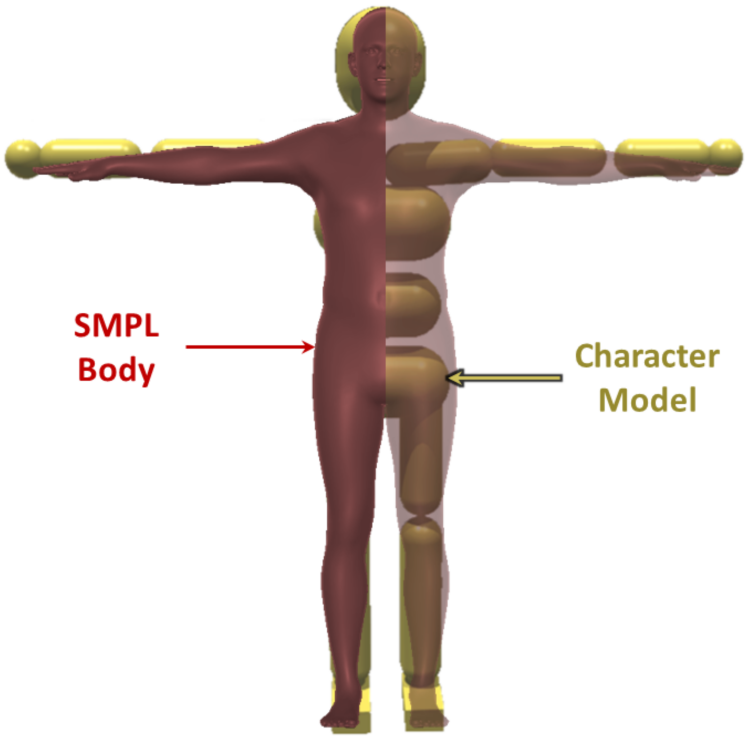


Figure 5.2: The base character model obtained from approximation of SMPL body (both $\beta(0)$ and $\beta(1)$ equals to zero).

$$\boldsymbol{R} = (1 - \theta)R_I + \theta R_E$$

When θ was set to 0, the policy prioritised accurate imitation of the reference motions, while higher values of θ place more emphasis on energy efficiency. This trade-off allowed the policy to flexibly adjust between highly accurate but possibly energy-intensive movements and smoother, more energy-efficient motions.

Motion Variations

To generalise the character's behaviour across different body types, we introduced variations in body shape during training by manipulating body shape parameters ($\boldsymbol{b}$). These parameters modified the character's base shape ($\boldsymbol{B}$) according to the SMPL body model [LMR+15a]. The SMPL model provides a parametric representation of the human body, thereby allowing us to adjust body dimensions along key axes (e.g., height and width) through the first two principal components $\beta(0)$ and $\beta(1)$.

By adjusting these beta parameters, a wide range of body shapes could be gen-

erated for the characters, from slender to broad physiques. These variations were crucial for creating our framework, as it generalises well across different body types. It also ensured that the framework could handle the biomechanical differences that arise from these variations.

To represent the character's body shape, we used a set of capsules that approximate the surface of the body, as shown in Figure 5.2. Each capsule was defined by its radius and axis length, which correspond to different body parts. A regressor mapped the SMPL model's shape parameters to the capsule parameters, aligning the capsules with the character's kinematic chain [BKL+16]. This capsule-based representation ensured that the generated characters maintained structural integrity and stability throughout their movements, regardless of body shape.

During the character generation process, random physical configurations were applied to the character based on these body shape parameters. However, special care was taken to ensure that the generated characters do not exhibit unrealistic behaviour in the environment, such as feet penetrating the ground. To address such behaviour, an initial posture configuration was calculated for each newly generated character, accounting for environmental factors to avoid such issues [WL19].

Balancing Imitation and Energy Efficiency

The total reward function $\boldsymbol{R}$ was influenced by the parameter θ, which allowed us to adjust the trade-off between imitation and energy efficiency. When θ was low, the framework focused on replicating the reference motion as closely as possible. As θ increased, the regularisation reward became more prominent, leading to more energy-efficient, albeit potentially less accurate, motions. This flexible trade-off was essential for generating variations in animations that balance both physical accuracy and efficiency.

For a given motion, we determined θ by first generating a rigid body character using the beta parameters for the base character $\boldsymbol{B}$. Then, we created new characters by varying the body shape parameters and computing the difference $\Delta\beta$ between the base character $\boldsymbol{B}$ and the newly generated character $\boldsymbol{B'}$. If $\Delta\beta = 0$, meaning the body shapes were identical, we set θ to 0 to prioritise imitation. For larger differences, we adjusted θ to reflect the increased emphasis on energy efficiency.

The framework dynamically adjusted θ based on the body shape variations and the character's physical requirements. This produced varied, biomechanically accurate movements across a wide range of body types while still optimising energy efficiency when appropriate. This balance created adaptable, lifelike characters capable of performing complex tasks in virtual environments.

5.2.1 Network Architecture and Training Setup

We trained our reinforcement learning policy, denoted as π_θ, using the Proximal Policy Optimisation (PPO) algorithm [SWD+17] within the Isaac Gym physics simulation environment [MWG+21]. Isaac Gym provides a GPU-accelerated simulator capable of parallel rollout across thousands of environments, making it well-suited for high-throughput policy learning across diverse character morphologies.

The PPO algorithm optimizes a *clipped surrogate objective function*, which is designed to stabilize policy updates by constraining the change in policy between iterations [SWD+17]. In standard policy gradient methods, large updates to the policy can lead to performance collapse. PPO addresses this by introducing a mechanism that restricts how far the updated policy is allowed to deviate from the previous policy during optimization.

Specifically, the surrogate objective function used in PPO is defined as:

$$L^{\text{CLIP}}(\theta) = \mathbb{E}_t \left[\min \left(r_t(\theta)\hat{A}_t,\ \text{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t \right) \right] \tag{5.1}$$

where: - $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ is the probability ratio between the new and old policies, - $\hat{A}_t$ is the estimated advantage at time step t, - ϵ is a small hyperparameter (typically 0.1–0.3) controlling the range of allowed deviation.

The clip$(\cdot)$ function ensures that if the policy update would push $r_t(\theta)$ outside the interval $[1-\epsilon, 1+\epsilon]$, the update is suppressed. This prevents the policy from making large steps that could degrade performance, while still allowing for effective exploration and learning. In our implementation, this clipping mechanism was key to maintaining training stability across characters with highly variable body shapes and physical parameters.

Policy Network Architecture

The policy network is implemented as a fully connected multilayer perceptron (MLP). It comprises three hidden layers with 256 units each, using ReLU activation functions. The final output layer predicts the mean μ of a Gaussian distribution over actions, while the log standard deviation $\log \sigma$ is maintained as a separate trainable parameter vector.

- **Input:** A concatenated vector of joint orientations (axis-angle or scalar), joint velocities, body shape parameters ($\boldsymbol{s_b}$), and future pose predictions ($\boldsymbol{s_s}$).
- **Output:** Mean values for $\Delta \boldsymbol{q}$, representing action modifications applied via PD control.

The value function network shares the same architecture but with a separate set of weights and outputs a scalar estimate of state value $V(s)$.

Training Configuration

Training was performed on a workstation equipped with:

- **GPU:** NVIDIA RTX 2090
- **CPU:** 10th Generation Intel® Core™ i7-10875H (8 cores)
- **RAM:** 32 GB
- **Software:** Python 3.8, PyTorch 1.12, Isaac Gym Preview 2

Simulation:

- **Environments:** 4096 simulated agents in parallel
- **Time step:** 1/60 seconds (60 Hz)
- **Simulation length:** 25 million steps per training run
- **Wall-clock time:** 10 hours per run

PPO Hyperparameters:

- Learning rate: 3×10^{-4}
- Clip ratio ϵ: 0.2
- Discount factor γ: 0.99
- GAE parameter λ: 0.95
- Entropy coefficient: 0.01
- Batch size: 32,768
- Epochs per update: 10
- Optimizer: Adam

This setup enabled efficient learning of a generalisable control policy across characters with diverse SMPL-based body shapes. The use of GPU-parallelised physics and batch processing significantly reduced training time compared to CPU-based environments.

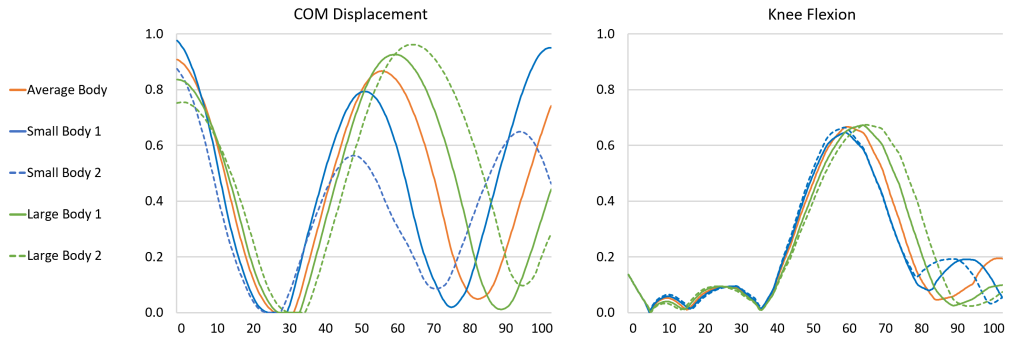


Figure 5.3: Biomechanical metrics for the average, small and large bodies, for one complete gait cycle: normalised root COM displacement (left); and normalised values of knee flexion angle (right).

5.3 Results

To evaluate the performance and robustness of our framework, we selected a set of motion capture data from the dataset provided by Hoyet et al. [HRZ+13]. This dataset includes actors with various BMI values, capturing a range of body shapes and sizes. We utilised the MoSh method [LMB14a] to extract beta shape parameters from the SMPL body model corresponding to each actor's body shape, which were then used to generate the virtual characters. We grouped the actors into two categories: "Small" for actors with lower BMI values and "Large" for those with higher BMI values. The actor with the average BMI was selected to serve as our base character ($\boldsymbol{B}$), from which a policy was trained. For four additional actors with varying body shapes, we modified the base character by adjusting the SMPL beta parameters, generating both "Small" and "Large" body shapes (details are provided in Table 5.1).

Table 5.1: Attributes of actors used to create body models (letters refer to Figure 5.4)

Category	Height	Weight	BMI	$\beta(0)$	$\beta(1)$
Small 1 (f)	167	61	21.9	1.59	-0.55
Small 2 (a)	179	76	23.7	0.04	-0.45
Average (g)	182	82	24.8	-0.35	-0.31
Large 1 (b)	172	88	29.7	0.77	0.75
Large 2 (c)	183	86	25.7	-0.50	-1.64

Our framework successfully handled body shape variations within a single trained policy. This was primarily due to the incorporation of shape-specific parameters in the state space and the use of a shape-agnostic policy that can generalise across different body shapes without retraining for each specific shape. This generalisation

contrasts with previous work, which typically requires retraining or fine-tuning for different bodies. Our method achieves comparable fidelity across shapes without this overhead.

The PPO algorithm operates by updating the policy based on a clipped surrogate objective function. Our approach was designed to strike a balance between exploration and exploitation during training by preventing the policy from deviating too much from the previous iteration while ensuring enough variability to find better solutions. This combination of exploration and stability allowed PPO to perform well in continuous control environments like ours, where the character's body shape changes dynamically.

Biomechanical Evaluation

To assess the effects of body shape on motion, we compared some key kinetic parameters across different body shapes. The focus was on the lower body, as we discussed in Section 2.1, the lower-limb kinematics are crucial for motion analysis. Given the importance of body mass in motion patterns, we examined how changes in body shape influence motion characteristics, particularly the pelvis trajectory and knee flexion during walking. Our results demonstrated distinct motion patterns for different body shapes, with noticeable variations in COM displacement and knee flexion.

Figure 5.3 (left) illustrates the trajectory of the pelvis for each character over a complete gait cycle. Notably, larger characters exhibited greater pelvis displacement compared to smaller ones. This is consistent with findings from prior studies, which suggest that body mass has a direct effect on COM displacement, impacting gait stability and energy consumption [MCM16]. For example, heavier individuals tend to have a larger COM displacement to maintain stability during motion, which increases energy expenditure. This effect was replicated in our physics-based simulation.

Similarly, we analyzed the knee flexion of the left leg, focusing on its range of motion throughout the gait cycle (Figure 5.3, right). Larger characters exhibited slower and less pronounced knee flexion compared to smaller characters. This result is in line with observations in medical research, which indicate that individuals with higher body mass often experience a reduced range of motion in their joints, particularly in the knees [BK07, MCM16]. This is likely due to the additional force required to move the larger mass, which affects both the speed and the amplitude of joint movements.

These differences underscore the role of body shape in determining motion style

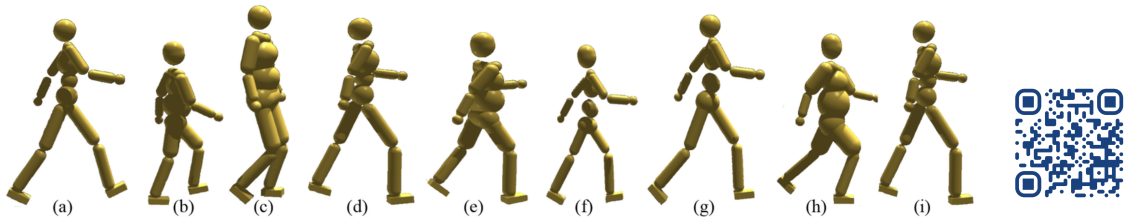


Figure 5.4: Human body representations with motion variations generated by the DRL framework, adapted to different body shape parameters. The full set of motion results can be accessed via the QR code or at: https://youtu.be/CnrKNBhSEuc.

and efficiency. The heavier, larger-bodied characters needed to expend more energy to move their limbs, and as a result, their motion patterns were characterised by reduced joint angles and slower speeds. Conversely, smaller characters, with less mass to move, exhibited quicker and more agile movements. This highlights the importance of incorporating body shape into the reward function, which allows the policy to account for variations in motion style and energy efficiency across different body types.

5.4 Discussion

In this chapter, we introduced a novel DRL framework for generating realistic motion for physics-based characters with varying body shapes. By incorporating body shape parameters into both the state representation and the reward function, our approach allowed for the creation of highly diverse characters while maintaining natural motion. This is particularly valuable for applications in gaming, animation, and virtual reality, where a wide range of body shapes and sizes need to be represented in a realistic and physically plausible manner.

Our results demonstrate the influence of body shape on motion patterns, especially in terms of biomechanical parameters like pelvis trajectory and knee flexion. As shown in our evaluations, larger characters tend to have slower and fewer movements compared to smaller characters. This suggests that incorporation of body mass and size factors generated animations that are both physically realistic and diverse.

An important contribution of our work is the energy-based reward function, which incentivises the policy to generate motions that are not only visually accurate but also energy-efficient. By incorporating the mass and inertia of the body into the reward function, we were able to simulate the physical effort required for different body shapes to move. This approach enabled the generation of varied motion styles that correspond to different body shapes. For example, a heavier character may

exhibit a more constrained motion style, reflecting the increased energy expenditure required to move a larger mass. In contrast, a lighter character would exhibit more fluid and agile movements, as less energy is needed to generate motion.

However, the current framework presents some challenges, Especially, when it comes to integration with existing animation pipelines. This lack of integration with popular physics-based game engines limits its accessibility. Another limitation is the simplified modelling of physics-based characters, where body shapes are approximated using capsules, leaving gaps between these capsules. This approximation relies on several assumptions to simulate motion, which compromises the accuracy of the character's movements. Introducing more precise character models would not only ensure that motions adhere more closely to biomechanical principles but also align better with human perception. Additionally, our current framework primarily focuses on body mass as the primary factors influencing motion. However, other physiological factors, such as muscle strength, flexibility, and centre of gravity, may also play a significant role in determining motion patterns. Incorporating these factors into the state representation and reward function could lead to even more realistic and diverse motion generation. For example, including muscle dynamics could allow for the simulation of fatigue or varying levels of physical fitness, which would further enhance the realism of the generated characters.

Another promising future direction would be the integration of auto-encoders or conditional variables further to control the style and quality of the generated motions. By incorporating these machine learning techniques, it would be possible to create a more flexible and customizable framework, where users could control both the body shape and motion style of the characters. This could have significant applications in personalised animation, allowing for the creation of characters that not only resemble a particular individual but also move in a way that reflects their unique physical characteristics.

In conclusion, our framework provided a strong foundation for generating diverse, physics-based character animations that account for body shape variations. Unlike prior methods that decouple shape and control or require training per character, our controller offers a scalable, physics-driven solution for motion generalisation. This makes it better suited for applications like crowd simulation or character personalisation, where shape diversity is key. The incorporation of biomechanical principles and energy-based rewards allows for the creation of physically plausible and visually diverse motions. However, we also explored alternative methods for introducing variation in existing animation pipelines. In the next section, we will examine a different technique that uses human anthropometric data to drive variations based on body shape.

Chapter 6

Balance Point Controller

Human perception and understanding of the virtual world are heavily influenced by the physical laws in nature. With this in mind, we aimed to replicate similar guiding principles for human motion in virtual environments. The system presented in this chapter referred to as the Balance Point Controller, adapts motion based on the character's physical attributes, such as body weight and height. Additionally, this system is compatible with game engine pipelines like Unity 3D, enhancing its practical application and accessibility.

The presented framework integrates both kinematics and dynamics, enabling it to adjust motion based on a character's body shape. It takes reference motion as input, along with the target body weight and height. Then, the physics-based character is constrained to follow this reference motion. The user-defined parameters of height and weight alter the character's inertial properties. This results in a new, adapted motion that better reflects these physical attributes. The framework is part of our exploration for research question **Q2**, and the evaluation of it explores the research question **Q3**, discussed in Section 1.1.

Prior work in balance control includes Zero Moment Point (ZMP)-based techniques [SKG03], which stabilise motion by keeping the projected centre of gravity within the support polygon. Mordatch et al. [MDLH10] introduced an inverted pendulum model, which simplify character dynamics into a reduced planning space. Zordan et al. [ZMCF05] have also shown how reactive feedback controllers can maintain balance during dynamic tasks. Unlike these approaches, our Balance Point Controller does not explicitly model ZMP or plan foot contacts. Instead, it modifies the internal mass distribution using anthropometric principles, allowing physical simulation to produce natural posture adjustments as a result of altered COM. This creates a lightweight mechanism for exploring shape-sensitive motion variations while remaining engine-friendly and interpretable.

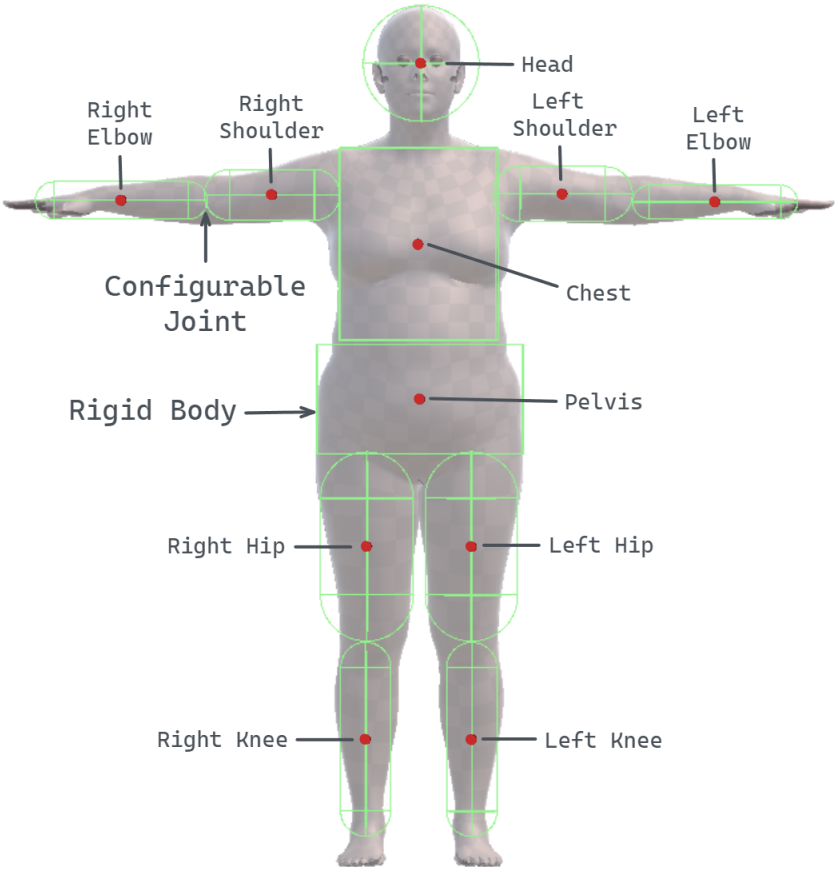


Figure 6.1: Ragdoll model with colliders, rigid bodies, and configurable joints in body segments. The red points depict the default COM position for the respective rigid body.

In the following sections, we will explore how the physics-based character is constructed, look into the details of the balance point control mechanism, and outline an experimental methodology for evaluating such frameworks.

6.1 Physics-based Character

The character model was implemented in the Unity™ game engine. These DoFs include both translational and rotational components. Specifically, the root segment (pelvis) is modeled with 6 DoFs—3 for global translation (x, y, z) and 3 for global rotation (represented using Euler angles). Only the rotational DoFs of the root (3 DoFs) are actuated by the PD controller, while the translational components are left unactuated and are governed by physics simulation (i.e., treated as a floating base). The remaining 16 DoFs correspond to rotational joints in the limbs (e.g., hips, knees, shoulders, elbows), all of which are fully actuated by the controller. No translational

joints are actuated. The model used capsule- or cuboid-shaped rigid bodies for the major limbs, along with a sphere for the head, as depicted in Figure 6.1. Such models are commonly referred to as ragdolls in game engine terminology.

In the character model, the mass of the rigid bodies was based on the anthropometric mass distribution, discussed later in Section 6.2. Each DoF represented the relative rotation or translation of two connected body segments by a revolute or prismatic joint, respectively. Since Unity uses different representations of joints, configurable joints were used to connect the two rigid body segments. These joints were configured to limit the motion in different axes, based on the limb type.

A PD controller is a type of feedback control system that is commonly used in engineering applications to stabilise a system and reduce errors. Our system was based on the control law that combines the proportional and derivative terms:

$$\mathrm{u(t)} = -k_p(e(t)) - k_d\dot{e}(t) \tag{6.1}$$

where $u(t)$ is the control signal, $e(t)$ is the error between the desired and actual state, and $\dot{e}(t)$ is the derivative of the error. k_p and k_d are the proportional and derivative gains, respectively. The gain values determined the relative contribution of each term to the control signal. There are stable PD controllers [TLT11] which avoid the manual tuning of gain values, but we used the traditional controller to have more control over the character.

We used the PD controller to compute control forces and torques to drive the *ragdoll* from its current pose towards the target pose. The torque and force were evaluated in every time step for every joint. The character was controlled by specifying the desired joint configuration, where $\mathbf{q}$ denotes the vector of joint angles, and $\dot{\mathbf{q}}$ the corresponding joint angular velocities. The PD controller then computed the joint torques $\boldsymbol{\tau}$ as follows:

$$\boldsymbol{\tau} = -k_p(\mathbf{q} - \bar{\mathbf{q}}) - k_d\dot{\boldsymbol{q}} \tag{6.2}$$

Here, $\bar{\mathbf{q}}$ represents the reference joint angles derived from the target motion. The terms k_p (0.155) and k_d (96.5) are the proportional and derivative gain values, respectively. The gain values were computed through a combination of trial and error and optimal control techniques.

We employed a PD controller to generate forces and torques for the simulation. The PD controller computed forces and torques based on the deviation between the character's current configuration and the reference motion, enabling smooth tracking of target poses on the ragdoll.

Regarding the gain values for the controller, we initially used a simple optimi-

sation algorithm to compute estimated values that worked well. Then, we further improved the controller through trial and error. First, we minimised a straightforward objective function, defined as the squared difference between the corresponding angles of the ragdoll motion and the reference motion. However, after obtaining estimated values, we manually fine-tuned the gain parameters to achieve stable and optimal motion for the base character. The final values of the controller's gain parameters were as follows:

Proportional Torque (k_{p_τ}): 0.16
Proportional Force (k_{p_f}): 30
Derivative Torque(k_{d_τ}): 0.002
Derivative Force(k_{d_f}): 0.01

Algorithm 1 outlines the optimisation loop used to tune the PD controller. We ran this procedure for a short motion sequence (1–2 gait cycles) and used the resulting gains as a baseline, which we subsequently refined manually to improve general stability across body shapes.

Algorithm 1: Optimisation of PD Controller Gains

Input: Target joint angles $\bar{q}_{1:T}$, initial gains k_p, k_d
Output: Optimised gains k_p^*, k_d^*

Initialise $k_p \leftarrow k_p^{(0)}$, $k_d \leftarrow k_d^{(0)}$;
for $i \leftarrow 1$ **to** N **do**
  Simulate character using current k_p, k_d;
  Compute pose error: $E_q \leftarrow \sum_{t=1}^{T} \|\bar{q}_t - q_t^{(i)}\|^2$;
  Compute velocity penalty: $E_{\dot{q}} \leftarrow \sum_{t=1}^{T} \|\dot{q}_t^{(i)}\|^2$;
  Total loss: $L \leftarrow E_q + \lambda E_{\dot{q}}$;
  Update gains using gradient descent:;
    $k_p \leftarrow k_p - \alpha \cdot \nabla_{k_p} L$;
    $k_d \leftarrow k_d - \alpha \cdot \nabla_{k_d} L$;
return k_p^*, k_d^*

The gains were obtained to effectively control the limb torques and world forces responsible for driving the base character for averaged motion. When the gain parameters were poorly selected for the base character, they led to motion divergence, affecting stability and accuracy. The chosen gain parameters worked well for all characters, ensuring stable motion.

6.2 Balance Point Control

The stability of a physics-based model is an important factor in achieving realism in character animation. The centre of mass (COM), sometimes referred to as the balance point, plays a key role in defining stability during movement, as it represents the point where the body's mass is considered to be concentrated (see Section 2.1 for further discussion). Each rigid body in a *ragdoll* has its own local COM, which follows the hierarchy of the ragdoll skeleton. The pelvis, located at the top of the hierarchy, influences all child branches. Therefore, the segment-level COM positions are related to the skeletal structure and must be carefully configured to promote plausible and balanced motion. These positions were estimated using anthropometric data to represent average COM distributions across segments.

While the PD gains used to control the character were fixed across all body configurations, variations in segment mass and geometry influenced how the physics engine handled dynamics. As a result, COM displacement indirectly affected joint torques and postural stability during motion tracking.

Table 6.1: Distribution of mass across body segments and the COM displacement along the longitudinal axis. The data are derived from the anthropometric studies conducted by McConville et al. and Young et al. [MCC+80, YCS+83].

Segment	**Weight factor** (X_w)	**COM factor** (X_l)
Pelvis	-	-0.088
Head	0.073	-0.034
Spine	0.507	-0.12
Upper Arm	0.026	0.013
Lower Arm	0.016	-0.46
Hand	0.007	-0.32
Upper Leg	0.103	-0.128
Lower Leg	0.043	-0.129
Foot	0.015	-0.051

Anthropometric datasets, such as those by McConville et al. and Young et al. [MCC+80, YCS+83], provide empirical values for estimating the distribution of mass and COM position within individual body segments. These studies show that the COM within a segment is typically offset from its geometric centre, depending on the segment's length and mass properties.

Table 6.1 lists the COM and weight factors used to model each limb segment. The mass of each rigid body in the *ragdoll* (with $N = 11$ segments) was computed as:

$$\boldsymbol{m_i'} = W \cdot X_{w_i} \tag{6.3}$$

where W is the total body weight.

The COM of each segment was estimated as an offset along the segment's longitudinal axis, defined from the base (proximal) joint of the segment. Let $\boldsymbol{c}_{\text{base},\boldsymbol{i}}$ denote the base position of segment i, and $\boldsymbol{d_i}$ its unit axis direction. Then:

$$\boldsymbol{c_i} = \boldsymbol{c}_{\text{base},\boldsymbol{i}} + (l_i \cdot X_{l_i}) \cdot \boldsymbol{d_i}, \quad \boldsymbol{c'_i} = \boldsymbol{c'}_{\text{base},\boldsymbol{i}} + (l'_i \cdot X_{l_i}) \cdot \boldsymbol{d'_i} \tag{6.4}$$

Here, l_i and l'_i refer to the limb lengths of the base and updated characters, respectively, while X_{l_i} is the anthropometric factor for that segment. This formulation ensures that characters with identical segment lengths and proportions yield the same COM positions.

The total skeletal COM was then computed as the mass-weighted sum of the segment-level COMs:

$$\boldsymbol{C} = \frac{\sum_{i=1}^{N} \boldsymbol{c_i} m_i}{\sum_{i=1}^{N} m_i}, \quad \boldsymbol{C'} = \frac{\sum_{i=1}^{N} \boldsymbol{c'_i} m'_i}{\sum_{i=1}^{N} m'_i} \tag{6.5}$$

The balance control factor $\boldsymbol{b}$ was defined as the Euclidean distance between the original and updated skeletal COMs:

$$\boldsymbol{b} = \|\boldsymbol{C} - \boldsymbol{C'}\| \tag{6.6}$$

For characters with identical height and symmetrical segment scaling, differences in total weight alone have a limited effect on the COM location in the horizontal plane. However, small vertical shifts or asymmetries may still impact balance and posture through dynamic interactions during control.

The effect of the balance control factor on posture is shown in Figure 6.2. When $\boldsymbol{b}$ was near zero, the ragdoll closely matched the reference kinematic pose. Larger values of $\boldsymbol{b}$—typically arising from height changes—led to more visible postural deviations. This demonstrates how shifts in COM, even without tuning PD gains, can influence the tracking fidelity and physical plausibility of character motion.

6.3 Testing and Validation

We assessed the physics-based framework using the character models with different body weights and heights as input, and the results were then compared with data obtained from studies on real human walking gait analysis.

As discussed by MacLean et al. [MCM16], there were some major differences in

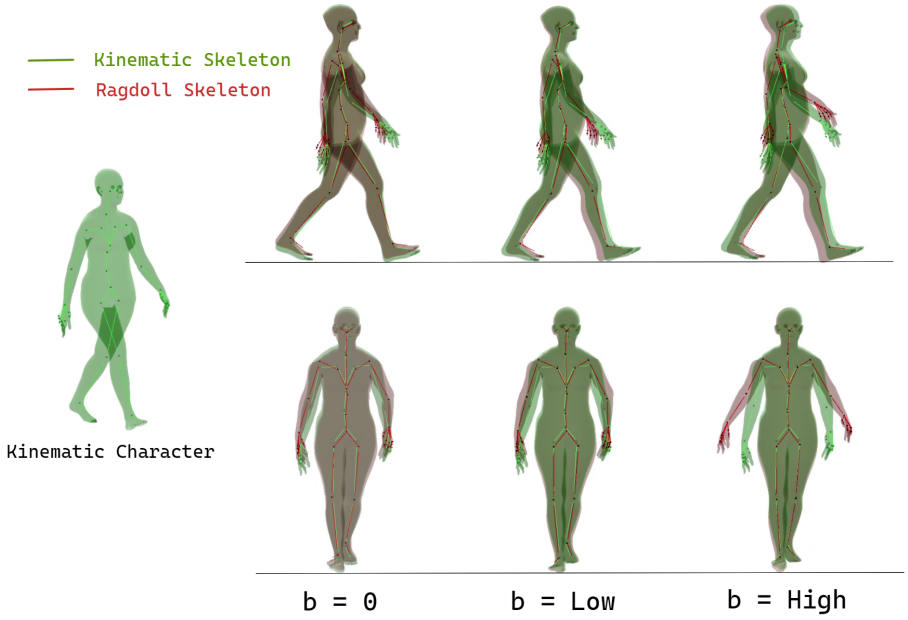


Figure 6.2: Effect of Balance Control factor on the posture. The green character represents the kinematic skeleton acting as the target posture, while the red character shows the ragdoll skeleton attempting to mimic the target. The posture of the ragdoll is influenced by the balance control factor.

the lower body extremities between different BMI groups as shown in Figure 6.3. The results from our method, shown in Figure 6.4, were time-normalised to 100% of the gait cycle and demonstrated the variability across characters generated with different body dimensions.

Due to balance point control, the trajectory of the pelvis was distinct for each character, as shown in Figure 6.4 (left). The total body mass of the character played a significant role in determining the COM position of the pelvis. The position of the COM was affected by the diverse body mass indices (BMI) of the characters.

The knee extension moment during one gait cycle varied across different BMI characters, as shown in Figure 6.4 (middle). The left knee moment values (N-m) of each character were normalised and plotted for comparison. It could be observed that the difference in moment values between higher and lower BMI characters was consistent with the values reported by MacLean et al. [MCM16]. However, the absolute peak values for higher BMI individuals did not exactly align with the peaks reported by Browning & Kram [BK07] and MacLean et al. [MCM16].

Figure 6.4 (right) shows the left leg's normalised knee flexion values (°). The peak in the curves indicates that characters with higher BMI have lower flexion compared to those with lower BMI, which suggests the impact of additional mass carried by the character. The patterns in the angle values are consistent with the

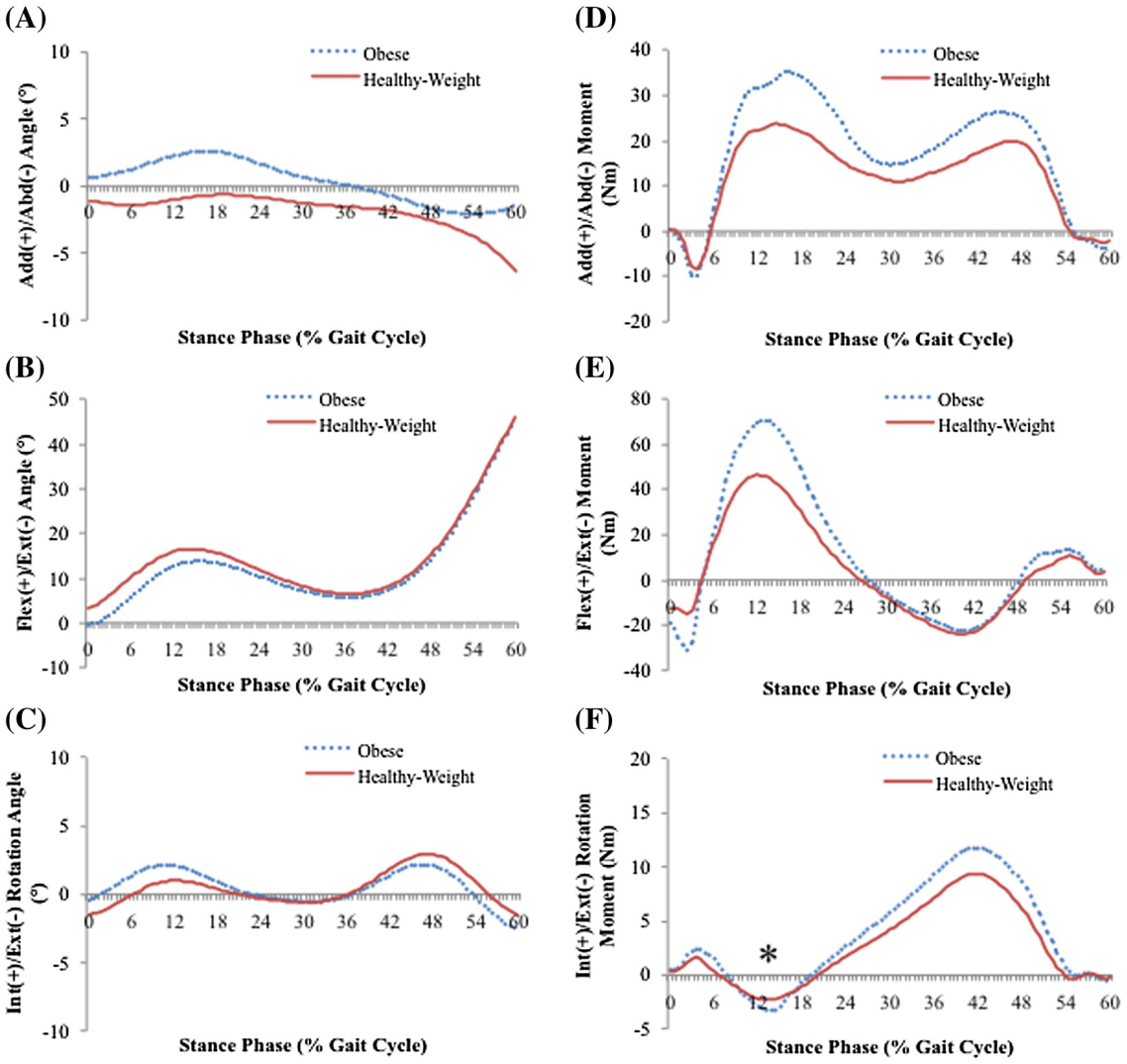


Figure 6.3: Average (A) adduction/abduction, (B) flexion/extension, (C) internal/external rotation angles, (D) adduction/abduction, (E) flexion/extension, (F) internal/external rotation moment waveforms for both participant groups at the matched speed condition. Adapted from [MCM16]

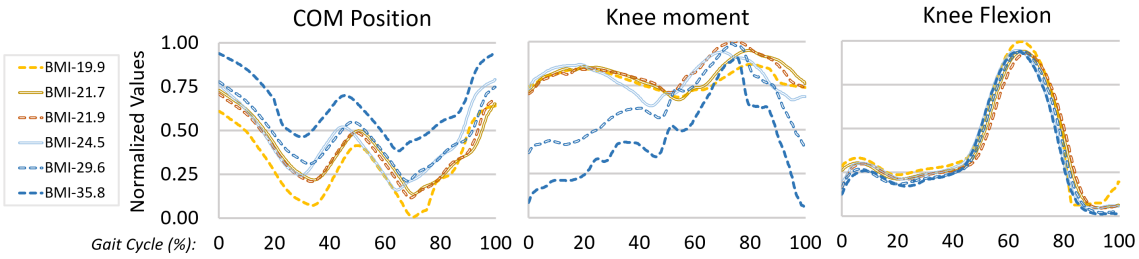


Figure 6.4: Comparison of biomechanical parameters across characters with different BMIs for one complete gait cycle – i) Root COM Position: normalised position for the level of displacement; ii) Knee Moment: normalised values of the Knee extension moment; iii) Knee Flexion: Normalised values of knee flexion angle.

results and observations reported by Browning & Kram [BK07] and MacLean et al. [MCM16].

The analysis of upper body parameters is limited due to several factors. Firstly, the lack of sufficient literature in this specific area restricts a detailed examination of the arms and head. However, previous studies have shown that the lower-limb and trunk model can effectively capture the kinematics of the overall body's COM with minimal errors [HMF18, VGRL10]. This suggests that the lower limb and trunk region have a greater influence on the properties of the COM.

The balance point control resulted in distinct and comparable gait patterns for characters with varying body dimensions. The comparison with real human gait analysis data validated these results. Overall, the method gave promising results which we evaluated through two experiments. The motion generated through this method is referred to as COM motion in the rest of the chapter.

6.4 Experiment 1: Evaluation with Quasi-average Motion

In this experiment, we aimed to evaluate the adapted or modified motions generated by the Balance Point Controller and create a framework that enables an effective evaluation of generated motions. In this way, we could examine whether the motions generated by the controller are perceived as natural or not. To control for style differences in motion data, we used an average motion within similar BMI groups as a baseline for this experiment, which we refer to as quasi-average motion. As a result, we could also assess the degree to which physics-based retargeting can effectively capture the effects of body shape on the perception of human motions.

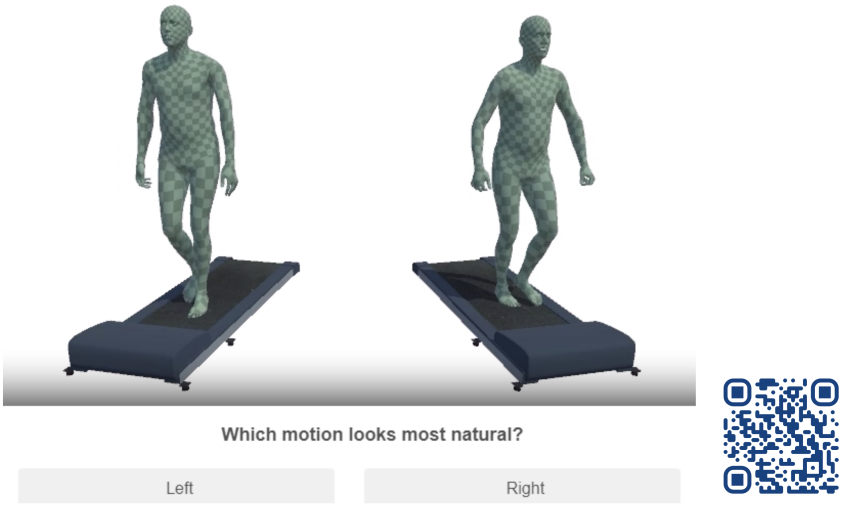


Figure 6.5: Experiment 1 example stimulus: Quasi-average and controller motion were presented on the same body shape and participants selected the one they found more natural. The full set of stimulus videos can be accessed via the QR code or at: https://youtu.be/9JexbzPHh3o.

6.4.1 Method

Stimuli

The stimuli used in this study were similar to those used in Section 3.2, and selected the same actors for body shape and motion. We used a technique similar to that outlined in [HRZ+13] to create the baseline or quasi-average motion. Due to the cyclic nature of walking and jogging motions, it was easier to find the contact phases. Then, we used Dynamic Time Warping to align and generate the quasi-average motion. To create the retargeted motions, we used the same baseline motion, which was an average motion within each similar BMI group, as controller input. Later, the controller modified it to match the skeletal dimensions of individuals in different BMI groups. The retargeted motions were generated using the method as described in Section 6.2. The SMPL-X bodies were used for both the retargeted motions and quasi-average motion. Each video clip lasted approximately 4 seconds and was shown from a third-person perspective. The perspective view remained consistent throughout the experiment, and the two motions were displayed simultaneously side-by-side on the same screen in a standardised format.

Participants

Twenty volunteers from uncontrolled backgrounds participated in the online survey. Among them, 11 identified as male, 8 as female, 1 as non-binary, and 1 preferred

not to disclose their gender. In terms of age, 8 participants were aged 18–29, 8 were aged 30–45, 3 were over 60, and 1 preferred not to disclose their age. Regarding experience with computer graphics systems, 11 participants reported expert-level experience, 5 had some experience (e.g., as users such as gamers), and 6 reported little or no prior experience. No personally identifying information was collected during the experiment for participant anonymity. We used Qualtrics, an online platform to administer the experiment. Participants viewed two videos side-by-side on the same screen, as shown in Figure 6.5, and were asked to choose which motion they found more natural by clicking on one of the two options, left or right.

Procedure

The experiment used a mixed design to investigate the effects of between-subjects factor SEX (Male, Female) and within-subjects dependent variables: GAIT (Walk, Jog), BODY type (High BMI [H1, H2], Low BMI [Lo1, Lo2]), and body MOTION (Hc, Hi, Lc, Lo). The Hi and Lo refer to the quasi-average motions of subjects with high and low BMI, respectively, while Hc and Lc refer to the retargeted motions adjusted to the respective quasi-average motions.

Similar to the experiment described in Chapter 3, we divided the study into two blocks, with each participant viewing either male or female animations. Within each block, there were two Gait conditions: walking and jogging. The order of these two gait conditions was randomised for each participant to minimise order effects. There were a total of two repetitions for each Gait condition. To prevent any potential perspective effects, we swapped the left and right sides of the videos during the repetitions, resulting in a total of four repetitions for each block. Each participant completed a total of 16 trials in each Gait block, with four trials. The entire survey took approximately 30 minutes to complete.

6.4.2 Results

Table 6.2 shows the effect of Gait, Sex, Body Type, and Body Motions on the selection of more natural motion. The main effect of motion (i.e., the naturalness of the motion) was significant, $F(3, 396.41) = 132.14$, $p < 0.001$, $\eta^2 = 0.88$. This indicates that the choice of motion significantly influenced the participants' perception of naturalness. The two-way interaction effect between motion and sex was also significant, $F(3, 23.02) = 7.67$, $p < 0.001$, $\eta^2 = 0.30$, indicating that the effect of motion on naturalness perception was different depending on which stimuli were rated by observers. The two-way interaction effect between gait and motion was also significant, $F(3, 79.30) = 26.43$, $p < 0.001$, $\eta^2 = 0.59$, which indicates that

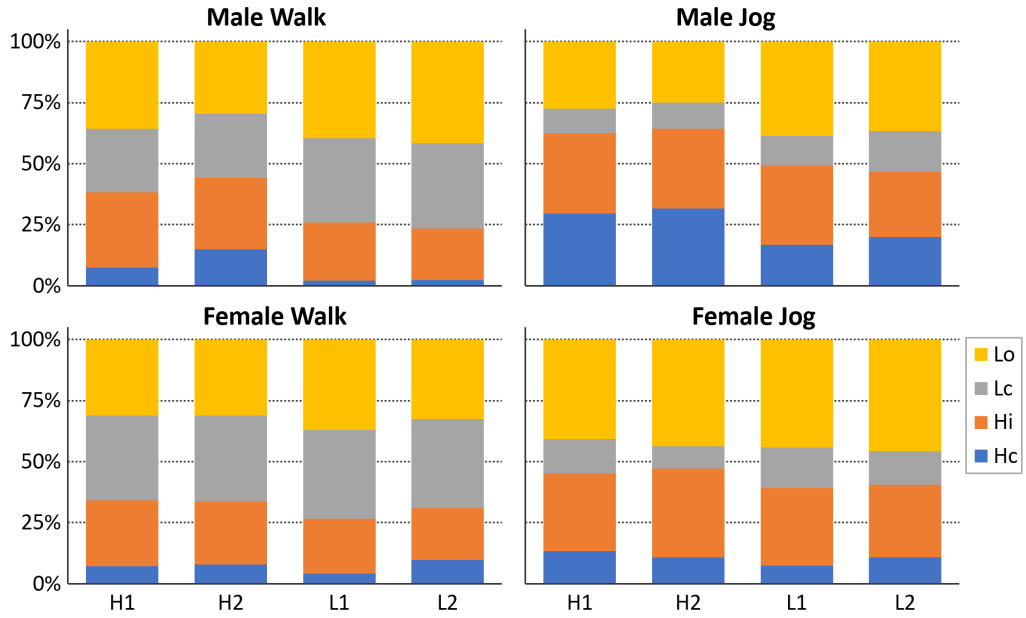


Figure 6.6: Experiment 1 Results: Percentage of time each motion was chosen as more natural.

the effect of gait on naturalness perception was different depending on the type of motion. The two-way interaction effect between body type and motion was also significant, $F(9, 84.81) = 9.42$, $p < 0.001$, $\eta^2 = 0.34$, suggesting that the effect of body type on naturalness perception was different depending on the type of motion. Furthermore, the three-way interaction effect between gait, motion, and sex was also significant, $F(3, 16.24) = 5.41$, $p < 0.001$, $\eta^2 = 0.23$. This interaction indicates that the interaction between gait, motion, and sex significantly influenced the participants' perception of naturalness. Finally, the three-way interaction effect between body type, motion, and sex was also significant, $F(9, 30.40) = 3.38$, $p < 0.001$, $\eta^2 = 0.16$. This result suggests that the interaction between body type, motion, and sex significantly influenced the participants' perception of naturalness.

The results suggest a preference for low motion in both male and female characters, especially for walking motions. This bias towards low average motion is further evident in the case of Lc, which also dominated even for higher BMI body shapes, which can be clearly seen in Figure 6.6. Conversely, Hi motion had a lower preference, even for higher BMI body shapes. The matrix table in Figure 6.7 shows all the possible combinations of controller and quasi-average motion during the experiment. The number in each box shows the percentage of the number of times column heading motions were selected by participants. It clearly demonstrates that Hc performed worse for both walking and jogging motion in both females and males. This distribution of motion preference across all conditions suggests a heavy bias

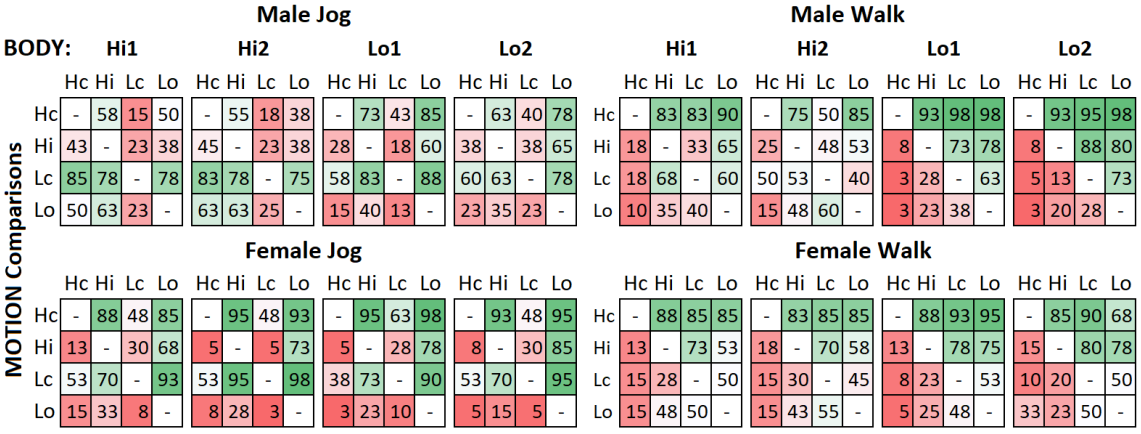


Figure 6.7: Experiment 1 Results: Each column heading shows, for the given motion, the percentage of times it was chosen for each of the row headings.

Table 6.2: Experiment 1 ANOVA: Significant effects with effect size (partial η^2), non-centrality (nc) and observed power ($\alpha = 0.05$).

Effect	**SS**	**dof**	**MS**	**F-Test**	p	η^2	**nc**	**power**
Main Effects								
MOTION	196117	3	65372.40	132.14	0.00	0.88	396.41	1.00
Two-way Interaction Effects								
MOTION*SEX	11386	3	3795.43	7.67	0.00	0.30	23.02	0.98
GAIT*MOTION	90855	3	30285.01	26.43	0.00	0.59	79.30	1.00
BODY*MOTION	21570	9	2396.70	9.42	0.00	0.34	84.81	1.00
Three-way Interaction Effects								
GAIT*MOTION*SEX	18610	3	6203.41	5.41	0.00	0.23	16.24	0.92
BODY*MOTION*SEX	7732	9	859.09	3.38	0.00	0.16	30.40	0.98

towards the Lo motion style, which is preferred more than Hi. Furthermore, the retargeting method did not perform well against quasi-average motion. However, the results clearly demonstrate the role of certain motion styles on participants' preferences and controller output. In another experiment, we tried to overcome the motion style bias and gain a deeper understanding of how different factors interact.

6.5 Experiment 2: Evaluation with Overall Average Motion

Our previous experiment demonstrated that motion style has an influence on the naturalness ratings in some scenarios. Here, we aimed to further investigate the rela-

tionship between human motions and body shapes and to assess the effect of physical properties on this relationship. Specifically, the experiment sought to evaluate the physics-based framework using pairs of side-by-side gait comparisons.

To ensure consistency and control over the motion style in our study, we utilised the average motion files from the same dataset previously used in Chapter 3 and Chapter 5. For this experiment, the average motion files were created by combining all the motions available in the motion database. The stimuli used in this experiment were similar to those employed in Experiment 1, utilising the same actors for body shape selection. Each video clip was presented from a third-person perspective and featured two motions displayed side-by-side on the same screen. Participants were tasked with selecting the motion they perceived as more natural, allowing us to assess their ability to recognise the same actor across different gaits.

6.5.1 Method

Stimuli

The experiment stimuli were generated using the same dataset as in Experiment 1 (Section 6.4). To ensure consistency between the experiments, we selected the same actors for the body shape parameters.

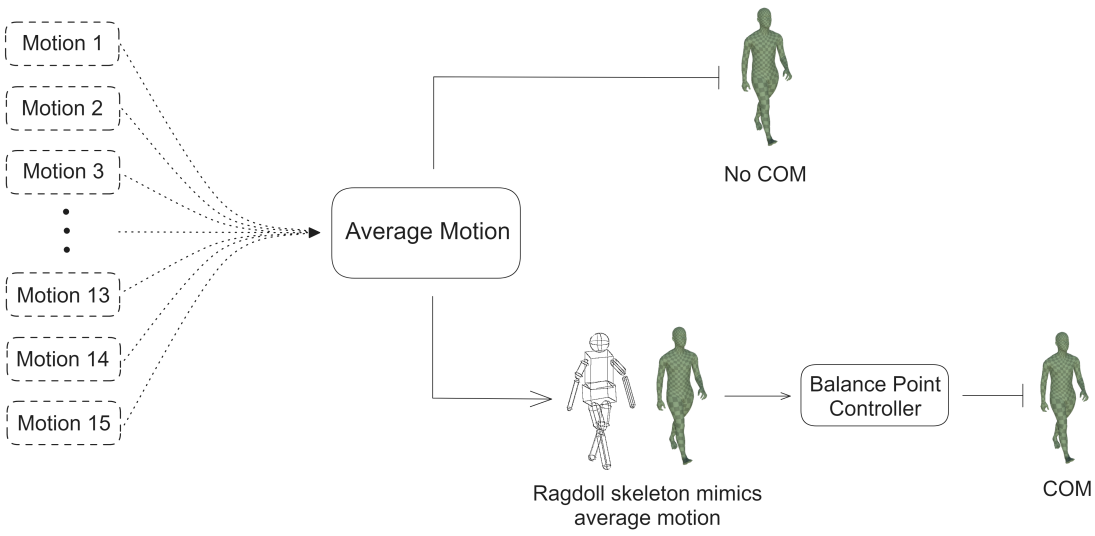


Figure 6.8: Experiment 2 Stimuli Generation Process: The process involves using average motion, where the *No COM* approach applies average motion directly to the meshed body. In the *COM* approach, a ragdoll skeleton mimics the average motion through retargeting, facilitated by the Balance Point Controller.

Similar to the last experiment, we established a baseline for the study. We incorporated the overall average motion files similar to the prior study conducted by

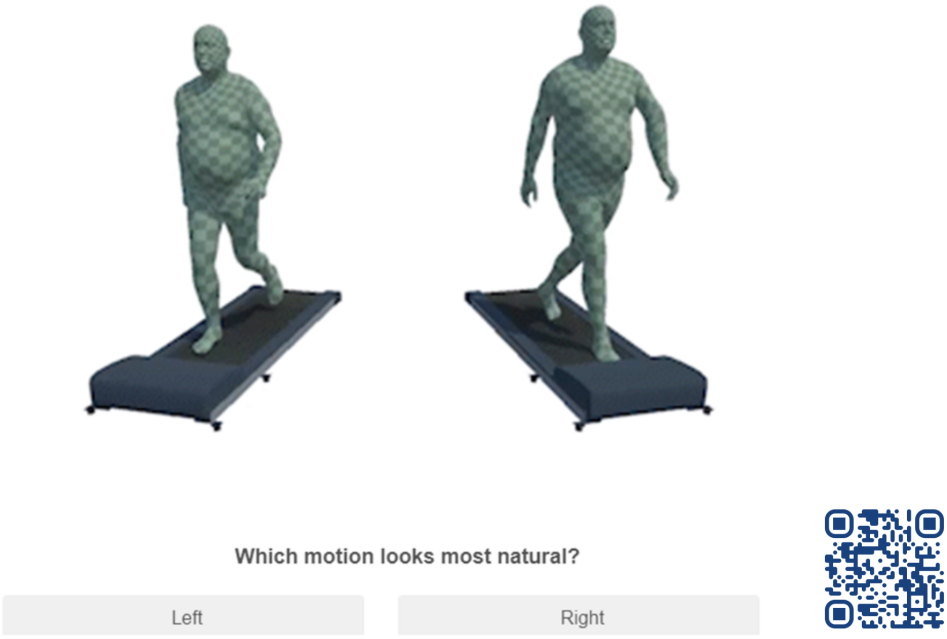


Figure 6.9: Experiment 2 example stimulus: Average and controller motion were presented on the same body shape and participants selected the one they found more natural. The full set of stimulus videos can be accessed via the QR code or at: https://youtu.be/YhKqrFWRZTY.

Hoyet et al. [HRZ+13]. This involved the implementation of Dynamic Time Warping, a technique used to align and derive the average motion patterns. Specifically, separate average motions were created for both the walk and jog gaits, using motion data from a total of 15 male actors and 15 female actors.

To generate retargeted motions, we adopted the same methodology as outlined in the previous experiment, in Section 6.4, utilising the SMPL-X body models. In Unity, the average motion was rendered using the SMPL-X Unity plugin, employing the normal animation system as shown in Figure 6.8 (No COM). For the physics-based retargeting method, we utilised the dual-rig approach to render the animation. The average motion was played normally in the animation system, with the mesh renderer turned off (master rig). Subsequently, a ragdoll with the same mesh was created to follow the movements of the master rig as shown in Figure 6.8 (COM). To generate the retargeted motion on top of the average motion, the weight and height of the actors were manually inputted.

Participants

We recruited sixteen new volunteers who had not previously participated in Experiment 1. This was done to minimise potential bias or carryover effects from the previous study and ensure fresh perspectives. Of sixteen participants, 11 identified as male and 5 as female. In terms of age distribution, 7 participants were aged 18–29 and 9 were aged 30–45. Regarding experience with computer graphics systems, 5 participants reported expert-level experience, 9 had some experience (e.g., as users such as gamers), and 2 reported little or no prior experience. We took precautions to ensure that none of the new participants had been involved in the motion capture session used to create the dataset. Recruitment was conducted through email and social media platforms, reaching individuals from diverse backgrounds. To administer the experiment, we used the Qualtrics platform, similar to the previous study. Before participation, all volunteers were provided with an informed consent form, which they voluntarily signed, and also provided demographic details as requested.

Procedure

The experiment employed a within-subjects design, enabling participants to assess all combinations of independent variables: GAIT, SEX, BODY, and METHOD. Each trial featured animations representing 2 GAIT (Walk or Jog) $\times$ 2 SEX (Male or Female) $\times$ 4 BODY shapes (high BMI H1, H2, low BMI L1, L2) $\times$ METHOD motions (No COM or COM). We organized the experiment into two blocks based on *SEX* (Male and Female), with sub-blocks for *GAIT* (Walk and Jog). Block and animation sequences were randomized to ensure variability and prevent order effects.

Before each *SEX* block, participants viewed subjects' body shapes in increasing BMI order for familiarization. Four practice videos were shown before each main block, distinct from the actual experiment. In stimuli, the two *METHOD* motions were presented side-by-side on matching body shapes, allowing for direct motion comparison. The motion's position (left and right) was counterbalanced. Each gait block included 8 unique videos (4 *BODY* types $\times$ 2 sides) shown twice randomly, resulting in 16 questions per block. Video clips lasted 4 seconds each. In total, there were 4 blocks (2 *SEX* and 2 *GAIT*), making the experiment last approximately 12-15 minutes. Participants viewed video clips and selected the more natural motion between the left and right sides, as in 6.5. They indicated their choice (left/right) after viewing the clip at least once, enabling an informed decision before proceeding. An example stimulus is shown in Figure 6.9.

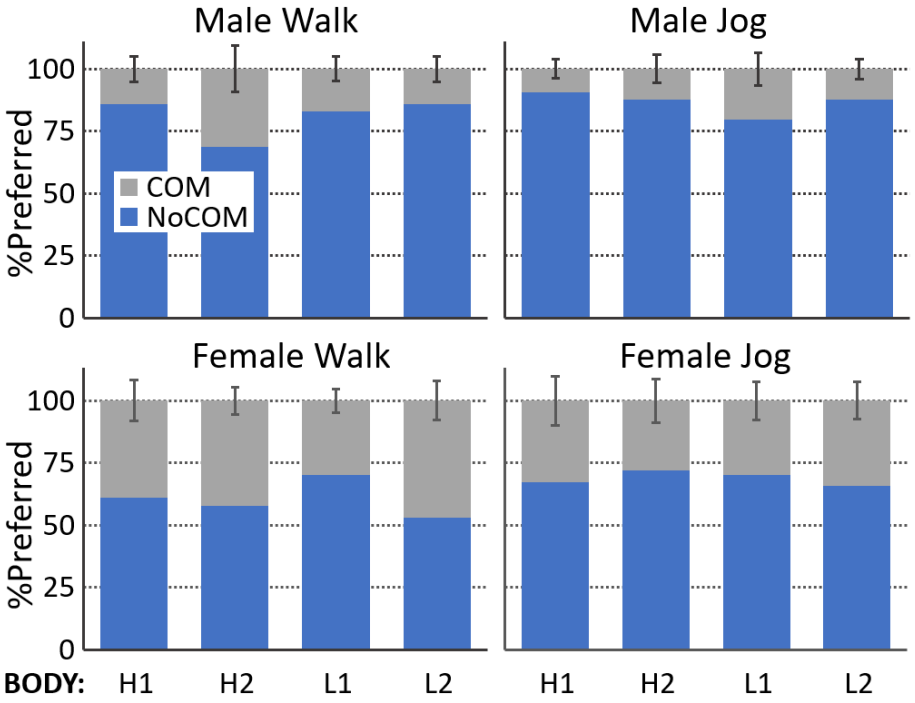


Figure 6.10: Experiment 2 Results: Percentage of time the No COM and COM motions were chosen.

6.5.2 Results

We conducted a repeated measures ANOVA to examine the effects of GAIT, SEX, BODY, and METHOD on participants' ratings of naturalness. The results are shown in Table 6.3. The main effect of METHOD, comparing the retargeted motion to the average motion, was found to be significant ($F(1,65) = 65.00$, $p < .001$, $\eta^2 = .81$, nc = 65.00, power = 1.00). Post-hoc analysis revealed that average motion received higher naturalness ratings compared to retargeted motion. However, there were no significant main effects of BODY or SEX on naturalness ratings.

Interestingly, a significant two-way interaction effect was observed between SEX and METHOD ($F(1,21.36) = 21.36$, $p < .001$, $\eta^2 = .59$, nc = 21.36, power = 0.99), indicating that the effect of body motions on naturalness ratings differed between male and female animations. Specifically, the COM method showed better performance for females compared to males, suggesting that the retargeted motion technique had a larger impact on males.

Additionally, there was also a significant two-way interaction effect between GAIT and METHOD ($F(1,6.51) = 6.51$, $p = .02$, $\eta^2 = .30$, nc = 6.51, power = 0.66). However, post-hoc analysis did not reveal a significant difference between the methods for walking and jogging motions. Overall, these results suggest that the choice of motion retargeting method influenced participants' ratings of naturalness,

Table 6.3: Experiment 2 ANOVA: significant effects for Method: NoCOM (NC), COM (C); Gait: Walk (W), Jog (J); and Sex (M, F), with effect size (partial η^2), non-centrality (nc) and observed power ($\alpha = 0.05$).

Effect	**SS**	**dof**	**MS**	**F-Test**	p	η^2	**nc**	**power**	**post-hoc (Tukey)**
Main Effects									
METHOD	297896	1	297895	65.00	< 0.001	0.81	65.00	1.00	$NC > C, p < 0.0005$
Two-way Interaction Effects									
SEX*METHOD	45942	1	45942	21.36	< 0.001	0.59	21.36	0.99	$NC*M > NC*F, p < .005$ $C*F > C*M, p < .05$
GAIT*METHOD	5981	1	5981	6.51	< 0.05	0.30	6.51	0.66	*not sig.*

with the average motion generally being perceived as more natural.

The main effect of METHOD significantly influences participants' ratings, indicating a substantial impact on their preferences. The average motion received higher naturalness ratings compared to the retargeted motion, as shown in Figure 6.10, aligning with Hoyet et al.'s previous research [HRZ+13], which found similar results regarding the attractiveness of average motions. The interaction between SEX and METHOD complicates the relationship between motion perception and physical properties. COM method performed better for females than males, suggesting differing effects of motion retargeting based on gender, potentially due to body proportions or movement patterns. Similarly, the GAIT and METHOD interaction highlights gait style's influence on ratings, but no significant differences emerged for walking and jogging motions in post-hoc analysis.

Runeson and Frykholm's research [RF83] confirms the ability to detect deceptive performances based on gender-atypical movements, emphasising human intuitions in assessing naturalness. Overall, the promising performance of the COM method for females suggests potential improvements, but it's essential to recognise the limitations of relying solely on inertial parameters and explore other factors in natural motion representation.

6.6 Discussion

In this chapter, we developed a physics-based framework called the *Balance Point Controller*, aimed at investigating motion consistency by modifying postural control parameters, particularly focusing on the COM. We validated the framework using

kinetic measures of locomotion, comparing them with human anthropometric data. Additionally, we designed experiment methodologies to evaluate the effectiveness of the framework.

In the first experiment, we generated quasi-average motions for each BMI category. The motions were aligned within the same BMI group, and an average representation of those motions was created. These quasi-average motions served as the input for our controller. Participants were asked to choose the most natural motion between the motion generated by the Balance Point Controller and the quasi-average motion. Results showed that the motions generated from the controller for high BMI characters performed worse for both walking and jogging across both male and female categories. The controller's motion did not perform well compared to the quasi-average motion, especially for higher BMI groups, suggesting a bias in participants' perception of motion style. This led us to conduct another experiment to address potential style bias by using the average of all motions, regardless of BMI.

The second experiment involved participants selecting the most natural motion between the Balance Point Controller's motion and the average motion from all the ground truth motion capture data for different BMI bodies. The average motions were perceived as more natural than retargeted motions. This aligns with the previous research by Hoyet et al. [HRZ+13] as they found that average motion is considered the most attractive and least distinctive. A significant interaction between character sex and the motion retargeting method revealed an intriguing relationship. The COM-based method seemed to perform better for females than males, possibly due to differences in body proportions or movement patterns. This finding underscores the importance of understanding sex-specific differences when evaluating naturalness in motion perception.

Experiment 2 demonstrated that modifying motions based solely on postural control parameters like COM is insufficient to ensure shape-motion consistency. It suggests that additional factors, such as motion style and human intuition about body shape, must be considered to enhance motion naturalness beyond relying on inertial properties. We collected participants' comments at the end of the experiments which highlighted recurring keywords, with "Speed" and "Arm movement" being the most frequently mentioned factors. Refer to Figure 6.11 for a detailed breakdown of comments.

While these studies advance our understanding of how virtual character motion is perceived, some limitations exist. First, the selection of actors with similar heights limited the variation in BMI, particularly for female actors in the Hi2 and Lo1 categories, potentially influencing the results. Future studies should address this by including a wider range of BMI and body shapes. Moreover, the research primarily

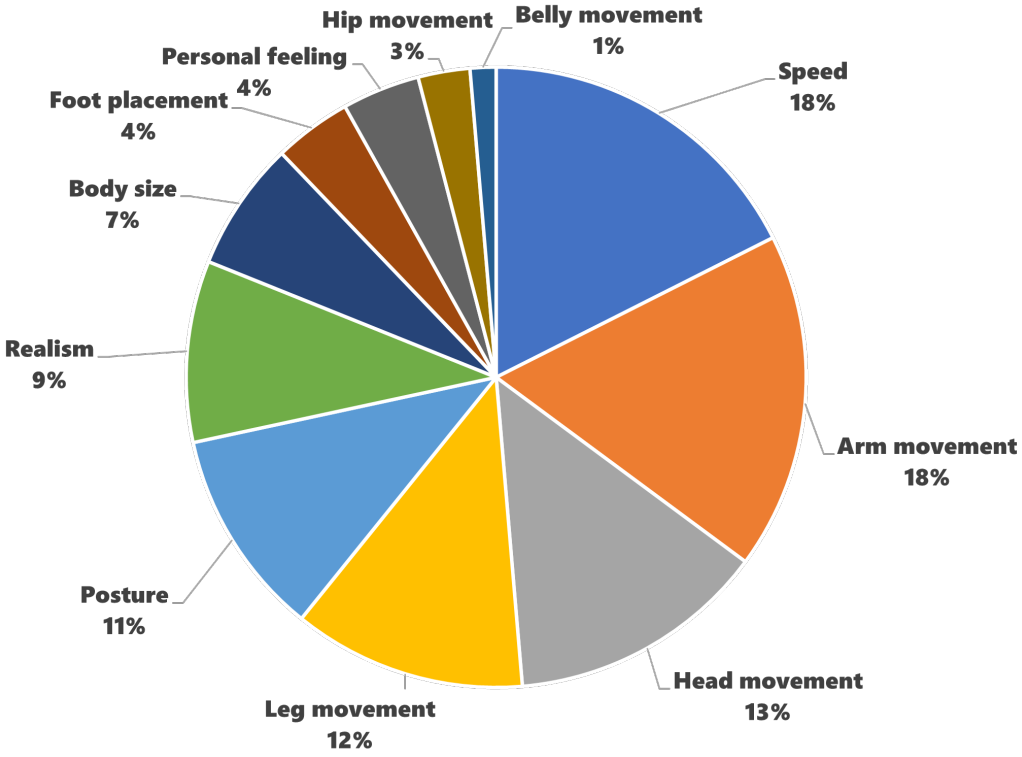


Figure 6.11: Factors mentioned in the comments and their frequencies.

focused on BMI-related body shape and motion consistency, without considering other physiological attributes like muscle, tendons, and fat, which could also play a crucial role.

Although our framework provides a useful starting point for generating more natural and realistic human motions, our findings suggest that additional factors such as individual variability, dynamic feedback, and cognitive elements may be necessary to create truly convincing virtual human motion. Participants' comments gathered at the end of the experiments highlighted recurring themes such as *Speed* and *Arm movement* as significant factors influencing their judgements. Other factors like *Head movement*, *Leg movement*, *Posture*, and *Realism* were also mentioned, indicating that these are key aspects participants pay attention to when evaluating character gait. The emphasis on speed and arm movement suggests that these elements play a vital role in shaping participants' perceptions of naturalness and consistency.

Future research could explore the influence of other factors on the perception of retargeted motion, such as gender, race, or age, to better understand the complexities involved in transferring motion across various body shapes. More sophisticated physics-based frameworks and machine learning techniques could also be developed to generate more natural-looking animations. Overall, our work highlights the need

to consider both body shape and motion style when creating realistic virtual characters. By further exploring the intricate relationship between body shape and motion perception, we can continue to improve the realism and effectiveness of virtual characters across various fields. To this end, we will make the complete dataset, including all motion and body shape data used in this research, openly accessible for further study.

In conclusion, the controller discussed in this section provides a solid foundation for further advancements. Some motions generated by the controller were comparable to the most attractive average motions. This suggests that they could be effectively used in scenarios where the focus is not on a single character. One such example is crowd simulation, where the introduction of physics-based characters could enhance the overall interactivity within the crowd. To further investigate the role of body shape in such setting, we conducted an experiment within crowd environments, which is described in the next chapter.

Chapter 7

Case Study II: How Body Shapes Change the Perception of Small-Scale Crowd Motions

This chapter builds on the findings of Chapter 3, where we explored how individual body shape influences perception in isolated avatars. Here, we extend that exploration to group contexts by examining how body shape diversity affects motion clone detection in small-scale crowds. This work further addresses Research Question 3 (see Section 1.1), which investigates how users perceive simulated motion in virtual environments.

The presence of realistic virtual avatars is a common feature in virtual environments, such as virtual reality (VR) games and simulations [MTT05]. However, achieving realism is more challenging when multiple avatars are present, such as in small-scale group or crowd settings. These scenarios are frequently encountered in everyday life, like people walking in parks or on sidewalks. With advancements in computer graphics and visual effects, these small groups are frequently recreated in virtual environments for use in movies and interactive applications. However, there is a need to strike a balance between realism and system performance. Researchers have explored various techniques to enhance the realism of virtual crowds while maintaining computational efficiency. This involves creating diversity in appearance such as body shape [SOWO17, SdSdSH19], texture [TLC02], colour [UCT04], and accessories [MYPT09]. In addition, motion realism has been enhanced through methods like motion blending [HK10] and secondary motion [HOKP16].

In ideal crowd scenarios, each virtual avatar should be distinct with unique movements to enhance realism. However, achieving this level of variety across all avatars is computationally expensive and inefficient for long-term applications. Therefore,

it is important to determine the level of variety needed for a natural-looking crowd. Extensive research has been conducted to develop techniques that introduce variety into crowd scenarios, particularly in crowd rendering [BPA16, DHOO05]. Another key aspect of crowd realism is the variety of 3D models used to depict virtual avatars. As noted by Shi et al. [SOWO17], shape and size are among the most significant and visually impactful features of any crowd representation. Out work explores the body shape feature of virtual avatars, specifically investigating how this feature influences perception in small-scale crowd scenarios. Previous research by McDonnell et al. [MLD+08] examined the identification of appearance clones using textures and coloured clothing. However, our focus is on assessing the impact of body shape on motion clones in the context of crowd perception.

In this study, we focus on two aspects of crowd realism: (1) body shape variation across avatars, and (2) motion diversity, manipulated through clone levels. We generated a set of physics-based avatars using a simple feedback control system, consistent with the Balance Point Controller framework introduced in Chapter 6, to simulate plausible walking motions. Using these avatars, we created crowd scenes with systematically varied numbers of unique motions and body shapes.

Our central hypothesis was that increasing body shape diversity would make it more difficult for users to detect motion clones, thus enhancing perceived realism. We also aimed to investigate whether the sex of the avatars affected clone detection accuracy.

In the sections that follow, we first discuss related work on body shape and crowd realism, then describe the experimental setup, stimuli generation, and participant procedure. We conclude with an analysis of results and a discussion of how this study contributes to a broader understanding of motion perception in virtual environments.

7.1 Body Shape and Crowd

The study of virtual avatar body shapes has also been an active research topic for many years, including both the reconstruction of 3D avatar body shapes and the perception of appearance. McDonnell et al. [MJH+07] investigated the factors influencing the perceived sex of virtual avatars and found that in the case of neutral walks, appearance plays a crucial role in determining the sex of an avatar. Similarly, Johnson and Tassinary [JT07] concluded that both appearance and motion significantly affect perceived attractiveness. These findings highlight the importance of appearance, particularly body shape, as a critical factor in perception. McDonnell et al. [MLH+09] further emphasized this by demonstrating that certain upper body parts, which are key to defining body shape, are especially influential in perception.

In terms of body modelling, researchers have developed statistical models using extensive collections of 3D body scans. We have already discussed some of these body models in detail; see Section 2.3. One of the most successful early models was SCAPE, introduced by Anguelov et al. [ASK+05], a deformable avatar body model. This was followed by Loper et al. [LMR+15b], who introduced a probabilistic body model that uses Principal Component Analysis (PCA) on body parameters derived from scanned real avatar data. The SMPL model, with its body shape and pose parameters, allows for the easy generation and animation of digital avatars with diverse body shapes.

Russell et al. [RSM+23] explored the perception of inconsistencies between body shape and motion by mismatching them, finding that participants could detect these inconsistencies during walking motions but struggled with other types of movements. The results from the experiment in Section 3.2 verified these findings, demonstrating that observers rated stimuli as more consistent when the body shape of an avatar matched the motion within the same BMI group. Our work builds on these studies by examining whether these results hold true within the context of motion clones and crowd animation.

However, not all research points to the ease of detecting inconsistencies. Kenny et al. [KMH+19] explored mismatches between body shape and motion in scenarios where virtual avatars interacted with objects in their environment. Their findings suggested that people could not identify inconsistencies in these complex interactions. This raises interesting questions about the conditions under which body shape and motion consistency are perceived and whether the dynamics of a crowd setting might influence this perception.

7.2 Experiment

This experiment extended our previous investigations on avatar body shape and motion perception (Chapter 3) to small-scale group settings. While earlier studies focused on isolated avatars, our goal here was to examine how body shape diversity affects the perception of motion clones within a crowd. Specifically, we asked whether varying the body shapes of avatars could mask or reveal motion similarity among characters performing the same motion.

To isolate perceptual effects from controller complexity, we used a simplified physics-based control framework. As established in Chapter 6, physics-based characters introduce dynamic, reactive motion aligned with real-world physical principles. However, instead of employing our Balance Point Controller or the more computationally intensive DRL framework (Chapter 5), we opted for a lightweight

feedback-based controller that closely mimics input reference motion. This allowed us to retain physical plausibility without introducing extra motion variation that might confound the experiment.

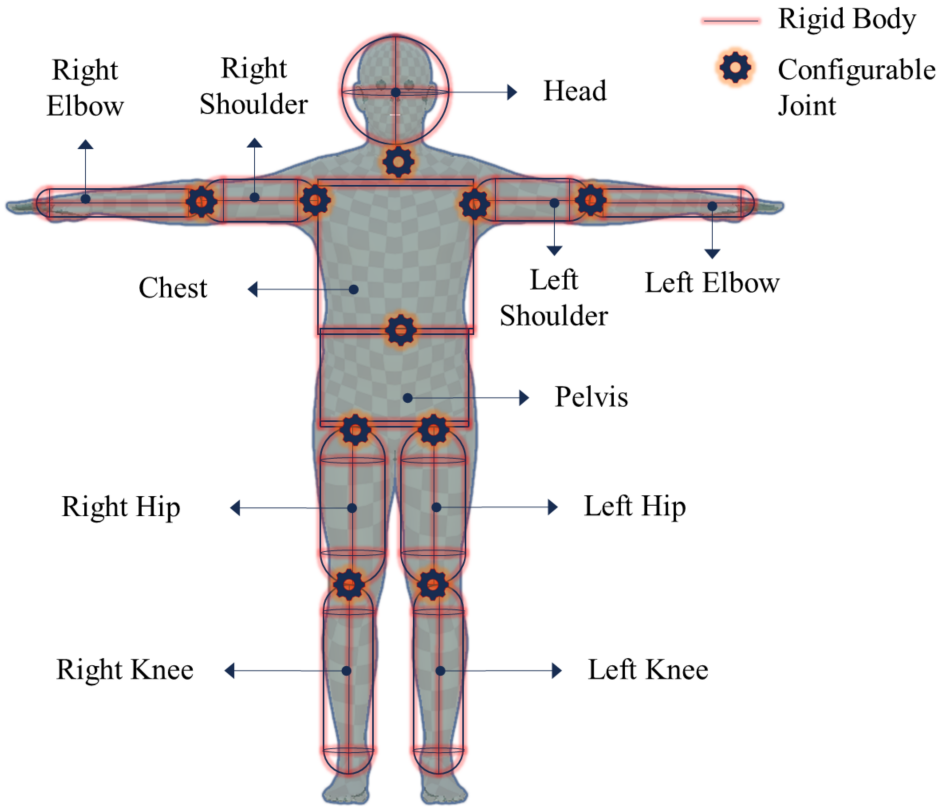


Figure 7.1: Ragdoll representation of an avatar with rigid bodies. This diagram shows the structure of a ragdoll created using primitive-shaped rigid bodies and configurable joints.

7.2.1 Physics-Based Model

As in Chapter 6, the avatar simulation was based on a dual-character setup: a *kinematic* character that served as the motion reference and a *dynamic* character (ragdoll) driven by physical forces. The ragdoll consisted of a hierarchy of rigid bodies (Figure 7.1) joined via Unity's configurable joints, enabling articulation through rotational and translational degrees of freedom.

Control was achieved using a Proportional-Derivative (PD) controller, a common method in physics-based character animation [ZH02, YLVdP07, SVCN+21]. For each joint, the controller minimized the deviation from the target pose using the equation:

$$\tau_{\mathbf{t}} = -k_p(\mathbf{q} - \bar{\mathbf{q}}) - k_d\dot{q}$$

Here, $\bar{\mathbf{q}}$ represents the desired (target) pose, $\mathbf{q}$ is the current joint pose, and $\dot{q}$ is the angular velocity. The stiffness (k_p) and damping (k_d) values were adapted from the Balance Point Controller to maintain stability. However, unlike in Chapter 6, we did not adjust the avatar's center of mass (COM), as our goal was to preserve reference motion fidelity rather than generate shape-specific adaptations.

7.2.2 Stimuli

To generate realistic stimuli for the perception study, we constructed small crowd scenes consisting of 12 virtual avatars. We selected 12 male and 12 female motion sequences from the dataset by Hoyet et al. [HRZ+13], which includes full-body motion capture and anthropometric metadata. We chose actors across a range of body weights and heights, stratified by Body Mass Index (BMI) into low, medium, and high groups. For each sex, four subjects were selected per BMI category to ensure diversity in both motion and shape, following a similar approach as used in Chapter 3.

As described in Section 3.2, the marker motion data was processed using MoSh++ [MGT+19b] to generate SMPL body meshes [PCG+19, RTB22], producing realistic avatars with consistent topology and deformable shape parameters. The resulting 3D meshes were animated using the bmlSUP Unity integration tool [BTT21] to enable playback of synchronized motion sequences.

The animation scenes were rendered in Unity using a clean white environment and consistent camera placement. To minimize bias from appearance realism and avoid the uncanny valley effect [MMK12, SWH18], we used a green checkerboard texture for all avatars. The aerial camera angle ensured all avatars were visible simultaneously, a strategy also used in prior crowd perception research [PO11, DMH+21, ANZ+21, KOOP11].

To discourage participants from focusing on fixed spatial locations, we introduced randomized avatar positions within the crowd for each stimulus. This spatial offsetting prevented participants from relying on positional memory to identify clone motions, thereby ensuring they attended to motion and shape features directly.

Baseline

The baseline scene consisted of virtual avatars where each character had a unique motion and body shape. The interplay between motion and body shape was critical, as the perception of a character's motion is influenced by its appearance, and certain motions may look different depending on the body shape performing them. As previous research has shown, humans are sensitive to even small discrepancies in

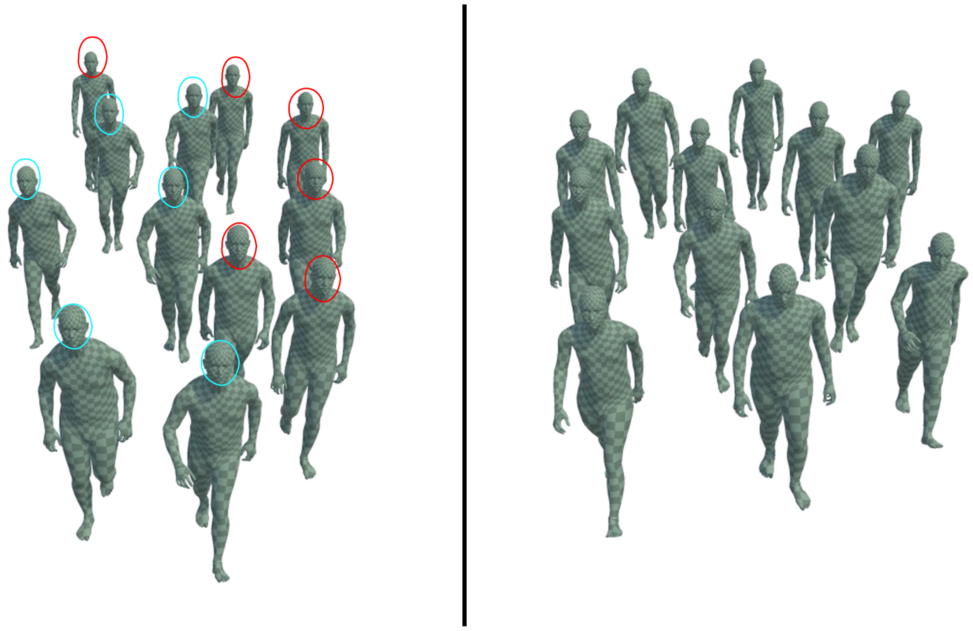


Figure 7.2: Stimuli Creation for the Experiment: **Motion Clone** (left): Featuring characters with two distinct motions. One motion is circled in green, and the other motion is in red (out-of-step motions). Each character has one of six different body shapes. **Baseline** (right): All characters have unique body shapes and are animated with unique motions.

virtual simulations [RSM+23]. To avoid bias, we randomized the combinations of body shapes and motions in the stimuli. None of the virtual avatars in the scene retained their original motion and body shape pairing, ensuring a fair comparison for motion clones.

The Fisher-Yates shuffle algorithm was used to randomize the lists of motions and body shapes. For each sequence, the shuffled list of motions was paired with the shuffled list of body shapes to assign the twelve characters. Unique sequences were generated by permuting the shuffled lists, and no repetitions were ensured by tracking previously used sequences. The experiment was designed to distribute all actors equally among the possible motions. The body shapes were divided into BMI sub-groups, reducing potential bias.

Motion Clone

The motion clone scenes featured virtual avatars with one or more motions applied to the twelve characters (see Figure 7.2). We explored four levels of motion (1 motion, 2 motions, 3 motions, and 6 motions). Additionally, we included different levels of body shape variation (1 body shape, 3 body shapes, 6 body shapes, and 12 body shapes). Therefore, each scene varied the number of unique motions and the

diversity of body shapes among the twelve virtual avatars.

To avoid bias from specific motion and body shape pairings, we randomized virtual avatars' body shape, motion, and position in each rendered scene. This approach ensured that each stimulus presented a unique combination of motions and body shapes among the twelve characters.

In cases where repeated motions were necessary due to lower motion levels (the maximum motion level possible was 6), we ensured that these repeated motions were out-of-step with each other to prevent the easy detection of motion clones. This resulted in a unique set of combinations for each stimulus.

7.3 Participants

Twenty-two participants from diverse backgrounds were recruited to perform the experiment online. Of these, 14 identified as male, 7 as female, and 1 preferred not to disclose their gender. The age distribution was as follows: 8 participants were aged 18–29, 12 were aged 30–45, 1 was aged 45–60, and 1 was over 60. In terms of relevant experience with computer graphics systems, 10 participants reported expert-level experience, 9 had some experience (e.g., as users such as gamers), and 3 reported little or no prior experience. Participants were recruited through internal mailing lists and social media platforms. None of these participants had been involved in the motion capture sessions used for stimuli creation. Participation was entirely voluntary, and each participant provided informed consent before participating in the experiment. They could provide demographic information and withdraw participation at any time during the experiment.

7.4 Procedure

Participants were presented with pairs of video clips showing crowd scenes side by side on their computer screens. On one side of the screen, each virtual avatar was animated using a unique synthesized motion, constituting the Baseline stimuli. On the other side, a clip from the motion clone set was displayed. Participants were tasked with detecting the motion clone by choosing the side where they believed all motions were different.

The experiment employed a mixed design to investigate the effects of the between-subjects factor *SEX* (Male and Female stimuli) and the within-subjects independent variables BODY (body shape level) and MOTION (motion clone level) on the detection of motion clones. Participants were divided into two groups based on *SEX*



Figure 7.3: Example Stimulus for Experiment: Participants were asked to indicate their confidence level regarding their decision on which side contains avatars with different motions. The full set of stimulus videos used in the experiment can be accessed via the QR code or at: https://youtu.be/dhj2ZqC7cTY

and were assigned to either the Male or Female group in a counterbalanced manner. Within each *SEX* group, there were combinations of 4 BODY levels (1 body shape, 3 body shapes, 6 body shapes, 12 body shapes) $\times$ 4 MOTION levels (1 motion, 2 motions, 3 motions, 6 motions), resulting in a total of 16 clip combinations.

As in the stimulus presentation, one side of the screen displayed the Baseline clip, while the other side displayed one of the motion clone clips. To control for side bias (left or right), the Baseline clip was shown on the left side for half of the motion clone clips (16 clips) and on the right side for the other half (16 clips). Participants were asked to click on the screen to start the videos, and both videos played simultaneously. This allowed participants to control the start of the video, but they could not pause or replay it. This design ensured that participants' choices

were based on their initial impressions and prevented them from intentionally finding patterns by pausing or replaying the videos. As already mentioned, to eliminate any familiarity effects related to the position or combination of motion and body shapes, the 32 motion clone clips had unique combinations for both sides. Within each *SEX* group, the video pairs were randomized for each participant.

At the start of the experiment, participants were shown example stimuli featuring crowds with different motions and identical motions. This was followed by a series of practice questions to familiarize participants with the task. These examples featured actors with similar body shapes and motion combinations as in the actual stimuli, but none of these actors' body shapes and motions were used in the main experiment.

The experiment consisted of 16 clips with the Baseline on the left and 16 with the Baseline on the right, with each set being repeated twice. This resulted in a total of 64 questions. Each video clip lasted 6 seconds. Participants were asked to watch the video clips and select the side they believed had all different motions. The answer options were presented on a 5-point Likert scale, allowing participants to express their confidence level: 'Definitely left', 'Probably left', 'Undecided', 'Probably right', and 'Definitely right'. An example stimulus for the experiment is shown in Figure 7.3. After selecting an answer, participants pressed the 'Next' button, which only appeared after the clip had finished playing. The total duration of the experiment was approximately 12-15 minutes on average.

7.5 Results

Figure 7.4 shows the mean ratings and standard deviations for participants' ability to detect motion clones across varying combinations of motions (1, 2, 3, and 6) and body shapes (1, 3, 6, and 12). Participants were asked to choose the side where they believed all the motions were different. The higher ratings indicate a lower detection rate of motion clones, and lower ratings reflect more accurate detection.

For both male and female avatars, no consistent pattern was found between body shape diversity and motion clone detection. This suggests that increasing the variety of body shapes did not always make it harder or easier to detect motion clones. For female avatars, the highest rating occurred when there was one motion and 12 body shapes (M = 3.682, SD = 0.791). This indicates the lowest ability to detect motion clones with one motion and 12 body shapes. On the other hand, the lowest rating for female avatars was observed with six motions and three body shapes (M = 2.682, SD = 0.852), where participants were better at detecting the clones.

In the case of male avatars, the highest rating, signifying the lowest clone detection accuracy, was recorded for one motion and one body shape (M = 3.818, SD =

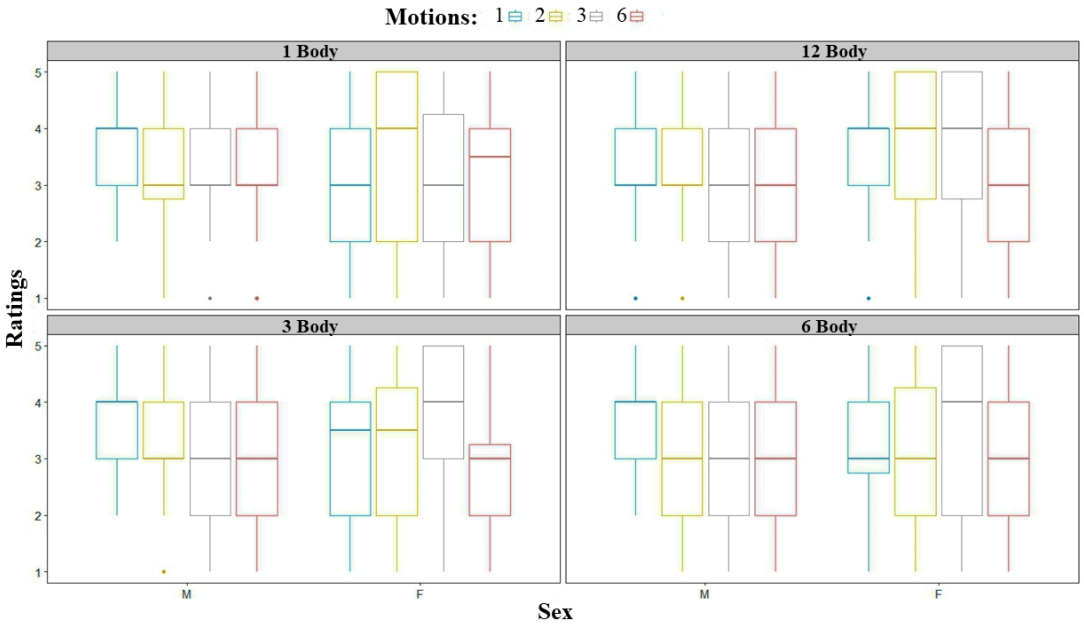


Figure 7.4: Mean Ratings from Participants: The boxplots represent the participants' mean ratings (with standard deviations) for their ability to detect stimuli with different motions under varied conditions of motion (1, 2, 3, and 6 motions) and body shapes (1, 3, 6, and 12 shapes), separated by sex (M = Male, F = Female). Higher ratings indicate lower motion clone detection accuracy, while lower ratings correspond to better performance in identifying distinct motions.

0.633). The lowest rating was found with three motions and 12 body shapes (M = 2.932, SD = 0.613), indicating better detection. These results show the complexity of how motion and body shape variety influence participants' perception of motion clones, with no clear linear relationship between them.

A repeated measures ANOVA was conducted to examine the effects of sex, number of motions, number of body shapes, and their interactions on participants' ratings of crowd realism. Table 7.1 provides the detailed results of a mixed-design ANOVA conducted to evaluate the effects of the independent variables: MOTION, BODY, and SEX on participants' ratings to detect motion clones. The analysis considered one-way, two-way, and three-way interaction effects.

The main effect of motion variety approached significance ($F(2.09, 41.85) = 3.05$, $p = 0.056$, $\eta^2 = 0.039$), indicating that as the variety of motions increased, participants were more likely to detect differences between the virtual avatars. Although the effect was not significant, the data suggests that motion is a key factor in creating believable and realistic crowd animations. This means greater motion diversity helps to reduce the detection of motion clones, making the virtual scene appear more natural.

In contrast, body shape diversity did not show a significant effect. This suggests

Table 7.1: ANOVA results for the experiment showing interaction effects of MOTION, BODY, and SEX on motion clone detection. The table presents degrees of freedom (dof), mean squares (MS), F-test values, effect sizes (η^2), and p-values for one-way, two-way, and three-way interaction effects.

Effect	**dof**	**MS**	**F-Test**	η^2	**p-value**
One-way Interaction Effects					
MOTION	2.09, 41.85	1.18	3.05	0.039	0.056
BODY	2.50, 50.08	0.31	0.47	0.002	0.668
SEX	1, 20	3.34	0.01	<0.001	0.937
Two-way Interaction Effects					
MOTION x BODY	5.27, 105.32	0.50	0.73	0.010	0.610
MOTION x SEX	2.09, 41.85	1.18	1.56	0.020	0.221
SEX x BODY	2.50, 50.08	0.31	1.72	0.007	0.182
Two-way Interaction Effects					
MOTION x BODY x SEX	5.27, 105.32	0.50	1.41	0.020	0.225

that variations in avatar body shapes did not strongly affect participants' ability to distinguish between unique and cloned motions. This finding indicates that visual elements like body shape alone may not be sufficient to change how people perceive motion realism in crowds. It also implies that while body shape diversity is important for making avatars visually distinct, it may not play a major role in detecting cloned movements compared to motion variety.

Similarly, the sex of the virtual avatars did not significantly impact clone detection. This means that whether the avatars were male or female had no influence on how participants identified motion clones. There were also no significant interaction effects between motion variety, body shape diversity, and avatar sex. This indicates that combining these factors did not affect the detection of motion clones either.

Overall, the results suggest that motion variety is far more important than body shape or sex in detecting motion clones in virtual crowds. While body shape and sex help with the overall visual realism, they do not significantly affect how people perceive motion clones. Motion diversity, on the other hand, plays a critical role in creating believable virtual environments.

7.6 Discussion

In this study, we explored the impact of body shape diversity on the detection of motion clones in the context of small-scale virtual crowds. We also focused on dynamically controlled virtual avatars for the stimuli. Based on the results from the

experiment in Chapter 3 on the role of body shape in virtual avatars, we initially hypothesised that greater diversity in body shape would make it more difficult to detect motion clones.

While the previous research has also highlighted the importance of motion variation in crowd simulations [ANZ+21], our study specifically investigated whether the body shape of virtual avatars affects the perception of crowd realism when different numbers of motion clones are used. The experiment was designed to simulate a variety of body shapes and motions, with virtual avatars categorized by Body Mass Index (BMI) and varying levels of motion and body diversity.

The results showed that neither the body shape nor sex had a significant effect on participants' ability to detect motion clones. Contrary to our hypothesis, body shape diversity did not significantly affect participants' ability to detect motion clones. This aligns with previous findings by Hoyet et al. [HOKP16] and Pravzak et al. [PO11], who found that motion variety had a greater perceptual impact than other visual features. Nonetheless, our results do not confirm this, and care should be taken in drawing conclusions beyond the scope of clone detection.

Although motion variety did not reach statistical significance, it showed a trend toward influencing clone detection accuracy. This suggests that with a larger sample size or a different experimental design, a clearer effect may emerge. Therefore, while our data do not conclusively show that motion variety plays a more important role than body shape, they do indicate a possible perceptual sensitivity to motion differences. Future work should investigate this further using additional or complementary measures.

After the experiment, we asked participants to provide some comments on what basis they made the decisions. The chart in Figure 7.5 illustrates the distribution of some key factors mentioned in the comments. "Arm/Hand movement" emerged as the primary factor. This could be attributed to a particular motion in female stimuli with a distinct arm swing. This observation may help explain the variation in detection rates across different conditions, although more controlled follow-up studies would be needed to isolate such effects.

Other important factors included "Speed," "Head movement," and "Shoulder movement." These comments give us a better understanding of what participants noticed during the experiment. By keeping these factors in mind, future designs of crowd scenarios can be made better.

Future research should also explore additional variables, such as clothing or facial features, that might interact with motion in forming crowd perception. Additionally, measuring perceived realism directly, alongside clone detection, would help determine whether and how visual diversity impacts overall crowd believability. Larger-

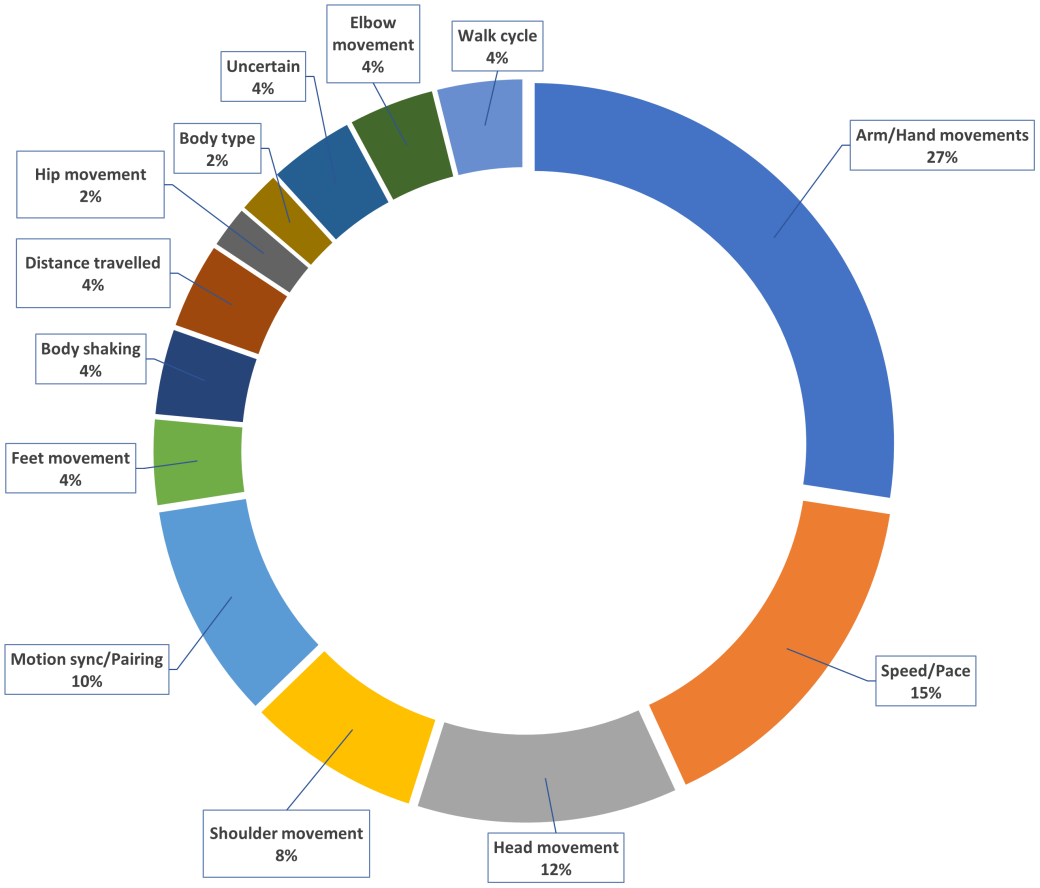


Figure 7.5: Distribution of key factors considered by participants when distinguishing between crowd motions in the experiment.

scale studies with more varied task demands and longer exposure durations could also clarify the interaction between visual and motion-based cues in complex crowd environments.

In conclusion, this study provides preliminary insight into how body shape and motion diversity may influence perception in virtual crowds. While body shape did not significantly affect clone detection, motion variety showed a trend that warrants further investigation. These findings suggest that clone detection may be more sensitive to motion-related cues than visual diversity alone, but future work is needed to confirm this and assess broader notions such as realism or plausibility in crowd behaviour.

Chapter 8

Conclusion

In this thesis, our aim was to explore how body shape influences the perception of motion in virtual avatars and to use our findings to guide the simulation of virtual character motion that is perceived to be natural. Essentially, this work represents an effort to reflect the diversity of real human experiences in virtual environments. In the introductory chapter, we outlined several research questions guiding this exploration. In this concluding chapter, we summarise the key findings, highlight the limitations encountered during our investigations, and propose potential directions for future research. We will now revisit the main themes of the thesis, reflecting on how each contributes to the understanding of avatar body shape and motion perception in virtual scenarios.

8.1 Avatar Body & Motion

Research Question 1: ***How do physiological factors, such as body shape, influence the motion perception of virtual avatars?***

The first research question was aimed at understanding the influence of physiological factors, particularly body shape, on how users perceive the motion of virtual avatars. We conducted two main studies to investigate the relationship between avatar body shape and motion perception, focusing on how inconsistencies between body shape and motion are perceived and how avatar embodiment influences user experience.

In the first study (Chapter 3), participants were asked to rate the consistency between avatar body shape and motion using locomotion (walking and jogging) as the motion type. We hypothesised that users would rate perceived consistency higher when the avatar's body shape aligns with that of the actor from whom the motion was captured. The results revealed that while a minority of participants consistently rated aligned body shapes and motions as more natural, body shape

played a significant role in how participants perceived these motions. Participants rated avatars with body shapes within the same BMI category as the actor in motion more favourably, indicating that consistency between body shape and movement is important for the perception of naturalness. The study also highlighted the influence of perceived attractiveness, with specific body shapes consistently receiving higher ratings, suggesting sexual dimorphism might play a role in motion perception.

The other study (Chapter 4) focused on embodiment in virtual reality (VR), where we examined how different levels of avatar self-representation (e.g., full-body avatar, partial-body avatar, without-body avatar representations) influenced the sense of embodiment in a virtual environment. Participants reported higher levels of embodiment, particularly ownership and agency, when using representations with visible hands, while the absence of any avatar diminished these feelings. Interestingly, no significant results were found concerning self-location, indicating that avatar presence did not strongly affect participants' sense of location within the virtual space. However, full-body avatar representation did not yield any significant improvements over partial-body avatar representation. These results suggest that users prioritise controllability over a complete avatar representation in task-based virtual environments.

Overall, the findings demonstrated that physiological factors like body shape significantly influence motion perception from a viewer's perspective. The alignment between body type and movement plays a key role in how natural or consistent avatars are perceived to be. These insights have practical applications such as gaming and virtual social spaces, where creating natural, personalised avatars can enhance user immersion.

8.2 Motion Controller

Research Question 2: ***Can physics-based simulations be used to generate variations in motion based on differences in body shape?***

For the second research question, we investigated the feasibility of using physics-based simulations to generate variations in motion that occur due to differences in body shape. We developed two controllers, a Body-Parameterised Motion Controller and a Balance Point Controller. These controllers simulate body-specific motions and create realistic and diverse avatar behaviours.

The Body-Parameterised Motion Controller is a DRL-based controller. It generates motions by adjusting physical attributes such as mass and limb size, allowing the simulation of realistic locomotion for avatars with varying body shapes. The con-

troller was trained using a reward system that balanced imitation and regularisation rewards, ensuring that the generated motions adhered to natural human movement patterns. Our findings indicated that mass played a critical role in creating motion diversity. The DRL framework was able to produce realistic and varied motions, with comparisons showing that the system effectively captured differences in body shape, as validated by metrics like gait parameters and motion dynamics.

Despite the promising results, the DRL controller faced some limitations. The integration of the framework into game engines was challenging due to its complexity. On the other hand, a simplified modelling of a physics-based avatar relies on various assumptions for simulation. This compromises the quality of generated motion and overall performance accuracy. These limitations point to areas for improvement, such as simplifying the framework for real-time applications and a biomechanically accurate model for simulating motions.

In addition to the DRL controller, we developed the Balance Point Controller, which adapted input motion based on an avatar's height and weight by adjusting the COM of limbs and the overall body. This controller produced motions closely aligned with real human gait patterns, proving lightweight and adaptable for real-time use in gaming and VR. However, it relied solely on inertial parameters, such as the COM, thereby limiting its ability to account for other physiological factors that could influence motion, such as joint flexibility or muscle strength. Future iterations of this controller could benefit from incorporating a wide range of parameters to improve motion realism and adaptability.

Overall, both controllers contributed to the field by demonstrating the potential of physics-based simulations in generating realistic, body-specific motions for virtual avatars. However, there remains room for refinement and further work through technological improvements.

8.3 Perception of Simulated Motion

Research Question 3: ***How are the resulting variations in human motion perceived by users in virtual environments?***

The third research question was focused on evaluating how users perceive simulated motions. We presented an evaluation framework using motions generated by the Balance Point Controller. The framework was used to evaluate the naturalness of these simulated motions through user feedback.

In the first study, participants were asked to compare motions generated by the Balance Point Controller with average motions that had previously been identified

as most attractive and least distinctive [HRZ+13]. The results showed that average motions were consistently rated as more natural, aligning with the previous research. However, the Balance Point Controller performed better for female avatars than male avatars, potentially due to differences in body proportions and movement patterns. The results also highlight that modifying motions based on postural control parameters alone does not guarantee shape/motion consistency, which suggests the need to consider additional factors and control parameters

Our second study was focused on the perception of small-scale crowds composed of avatars with varying body shapes. Participants were tasked with choosing, from a pair of videos, the one in which they believed all motions were from different actors, essentially testing their ability to detect motion clones. Contrary to our hypothesis, we found no significant effect of body shape on participants' ability to detect motion clones. However, the level of motion variety reached near-significant results. This suggests that motion characteristics, rather than body shape, play a more significant role in clone detection.

The results aligned with previous research findings that motion variety has a bigger impact on perception than visual elements, in this case, body shape. Nonetheless, these dynamically controlled physics-based avatars can be used in crowd simulation, adding reactive and responsive elements in the form of secondary motions, such as shoulder motion. However, body shape diversity alone is insufficient to improve crowd realism.

8.4 Limitations

Several limitations became apparent throughout this research, affecting both the scope of the findings and the generalizability of the results. The most significant limitation is the lack of diversity in the stimuli from the selected dataset. In particular, the avatars' body shapes, sexes, and ethnicities were not diverse enough. While the studies aimed to explore the role of body shape in motion perception, the limited representation of female avatars and the relatively homogeneous nature of the stimuli restricted the ability to generalise findings across different body types. This is especially important when considering that body shape plays a central role in how users perceive motion, and factors such as ethnicity, gender, and age may also influence perception in ways not captured by our experiments.

Another limitation involves the scope of the perceptual studies presented in Chapter 3. While the research explored how participants rated consistency between body shape and motion, as well as embodiment in virtual environments, the perceptual studies were limited by the types of motion analysed. The focus on

locomotion (e.g., walking and jogging) was necessary to maintain control over the experiment. Still, this small subset of possible human motions does not capture the full range of motion typically used for virtual avatars. The exploration of complex or dynamic movements could further highlight differences in body shape perception. Furthermore, body shape in this work was primarily defined using BMI. While BMI offers a convenient and widely used metric for categorising body size, it has important limitations. It does not account for variations in muscle mass, fat distribution, or bone density. For example, individuals with high muscle mass, such as athletes, may be classified as overweight or obese despite having low body fat levels. Likewise, two individuals with identical BMI values may have substantially different body compositions [RCSSJ+08]. These limitations are especially relevant in our context, as virtual avatars generated solely based on BMI may not accurately reflect the diversity of real human body shapes. Future work could incorporate additional physiological descriptors or body composition metrics to improve anatomical realism and perceptual validity. In the study of controller evaluation, presented in Chapter 6, the duration of the video stimuli may have constrained participants' ability to fully engage with and assess the motions. Some participants noted that the short display duration of videos made it difficult to form a complete judgement about the naturalness of the motion, which could have impacted the reliability of their ratings.

Technical limitations also emerged in the work on motion controllers, described in Chapters 5 and 6, particularly regarding the implementation of the Body-parameterized Controller and Balance Point Controller. While both controllers were capable of generating varied, realistic motions based on body shape, integrating them into real-time applications like game engines was challenging due to the complexity and computational demands of these models. The Balance Point Controller, although lightweight and compatible with some systems, relied only on inertial parameters like the COM and did not incorporate any other physiological factors, such as joint flexibility or muscle dynamics. These limitations reduce the adaptability of the controllers in producing a wider range of realistic motions.

Furthermore, in the study on avatar self-representation in VR, presented in Chapter 3, some technical limitations emerged. The mismatched positional tracking between the HMD and motion capture suit could have impacted the embodiment experience for participants. This was also clear from the non-significant results for Self-Location ratings. A more improved combination of these systems could provide a much better understanding of the role of body shape in avatar representation on the feeling of embodiment.

Lastly, the findings related to motion clone detection in crowds, presented in Chapter 7, may have been influenced by the experimental design. Similar to the

studies presented in Chapter 6, one of the limitations of this study was again the short duration of video stimuli. This may have restricted participants' ability to thoroughly assess and compare motions. Future studies should consider extending the duration of stimuli or altering the presentation method to allow for a more in-depth analysis of motion perception. Additionally, providing fixed starting positions for the avatars in the crowd, rather than changing their positions in each stimulus, could potentially provide more controlled results for future investigations.

8.5 Future Work

Building on the findings and limitations of this thesis, there are several promising avenues for future research that could expand and refine our understanding of body shape, motion perception, and physics-based simulations in virtual environments. Incorporating stimuli from a more diverse dataset would help in a better understanding of the relationship between body shape and motion perception. Future studies should include a wider range of body shapes, ethnic backgrounds, and gender identities to better understand how these factors influence motion perception. By increasing this diversity of avatars used in perceptual studies, findings would be more inclusive and representative of a broader audience.

In addition to expanding the range of avatars, future work should explore a broader variety of motions. While locomotion, i.e., walking and jogging, was an effective starting point, more complex actions, such as jumping, running, or dancing, could provide further insights into how body shape influences the perception of dynamic, high-energy movements. This would not only enhance our understanding of motion perception, but also broaden the applicability of the findings to various domains such as gaming, virtual training, and rehabilitation, where a wide range of human movements are often simulated.

For the technical aspects, future work could focus on improving the efficiency and scalability of the Body-parameterised Motion and Balance Point controllers. Simplifying the DRL framework to make it more accessible for integration into real-time applications, such as video games and VR environments, would significantly enhance its utility. Additionally, exploring alternative control systems beyond the basic PD controller used in Chapter 6 could lead to more adaptive and natural motion simulations. Integrating more advanced biomechanical models or machine learning techniques could allow for more responsive avatars that adjust their movements dynamically based on changes in the virtual environment or user input.

Moreover, future studies should incorporate more complex physiological parameters into the simulations. Factors such as muscle dynamics, joint flexibility, and

external forces (e.g., gravity, friction) could be included to create even more realistic and nuanced avatars. Incorporating these elements would help bridge the gap between simulated and real human motion, thus enhancing the believability of virtual avatars and increasing user immersion.

Another key area for future research is further perceptual evaluation studies to assess how users perceive motions generated by physics-based controllers, especially the DRL framework, to generate valuable feedback on the effectiveness of these systems. Understanding how users judge the naturalness and believability of these motions in different contexts will be essential for further refining motion generation techniques.

Finally, future research should also explore the role of gender and other demographic factors in motion perception. The gender-based differences observed in the perception of naturalness in female avatars suggest that physiological and social factors may play a more significant role than previously thought. Investigating how these differences manifest across diverse user groups could lead to the development of more personalised and adaptive virtual avatars that better cater to individual needs and preferences.

By addressing these areas, future research can build on the contributions of this thesis to further enhance the realism and diversity of virtual avatars, pushing the boundaries of what is possible in virtual environments.

8.6 Final Remarks

In conclusion, this thesis has made significant contributions to the understanding of how body shape influences motion perception and how physics-based simulations can generate these variations. By addressing the three research questions, we have provided valuable insights into avatar realism, and motion generation in virtual environments. While the work presented here offers foundational knowledge, there are many exciting opportunities for future research that could further refine and expand upon these findings. As virtual environments continue to grow in popularity and application, enhancing avatar realism and motion perception will remain critical for creating immersive, believable, and personalised experiences for users.

Bibliography

[AA00] Thomas P Andriacchi and Eugene J Alexander. Studies of human locomotion: past, present and future. *Journal of biomechanics*, 33(10):1217–1224, 2000.

[AAKC13] Rami Ali Al-Asqhar, Taku Komura, and Myung Geol Choi. Relationship descriptors for interactive motion adaptation. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 45–53, 2013.

[ABRB+18] Mazen Al Borno, Ludovic Righetti, Michael J Black, Scott L Delp, Eugene Fiume, and Javier Romero. Robust physics-based motion retargeting with realistic body shapes. In *Computer Graphics Forum*, volume 37, pages 81–92. Wiley Online Library, 2018.

[AG85] William W Armstrong and Mark W Green. The dynamics of articulated rigid bodies for purposes of animation. *The visual computer*, 1:231–240, 1985.

[AMH01] Amr Ahmed, Farzin Mokhtarian, and Adrian Hilton. Parametric motion blending through wavelet analysis. In *Eurographics (Short Presentations)*, 2001.

[AMY08] Daijiro Abe, Satoshi Muraki, and Akira Yasukouchi. Ergonomic effects of load carriage on energy cost of gradient walking. *Applied ergonomics*, 39(2):144–149, 2008.

[ANZ+21] Robin Adili, Benjamin Niay, Katja Zibrek, Anne-Hélène Olivier, Julien Pettré, and Ludovic Hoyet. Perception of motion variations in large-scale virtual human crowds. In *Proceedings of the 14th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pages 1–7, 2021.

[AOG77] TP Andriacchi, JA Ogle, and JO Galante. Walking speed as a basis for normal and abnormal gait measurements. *Journal of biomechanics*, 10(4):261–268, 1977.

[ARC22] Eduardo Alvarado, Damien Rohmer, and Marie-Paule Cani. Generating upper-body motion for real-time characters making their

way through dynamic environments. In *Computer Graphics Forum*, volume 41, pages 169–181. Wiley Online Library, 2022.

[ASK+05] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005.

[AT00] Amaury Aubel and Daniel Thalmann. Realistic deformation of human body shapes. In *Computer Animation and Simulation 2000: Proceedings of the Eurographics Workshop in Interlaken, Switzerland, August 21–22, 2000*, pages 125–135. Springer, 2000.

[Ayy97] Ed Ayyappa. Normal human locomotion, part 1: Basic concepts and terminology. *JPO: Journal of Prosthetics and Orthotics*, 9(1):10–17, 1997.

[BAP+17] Ilias Bergström, Sérgio Azevedo, Panos Papiotis, Nuno Saldanha, and Mel Slater. The plausibility of a string quartet performance in virtual reality. *IEEE transactions on visualization and computer graphics*, 23(4):1352–1359, 2017.

[BBG08] Jeremy N Bailenson, Jim Blascovich, and Rosanna E Guadagno. Self-representations in immersive virtual environments 1. *Journal of Applied Social Psychology*, 38(11):2673–2690, 2008.

[BC98] Matthew Botvinick and Jonathan Cohen. Rubber hands 'feel'touch that eyes see. *Nature*, 391(6669):756–756, 1998.

[BCHF19] Kevin Bergamin, Simon Clavet, Daniel Holden, and James Richard Forbes. Drecon: data-driven responsive control of physics-based characters. *ACM Transactions On Graphics (TOG)*, 38(6):1–11, 2019.

[BF50] B Bresler and JP Frankel. The forces and moments in the leg during level walking. *Transactions of the American Society of Mechanical Engineers*, 72(1):27–36, 1950.

[BFC+09] Jeanne H Brockmyer, Christine M Fox, Kathleen A Curtiss, Evan McBroom, Kimberly M Burkhart, and Jacquelyn N Pidruzny. The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of experimental social psychology*, 45(4):624–634, 2009.

[BH00] Matthew Brand and Aaron Hertzmann. Style machines. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 183–192, 2000.

[Bio97] Frank Biocca. The cyborg's dilemma: Progressive embodiment in virtual environments. *J. Computer-Mediated Communication*, 3, 01 1997.

[BK07] Raymond C Browning and Rodger Kram. Effects of obesity on the biomechanics of walking at different speeds. *Medicine & Science in Sports & Exercise*, 39(9):1632–1641, 2007.

[BKFK20] Jennifer Brade, Alexander Kögel, Christian Fuchs, and Philipp Klimant. Impact of first person avatar representation in assembly simulations on perceived presence and acceptance. In *VISIGRAPP (1: GRAPP)*, pages 17–24, 2020.

[BKL+16] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J. Black. Keep it smpl: Automatic estimation of 3d human pose and shape from a single image, 2016.

[BM09] Olaf Blanke and Thomas Metzinger. Full-body illusions and minimal phenomenal selfhood. *Trends in cognitive sciences*, 13(1):7–13, 2009.

[BPA16] Alejandro Beacco, Nuria Pelechano, and Carlos Andújar. A survey of real-time crowd rendering. In *Computer Graphics Forum*, volume 35, pages 32–50. Wiley Online Library, 2016.

[BPCS+11] Janusz Blaszczyk, Michal Plewa, Joanna Cieslinska-Swider, Bogdan Bacik, Barbara Zahorska-Markiewicz, and Andrzej Markiewicz. Impact of excess body weight on walking at the preferred speed. *Acta neurobiologiae experimentalis*, 71(4):528–540, 2011.

[BRRP97] Bobby Bodenheimer, Chuck Rose, Seth Rosenthal, and John Pella. The process of motion capture: Dealing with the data. In *Computer Animation and Simulation'97: Proceedings of the Eurographics Workshop in Budapest, Hungary, September 2–3, 1997*, pages 3–18. Springer, 1997.

[BTT21] Adam O Bebko, Anne Thaler, and Nikolaus F Troje. bmlsup–a smpl unity player. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 573–574. IEEE, 2021.

[BWBM20] Jean Basset, Stefanie Wuhrer, Edmond Boyer, and Franck Multon. Contact preserving shape transfer: Retargeting motion from one shape to another. *Computers & Graphics*, 89:11–23, 2020.

[BWF02] Sarah-Jayne Blakemore, Daniel M Wolpert, and Christopher D Frith. Abnormalities in the awareness of action. *Trends in cognitive sciences*, 6(6):237–242, 2002.

[CK77] James E Cutting and Lynn T Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society*, 9:353–356, 1977.

[CL80] Peter R Cavanagh and Mario A Lafortune. Ground reaction forces in distance running. *Journal of biomechanics*, 13(5):397–406, 1980.

[Cla15] David J Clark. Automaticity of walking: functional significance, mechanisms, measurement and rehabilitation strategies. *Frontiers in human neuroscience*, 9:246, 2015.

[CMCY69] Charles E Clauser, John T Mc Conville, and John W Young. Weight, volume, and center of mass of segments of the human body. Technical report, National Technical Information Service: Springfield, VA, USA, 1969.

[CMM+18] Nuttapong Chentanez, Matthias Müller, Miles Macklin, Viktor Makoviychuk, and Stefan Jeschke. Physics-based motion capture imitation with deep reinforcement learning. pages 1–10, 11 2018.

[CMU] CMU. Carnegie mellon graphics lab motion capture database. `http://mocap.cs.cmu.edu/`. 2000.

[CTZ76] Giovanni A Cavagna, H Thys, and A Zamboni. The sources of external work in level walking and running. *The Journal of physiology*, 262(3):639–657, 1976.

[Dar72] Charles Darwin. *The expression of the emotions in man and animals*. London: John Murray. 1st edition, 1872.

[Dem55] Wilfrid Taylor Dempster. The anthropometry of body action. 1955.

[DHOO05] Simon Dobbyn, John Hamill, Keith O'Conor, and Carol O'Sullivan. Geopostors: a real-time geometry/impostor crowd rendering system. In *Proceedings of the 2005 symposium on Interactive 3D graphics and games*, pages 95–102, 2005.

[dL96] Paolo de Leva. Adjustments to zatsiorsky-seluyanov's segment inertia parameters. *Journal of Biomechanics*, 29(9):1223–1230, 1996.

[DL04] Ruth Dickstein and Yocheved Laufer. Light touch and center of mass stability during treadmill locomotion. *Gait & posture*, 20(1):41–47, 2004.

[DM88] P Das and G McCollum. Invariant structure in locomotion. *Neuroscience*, 25(3):1023–1034, 1988.

[DMH+21] Beatríz Cabrero Daniel, Ricardo Marques, Ludovic Hoyet, Julien Pettré, and Josep Blat. A perceptually-validated metric for crowd trajectory quality evaluation. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 4(3):1–18, 2021.

[DPPSRJ12] Hannah Del Porto, Celia Pechak, Darla Smith, and Rebecca Reed-Jones. Biomechanical effects of obesity on balance. *International Journal of Exercise Science*, 5(4):301–320, 2012.

[Dur08] Jennifer L Durkin. Measurement and estimation of human body segment parameters. In *Routledge Handbook of Biomechanics and Human Movement Science*, pages 197–213. Routledge, 2008.

[EC17] J. Favre E.F. Chehab, T.P. Andriacchi. Speed, age, sex, and body mass index provide a rigorous basis for comparing the kinematic and kinetic profiles of the lower extremity during walking. *Journal of Biomechanics*, 2017.

[EKL+15] Ali Ezzati, Mindy J Katz, Michael L Lipton, Richard B Lipton, and Joe Verghese. The association of brain structure with gait velocity in older adults: a quantitative volumetric analysis of brain mri. *Neuroradiology*, 57:851–861, 2015.

[EMFM21] James Coleman Eubanks, Alec G Moore, Paul A Fishwick, and Ryan P McMahan. A preliminary embodiment short questionnaire. *Frontiers in Virtual Reality*, 2:647896, 2021.

[FFG+01] Nicolas Franck, Chlöé Farrer, Nicolas Georgieff, Michel Marie-Cardine, Jean Daléry, Thierry d'Amato, and Marc Jeannerod. Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, 158(3):454–459, 2001.

[FHKS12] Andrew Feng, Yazhou Huang, Marcelo Kallmann, and Ari Shapiro. An analysis of motion blending techniques. In *Motion in Games: 5th International Conference, MIG 2012, Rennes, France, November 15-17, 2012. Proceedings 5*, pages 232–243. Springer, 2012.

[FJ22] William M Felton and Russell E Jackson. Presence: A review. *International Journal of Human–Computer Interaction*, 38(1):1–18, 2022.

[FJD22] Xiaojun Fan, Xinyu Jiang, and Nianqi Deng. Immersive technology: A meta-analysis of augmented/virtual reality applications and their impact on tourism experience. *Tourism Management*, 91:104534, 2022.

[FSV09] Nasim Foroughi, Richard Smith, and Benedicte Vanwanseele. The association of external knee adduction moment with biomechanical variables in osteoarthritis: a systematic review. *The Knee*, 16(5):303–309, 2009.

[FVdPT01] Petros Faloutsos, Michiel Van de Panne, and Demetri Terzopoulos. Composable controllers for physics-based character animation. In

Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 251–260, 2001.

[Gal00] Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21, 2000.

[GB21] Nima Ghorbani and Michael J. Black. SOMA: Solving optical marker-based mocap automatically. In *Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021.

[GBT04] Pascal Glardon, Ronan Boulic, and Daniel Thalmann. Pca-based walking engine using motion capture data. In *Proceedings Computer Graphics International, 2004.*, pages 292–298. IEEE, 2004.

[GFP18] Mar Gonzalez-Franco and Tabitha C Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018.

[Gle98] Michael Gleicher. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 33–42, 1998.

[GMB+22] Guilherme Gonçalves, Miguel Melo, Luís Barbosa, José Vasconcelos-Raposo, and Maximino Bessa. Evaluation of the impact of different levels of self-representation and body tracking on the sense of presence and embodiment in immersive vr. *Virtual Reality*, 26(1):1–14, 2022.

[GME96] Patricia A Goldie, Thomas A Matyas, and Owen M Evans. Deficit and change in gait velocity during rehabilitation after stroke. *Archives of physical medicine and rehabilitation*, 77(10):1074–1082, 1996.

[Gol02] H. Goldstein. *Classical Mechanics*. Pearson Education, 2002.

[GP12] Thomas Geijtenbeek and Nicolas Pronost. Interactive character animation using simulated physics: A state-of-the-art review. In *Computer graphics forum*, volume 31, pages 2492–2515. Wiley Online Library, 2012.

[GVDPVDS13] Thomas Geijtenbeek, Michiel Van De Panne, and A Frank Van Der Stappen. Flexible muscle-based locomotion for bipedal creatures. *ACM Transactions on Graphics (TOG)*, 32(6):1–11, 2013.

[GY03] Weidong Geng and Gino Yu. Reuse of motion capture data in animation: A review. In *Computational Science and Its Applications—ICCSA 2003: International Conference Montreal, Canada, May 18–21, 2003 Proceedings, Part III 3*, pages 620–629. Springer, 2003.

[GZA09] Uri Givon, Gabriel Zeilig, and Anat Achiron. Gait analysis in multiple sclerosis: characterization of temporal–spatial parameters using gaitrite functional ambulation system. *Gait & posture*, 29(1):138–142, 2009.

[GZM20] Georgios Giarmatzis, Evangelia I Zacharaki, and Konstantinos Moustakas. Real-time prediction of joint forces by motion capture and machine learning. *Sensors*, 20(23):6933, 2020.

[HBL+17] Yinghao Huang, Federica Bogo, Christoph Lassner, Angjoo Kanazawa, Peter V Gehler, Javier Romero, Ijaz Akhter, and Michael J Black. Towards accurate marker-less human shape and pose estimation over time. In *2017 international conference on 3D vision (3DV)*, pages 421–430. IEEE, 2017.

[HK10] Yazhou Huang and Marcelo Kallmann. Motion parameterization with inverse blending. In *Motion in Games: Third International Conference, MIG 2010, Utrecht, The Netherlands, November 14-16, 2010. Proceedings 3*, pages 242–253. Springer, 2010.

[HKM+94] Larry F Hodges, Rob Kooper, Thomas C Meyer, Johannes J Hans De Graaff, Barbara O Rothbaum, Dan Opdyke, James S Williford, and Max M North. Presence as the defining factor in a vr application. 1994.

[HKPP20] Daniel Holden, Oussama Kanoun, Maksym Perepichka, and Tiberiu Popa. Learned motion matching. *ACM Transactions on Graphics (TOG)*, 39(4):53–1, 2020.

[HKS17] Daniel Holden, Taku Komura, and Jun Saito. Phase-functioned neural networks for character control. *ACM Transactions on Graphics (TOG)*, 36(4):1–13, 2017.

[HKT10] Edmond SL Ho, Taku Komura, and Chiew-Lan Tai. Spatial relationship preserving character motion adaptation. In *ACM SIGGRAPH 2010 papers*, pages 1–8. 2010.

[HMF18] Kathryn L Havens, Tatri Mukherjee, and James M Finley. Analysis of biases in dynamic margins of stability introduced by the use of simplified center of mass estimates during walking and turning. *Gait & posture*, 59:162–167, 2018.

[HOKP16] Ludovic Hoyet, Anne-Hélène Olivier, Richard Kulpa, and Julien Pettré. Perceptual effect of shoulder motions on crowd animations. *ACM Transactions on Graphics (TOG)*, 35(4):1–10, 2016.

[HP97] Jessica K Hodgins and Nancy S Pollard. Adapting simulated behaviors for new characters. In *Proceedings of the 24th annual conference*

on Computer graphics and interactive techniques, pages 153–162, 1997.

[HRE+08] Chris Hecker, Bernd Raabe, Ryan W Enslow, John DeWeese, Jordan Maynard, and Kees van Prooijen. Real-time motion retargeting to highly varied user-created morphologies. *ACM Transactions on Graphics (TOG)*, 27(3):1–11, 2008.

[HRZ+13] Ludovic Hoyet, Kenneth Ryall, Katja Zibrek, Hwangpil Park, Jehee Lee, Jessica Hodgins, and Carol O'sullivan. Evaluating the distinctiveness and attractiveness of human motions on realistic virtual bodies. *ACM Transactions on Graphics (TOG)*, 32(6):1–11, 2013.

[HS07] William EH Harcourt-Smith. The origins of bipedal locomotion. *Handbook of paleoanthropology*, 3:1483–1518, 2007.

[HSK16] Daniel Holden, Jun Saito, and Taku Komura. A deep learning framework for character motion synthesis and editing. *ACM Transactions on Graphics (TOG)*, 35(4):1–11, 2016.

[HWS+15] Tilo Hartmann, Werner Wirth, Holger Schramm, Christoph Klimmt, Peter Vorderer, André Gysbers, Saskia Böcking, Niklas Ravaja, Jari Laarni, Timo Saari, et al. The spatial presence experience scale (spes). *Journal of Media Psychology*, 2015.

[IE+53] Verne T Inman, Howard D Eberhart, et al. The major determinants in normal and pathological gait. *Jbjs*, 35(3):543–558, 1953.

[Inm66] Verne T Inman. Human locomotion. *Canadian Medical Association Journal*, 94(20):1047, 1966.

[JAA+16] Eakta Jain, Lisa Anthony, Aishat Aloba, Amanda Castonguay, Isabella Cuba, Alex Shaw, and Julia Woodward. Is the motion of a child perceivably different from the motion of an adult? *ACM Transactions on Applied Perception (TAP)*, 13(4):1–17, 2016.

[Joh73] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, 14:201–211, 1973.

[JT03] Daniel Jokisch and Nikolaus F Troje. Biological motion as a cue for the perception of size. *Journal of Vision*, 3(4):1–1, 2003.

[JT07] Kerri L Johnson and Louis G Tassinary. Compatibility of basic social perceptions determines perceived attractiveness. *Proceedings of the National Academy of Sciences*, 104(12):5246–5251, 2007.

[KAB20] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. Vibe: Video inference for human body pose and shape estimation. In

Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5253–5263, 2020.

[KAK+22] Ariel Kwiatkowski, Eduardo Alvarado, Vicky Kalogeiton, C. Karen Liu, Julien Pettré, Michiel van de Panne, and Marie-Paule Cani. A Survey on Reinforcement Learning Methods in Character Animation. *arXiv e-prints*, page arXiv:2203.04735, March 2022.

[Kal08] Marcelo Kallmann. Analytical inverse kinematics with body posture control. *Computer animation and virtual worlds*, 19(2):79–91, 2008.

[KBH+18] Kangsoo Kim, Luke Boelling, Steffen Haesler, Jeremy Bailenson, Gerd Bruder, and Greg F Welch. Does a digital assistant need a body? the influence of visual embodiment and social behavior on the perception of intelligent virtual agents in ar. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 105–114. IEEE, 2018.

[KdQMS+04] Inès A Kramers-de Quervain, Roland Müller, A Stacoff, Dieter Grob, and Edgar Stüssi. Gait analysis in patients with idiopathic scoliosis. *European Spine Journal*, 13:449–456, 2004.

[KGP02] Lucas Kovar, Michael Gleicher, and Frédéric Pighin. Motion graphs. 21(3):473–482, 2002.

[KGS12] Konstantina Kilteni, Raphaela Groten, and Mel Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[KHT16] Malte Klüver, Heiko Hecht, and Nikolaus F Troje. Internal consistency predicts attractiveness in biological motion walkers. *Evolution and Human Behavior*, 37(1):40–46, 2016.

[KKC+19] Jun-Ho Kim, Hyun-Jung Kim, Haluk Celik, Joo-Hwan Kim, and Dae-Hee Lee. Change in adduction moment following medial open wedge high tibial osteotomy: a meta-analysis. *BMC Musculoskeletal Disorders*, 20:1–7, 2019.

[KMA05] Richard Kulpa, Franck Multon, and Bruno Arnaldi. Morphology-independent representation of motions for interactive human-like animation. In *Eurographics*, 2005.

[KMH+19] Sophie Kenny, Naureen Mahmood, Claire Honda, Michael J Black, and Nikolaus F Troje. Perceptual effects of inconsistency in human animations. *ACM Transactions on Applied Perception (TAP)*, 16(1):1–18, 2019.

[KOOP11] Richard Kulpa, Anne-Hélène Olivierxs, Jan Ondřej, and Julien Pettré. Imperceptible relaxation of collision avoidance constraints in

virtual crowds. In *Proceedings of the 2011 SIGGRAPH Asia Conference*, pages 1–10, 2011.

[KS14] Elena Kokkinara and Mel Slater. Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception*, 43(1):43–58, 2014.

[LBS13] Joan Llobera, Kristopher J Blom, and Mel Slater. Telling stories within immersive virtual environments. *Leonardo*, 46(5):471–476, 2013.

[LC12] Tung-Wu Lu and Chu-Fen Chang. Biomechanics of human movement and its clinical applications. *The Kaohsiung journal of medical sciences*, 28(2):S13–S25, 2012.

[LCB06] Benoît Le Callennec and Ronan Boulic. Interactive motion deformation with prioritized constraints. *Graphical Models*, 68(2):175–193, 2006.

[LCC19] George HZ Liu, Michael ZQ Chen, and Yonghua Chen. When joggers meet robots: the past, present, and future of research on humanoid robots. *Bio-Design and Manufacturing*, 2(2):108–118, 2019.

[LCR+02] Jehee Lee, Jinxiang Chai, Paul SA Reitsma, Jessica K Hodgins, and Nancy S Pollard. Interactive control of avatars animated with human motion data. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 491–500, 2002.

[LDMH13] Sébastien Lobet, Christine Detrembleur, Firas Massaad, and Cedric Hermans. Three-dimensional gait analysis can shed new light on walking in patients with haemophilia. *The Scientific World Journal*, 2013(1):284358, 2013.

[LF98] Cynthia R Lee and Claire T Farley. Determinants of the center of mass trajectory in human walking and running. *Journal of experimental biology*, 201(21):2935–2944, 1998.

[LH17] Libin Liu and Jessica Hodgins. Learning to schedule control fragments for physics-based characters using deep q-learning. *ACM Transactions on Graphics (TOG)*, 36(3):1–14, 2017.

[LH18] Libin Liu and Jessica Hodgins. Learning basketball dribbling skills using trajectory optimization and deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, 37(4):1–14, 2018.

[LHP05] C Karen Liu, Aaron Hertzmann, and Zoran Popović. Learning physics-based motion style with nonlinear inverse optimization. *ACM Transactions on Graphics (TOG)*, 24(3):1071–1081, 2005.

[LJ15] Matthew Lombard and Matthew T Jones. Defining presence. *Immersed in media: Telepresence theory, measurement & technology*, pages 13–34, 2015.

[LJ16] Lorraine Lin and Sophie Jörg. Need a hand? how appearance affects the virtual hand illusion. In *Proceedings of the ACM symposium on applied perception*, pages 69–76, 2016.

[LMB14a] Matthew Loper, Naureen Mahmood, and Michael J Black. Mosh: motion and shape capture from sparse markers. *ACM Trans. Graph.*, 33(6):220–1, 2014.

[LMB14b] Matthew M. Loper, Naureen Mahmood, and Michael J. Black. MoSh: Motion and shape capture from sparse markers. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 33(6):220:1–220:13, November 2014.

[LMR+15a] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015.

[LMR+15b] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.

[LMT01] Won-Sook Lee and Nadia Magnenat-Thalmann. Virtual body morphing. In *Proceedings Computer Animation 2001. Fourteenth Conference on Computer Animation (Cat. No. 01TH8596)*, pages 158–166. IEEE, 2001.

[LMT08] Etienne Lyard and Nadia Magnenat-Thalmann. Motion adaptation based on character shape. *Computer Animation and Virtual Worlds*, 19(3-4):189–198, 2008.

[LNA+19] Lorraine Lin, Aline Normoyle, Alexandra Adkins, Yu Sun, Andrew Robb, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. The effect of hand size and interaction modality on the virtual hand illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 510–518. IEEE, 2019.

[LP12] Sergey Levine and Jovan Popović. Physically plausible simulation for character animation. In *Proceedings of the 11th ACM SIGGRAPH/Eurographics conference on Computer Animation*, pages 221–230, 2012.

[LRG+17] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the*

23rd ACM symposium on virtual reality software and technology, pages 1–10, 2017.

[LSJC14] Chiung-ju Liu, Deepika M Shiroy, Leah Y Jones, and Daniel O Clark. Systematic review of functional training on muscle strength, physical functioning, and activities of daily living in older adults. *European review of aging and physical activity*, 11:95–106, 2014.

[LYVdP+10] Libin Liu, KangKang Yin, Michiel Van de Panne, Tianjia Shao, and Weiwei Xu. Sampling-based contact-rich motion control. In *ACM SIGGRAPH 2010 papers*, pages 1–10. 2010.

[MAS94] AE Minetti, Luca Paolo Ardigo, and Saibene. The transition between walking and running in humans: metabolic and mechanical aspects at different gradients. *Acta physiologica scandinavica*, 150(3):315–323, 1994.

[MC90] Thomas A McMahon and George C Cheng. The mechanics of running: how does stiffness couple with speed? *Journal of biomechanics*, 23:65–78, 1990.

[MCC+80] John T McConville, Charles E Clauser, Thomas D Churchill, Jaime Cuzzi, and Ints Kaleps. Anthropometric relationships of body and body segment moments of inertia. Technical report, ANTHROPOLOGY RESEARCH PROJECT INC YELLOW SPRINGS OH, 1980.

[MCM16] Kathleen FE MacLean, Jack P Callaghan, and Monica R Maly. Effect of obesity on knee joint biomechanics during gait in young adults. *Cogent Medicine*, 3(1):1173778, 2016.

[MCRTB10] Betty J Mohler, Sarah H Creem-Regehr, William B Thompson, and Heinrich H Bülthoff. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence*, 19(3):230–242, 2010.

[MDLH10] Igor Mordatch, Martin De Lasa, and Aaron Hertzmann. Robust physics-based locomotion using low-dimensional planning. In *ACM SIGGRAPH 2010 papers*, pages 1–8. 2010.

[Men00] Alberto Menache. *Understanding motion capture for computer animation and video games*. Morgan kaufmann, 2000.

[MFC20] Andy Miah, Alex Fenton, and Simon Chadwick. Virtual reality and sports: The rise of mixed, augmented, immersive, and esports experiences. *21st century sports: How technologies will change sports in the digital age*, pages 249–262, 2020.

[MGT+19a] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pages 5442–5451, October 2019.

[MGT+19b] Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5442–5451, 2019.

[MHG+18] Josh Merel, Leonard Hasenclever, Alexandre Galashov, Arun Ahuja, Vu Pham, Greg Wayne, Yee Whye Teh, and Nicolas Heess. Neural probabilistic motor primitives for humanoid control. *arXiv preprint arXiv:1811.11711*, 2018.

[MHLC+22] Lucas Mourot, Ludovic Hoyet, François Le Clerc, François Schnitzler, and Pierre Hellier. A survey on deep learning for skeleton-based human animation. In *Computer Graphics Forum*, volume 41, pages 122–157. Wiley Online Library, 2022.

[MJH+07] Rachel McDonnell, Sophie Jörg, Jessica K Hodgins, Fiona Newell, and Carol O'Sullivan. Virtual shapers & movers: form and motion affect sex perception. In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, pages 7–10, 2007.

[MKHK09] Franck Multon, Richard Kulpa, Ludovic Hoyet, and Taku Komura. Interactive animation of virtual humans based on motion capture data. *Computer Animation and Virtual Worlds*, 20(5-6):491–500, 2009.

[MLD+08] Rachel McDonnell, Michéal Larkin, Simon Dobbyn, Steven Collins, and Carol O'Sullivan. Clone attack! perception of crowd variety. In *ACM SIGGRAPH 2008 papers*, pages 1–8. 2008.

[MLH+09] Rachel McDonnell, Michéal Larkin, Benjamín Hernández, Isaac Rudomin, and Carol O'Sullivan. Eye-catching crowds: saliency based selective variation. *ACM Transactions on Graphics (TOG)*, 28(3):1–10, 2009.

[MMK12] Masahiro Mori, Karl F MacDorman, and Norri Kageki. The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2):98–100, 2012.

[MNMO08] J Martínez, JC Nebel, D Makris, and C Orrite. Tracking human body parts using particle filters constrained by human biomechanics. In *Proc. BMVC*, 2008.

[MS13] Antonella Maselli and Mel Slater. The building blocks of the full body ownership illusion. *Frontiers in human neuroscience*, 7:83, 2013.

[MSK22] Ian Mason, Sebastian Starke, and Taku Komura. Real-time style modelling of human locomotion via feature-wise transformations and local motion phases. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 5(1):1–18, 2022.

[MST19] Ayelet Melzer, Tal Shafir, and Rachelle Palnick Tsachor. How do we recognize emotion from movement? specific motor components contribute to the recognition of each emotion. *Frontiers in psychology*, 10:392097, 2019.

[MTD+15] Christian Mandery, Ömer Terlemez, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. The kit whole-body human motion database. In *2015 International Conference on Advanced Robotics (ICAR)*, pages 329–336. IEEE, 2015.

[MTM+18] Simone C Mölbert, Anne Thaler, Betty J Mohler, Stephan Streuber, Javier Romero, Michael J Black, Stephan Zipfel, H-O Karnath, and Katrin E Giel. Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: Attitudinal components rather than visual body size estimation are distorted. *Psychological medicine*, 48(4):642–653, 2018.

[MTT05] Nadia Magnenat-Thalmann and Daniel Thalmann. *Handbook of virtual humans*. John Wiley & Sons, 2005.

[MWG+21] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.

[MYPT09] Jonathan Maim, Barbara Yersin, Julien Pettre, and Daniel Thalmann. Yaq: an architecture for real-time navigation and rendering of varied crowds. *IEEE Computer Graphics and Applications*, 29(4):44–53, 2009.

[MYT+21] Li-Ke Ma, Zeshi Yang, Xin Tong, Baining Guo, and KangKang Yin. Learning and exploring motor skills with spacetime bounds, 2021.

[NODLL20] Estelle Nakul, Nicolas Orlando-Dessaints, Bigna Lenggenhager, and Christophe Lopez. Measuring perceived self-location in virtual reality. *Scientific Reports*, 10(1):6802, 2020.

[NOZ+20] Benjamin Niay, Anne-Hélène Olivier, Katja Zibrek, Julien Pettré, and Ludovic Hoyet. Walk Ratio: Perception of an Invariant Parameter of Human Walk on Virtual Characters. In *SAP 2020 - ACM Symposium on Applied Perception*, pages 1–9, Virtual Event, United States, September 2020. ACM.

[O+95] World Health Organization et al. *Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee*. World Health Organization, 1995.

[OBPB22] Magalie Ochs, Jérémie Bousquet, Jean-Marie Pergandi, and Philippe Blache. Multimodal behavioral cues analysis of the sense of presence and social presence during a social interaction with a virtual patient. *Frontiers in Computer Science*, 4:746804, 2022.

[ONA+21] Akimi Oyanagi, Takuji Narumi, Kazuma Aoyama, Kenichiro Ito, Tomohiro Amemiya, and Michitaka Hirose. Impact of long-term use of an avatar to ivbo in the social vr. In *International Conference on Human-Computer Interaction*, pages 322–336. Springer, 2021.

[Org95] World Health Organization. Physical status: The use and interpretation of anthropometry. *WHO Technical Report Series*, 854, 1995.

[OSK+04] Michael S Orendurff, Ava D Segal, Glenn K Klute, Jocelyn S Berge, Eric S Rohr, and Nancy J Kadel. The effect of walking speed on center of mass displacement. *Journal of Rehabilitation Research & Development*, 41(6), 2004.

[PALvdP18] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics (TOG)*, 37(4):1–14, 2018.

[Pat97] Aftab E Patla. Understanding the roles of vision in the control of human locomotion. *Gait & posture*, 5(1):54–69, 1997.

[PBYVDP17] Xue Bin Peng, Glen Berseth, KangKang Yin, and Michiel Van De Panne. Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning. *Acm transactions on graphics (tog)*, 36(4):1–13, 2017.

[PCG+19] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10975–10985, 2019.

[PKL+22] Andreas Panayiotou, Theodoros Kyriakou, Marilena Lemonari, Yiorgos Chrysanthou, and Panayiotis Charalambous. Ccp: Configurable crowd profiles. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022.

[PKM+18] Xue Bin Peng, Angjoo Kanazawa, Jitendra Malik, Pieter Abbeel, and Sergey Levine. Sfv: Reinforcement learning of physical skills from videos. *ACM Transactions On Graphics (TOG)*, 37(6):1–14, 2018.

[PMA+21] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (TOG)*, 40(4):1–20, 2021.

[PO11] Martin Pražák and Carol O'Sullivan. Perceiving human motion variety. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, pages 87–92, 2011.

[PSS02] Sang Il Park, Hyun Joon Shin, and Sung Yong Shin. On-line locomotion generation based on motion blending. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 105–111, 2002.

[PZZD18] Georgios Pavlakos, Luyang Zhu, Xiaowei Zhou, and Kostas Daniilidis. Learning to estimate 3d human pose and shape from a single color image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 459–468, 2018.

[RCSSJ+08] Abel Romero-Corral, Virend K Somers, Justo Sierra-Johnson, Randal J Thomas, ML Collazo-Clavell, JEC Korinek, Thomas G Allison, JA Batsis, FH Sert-Kuniyoshi, and Francisco Lopez-Jimenez. Accuracy of body mass index in diagnosing obesity in the adult general population. *International journal of obesity*, 32(6):959–966, 2008.

[RDP99] Kathleen M Robinette, Hans Daanen, and Eric Paquet. The caesar project: a 3-d surface anthropometry survey. In *Second international conference on 3-D digital imaging and modeling (cat. No. PR00062)*, pages 380–386. IEEE, 1999.

[RF83] Sverker Runeson and Gunilla Frykholm. Kinematic specification of dynamics as an informational basis for person-and-action perception: expectation, gender recognition, and deceptive intention. *Journal of experimental psychology: general*, 112(4):585, 1983.

[RIKA08] Brian Ries, Victoria Interrante, Michael Kaeding, and Lee Anderson. The effect of self-embodiment on distance perception in immer-

sive virtual environments. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 167–170, 2008.

[RL20] Daniel Roth and Marc Erich Latoschik. Construction of the virtual embodiment questionnaire (veq). *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3556, 2020.

[RMFW20] Jaziar Radianti, Tim A Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147:103778, 2020.

[RMP17] Mary Roberts, David Mongeon, and Francois Prince. Biomechanical parameters for gait analysis: a systematic review of healthy human gait. *Phys. Ther. Rehabil*, 4(6), 2017.

[RNC18] Tyler Rose, Chang S Nam, and Karen B Chen. Immersion of virtual reality for rehabilitation-review. *Applied ergonomics*, 69:153–161, 2018.

[Row19] Rebecca Rowe. Shaping girls: analyzing animated female body shapes. *Animation*, 14(1):22–36, 2019.

[RSM+22] Joseph Russell, Sunia Saboor, Manmeet Makkar, Arpita Barua, and Anne Thaler. Detection of inconsistency between shape and motion in realistic female and male animation. *STEM Fellowship Journal*, (0):1–8, 2022.

[RSM+23] Joseph Russell, Sunia Saboor, Manmeet Makkar, Arpita Barua, and Anne Thaler. Detection of inconsistency between shape and motion in realistic female and male animation. *STEM Fellowship Journal*, 8(1):17–23, 2023.

[RTB17] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: modeling and capturing hands and bodies together. *ACM Transactions on Graphics*, 36(6):1–17, November 2017.

[RTB22] Javier Romero, Dimitrios Tzionas, and Michael J Black. Embodied hands: Modeling and capturing hands and bodies together. *arXiv preprint arXiv:2201.02610*, 2022.

[S+92] Thomas B Sheridan et al. Musings on telepresence and virtual presence. *Presence Teleoperators Virtual Environ.*, 1(1):120–125, 1992.

[SBB07] Leonid Sigal, Alexandru Balan, and Michael Black. Combined discriminative and generative articulated pose and non-rigid shape estimation. *Advances in neural information processing systems*, 20, 2007.

[Sch03] Daniel Schmitt. Insights into the evolution of human bipedalism from experimental studies of humans and other primates. *Journal of Experimental Biology*, 206(9):1437–1448, 2003.

[Sch09] Thomas W Schubert. A new conception of spatial presence: Once again, with feeling. *Communication Theory*, 19(2):161–187, 2009.

[SdSdSH19] Gabriel Caixeta Silva, André Tavares da Silva, and Marcelo da Silva Hounsell. Crowd generation using morphological obesity criteria. In *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 81–90. IEEE, 2019.

[SFR01] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001.

[SGR+03] Federica Sibella, Manuela Galli, Marianna Romei, A Montesano, and Marcello Crivellini. Biomechanical analysis of sit-to-stand movement in normal and obese subjects. *Clinical biomechanics*, 18(8):745–750, 2003.

[SIE53] J.B. Saunders, V.T. Inman, and H.D. Eberhart. The major determinants in normal and pathological gait. *Jbjs*, 35(3):543–558, 1953.

[SKG03] Hyun Joon Shin, Lucas Kovar, and Michael Gleicher. Physical touch-up of human motions. In *Pacific Graphics*, volume 3, page 194, 2003.

[SKGD20] Stella Sylaiou, Vlasios Kasapakis, Damianos Gavalas, and Elena Dzardanova. Avatars as storytellers: affective narratives in virtual museums. *Personal and Ubiquitous Computing*, 24:829–841, 2020.

[SKHH19] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–12, 2019.

[SKNG23] Kim Szolin, Daria J Kuss, Filip M Nuyens, and Mark D Griffiths. "i am the character, the character is me": A thematic analysis of the user-avatar relationship in videogames. *Computers in Human Behavior*, 143:107694, 2023.

[SKR+20] Bernd J Stetter, Frieder C Krafft, Steffen Ringhof, Thorsten Stein, and Stefan Sell. A machine learning and wearable sensor based approach to estimate external knee flexion and adduction moments during various locomotion tasks. *Frontiers in bioengineering and biotechnology*, 8:9, 2020.

[SKTDR98] Meera Saini, D Casey Kerrigan, Mandyam A Thirunarayan, and M Duff-Raffaele. The vertical displacement of the center of mass during walking: a comparison of four measurement methods. 1998.

[Sla03] Mel Slater. A note on presence terminology. *Presence connect*, 3(3):1–5, 2003.

[Sla09] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[SM03] Franco Saibene and Alberto E Minetti. Biomechanical and physiological aspects of legged locomotion in humans. *European journal of applied physiology*, 88:297–316, 2003.

[SMC+00] Monique M Samson, IB Meeuwsen, Alan Crowe, JA Dessens, Sijmen A Duursma, and HJ Verhaar. Relationships between physical performance measures, age, height and body weight in healthy adults. *Age and ageing*, 29(3):235–242, 2000.

[SMK22] Sebastian Starke, Ian Mason, and Taku Komura. Deepphase: Periodic autoencoders for learning motion phase manifolds. *ACM Transactions on Graphics (TOG)*, 41(4):1–13, 2022.

[SOWO17] Yinxuan Shi, Jan Ondřej, He Wang, and Carol O'Sullivan. Shape up! perception based body shape variation for data-driven crowds. In *2017 ieee virtual humans and crowds for immersive environments (vhcie)*, pages 1–7. IEEE, 2017.

[SSC10] Mel Slater, Bernhard Spanlang, and David Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM transactions on graphics (TOG)*, 29(4):1–9, 2010.

[SSMM98] Mel Slater, Anthony Steed, John McCarthy, and Francesco Maringelli. The influence of body movement on subjective presence in virtual environments. *Human factors*, 40(3):469–477, 1998.

[SSV16] Mel Slater and Maria V Sanchez-Vives. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3:74, 2016.

[SUC95] Mel Slater, Martin Usoh, and Yiorgos Chrysanthou. The influence of dynamic shadows on presence in immersive virtual environments. In *Virtual Environments' 95: Selected papers of the Eurographics Workshops in Barcelona, Spain, 1993, and Monte Carlo, Monaco, 1995*, pages 8–21. Springer, 1995.

[SVCN+21] Danilo B da Silva, Creto A Vidal, Joaquim B Cavalcante-Neto, Italo NS Pessoa, and Rubens F Nunes. Topple-free foot strategy applied to real-time motion capture data using kinect sensor. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, pages 98–106, 2021.

[SVKS19] Shubham Sharma, Shubhankar Verma, Mohit Kumar, and Lavanya Sharma. Use of motion capture in 3d animation: motion capture systems, challenges, and recent trends. In *2019 international conference on machine learning, big data, cloud and parallel computing (comitcon)*, pages 289–294. IEEE, 2019.

[SWD+17] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.

[SWH18] Valentin Schwind, Katrin Wolf, and Niels Henze. Avoiding the uncanny valley in virtual character design. *interactions*, 25(5):45–49, 2018.

[SY05] Atsushi Sato and Asako Yasuda. Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, 94(3):241–255, 2005.

[SZKZ20] Sebastian Starke, Yiwei Zhao, Taku Komura, and Kazi Zaman. Local motion phases for learning multi-contact character movements. *ACM Transactions on Graphics (TOG)*, 39(4):54–1, 2020.

[TBM+20] Anne Thaler, Andreas Bieg, Naureen Mahmood, Michael J Black, Betty J Mohler, and Nikolaus F Troje. Attractiveness and confidence in walking style of male and female virtual characters. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 678–679. IEEE, 2020.

[TJ95] Frank Thomas and Ollie Johnston. *The illusion of life: Disney animation*. Disney Editions New York, 1995.

[TLC02] Franco Tecchia, Céline Loscos, and Yiorgos Chrysanthou. Image-based crowd rendering. *IEEE computer graphics and applications*, 22(2):36–43, 2002.

[TLT11] Jie Tan, Karen Liu, and Greg Turk. Stable proportional-derivative controllers. *IEEE Computer Graphics and Applications*, 31(4):34–44, 2011.

[Tro02] Nikolaus F Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of vision*, 2(5):2–2, 2002.

[Tro03] Nikolaus F Troje. Cat walk and western hero–motion is expressive. *IGSN report*, pages 40–43, 2003.

[Tro08] Nikolaus F Troje. Retrieving information from human movement patterns. *Understanding events: How humans see, represent, and act on events*, 1:308–334, 2008.

[Tsa10] Manos Tsakiris. My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia*, 48(3):703–712, 2010.

[UCAS00] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. Using presence questionnaires in reality. *Presence*, 9(5):497–503, 2000.

[UCT04] Branislav Ulicny, Pablo de Heras Ciechomski, and Daniel Thalmann. Crowdbrush: interactive authoring of real-time crowd scenes. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 243–252, 2004.

[Vau92] C.L. Vaughan. *Dynamics of Human Gait*. Human Kinetics Books, 1992.

[VGRL10] Jos Vanrenterghem, Donna Gormley, Mark Robinson, and Adrian Lees. Solutions for representing the whole-body centre of mass in side cutting manoeuvres based on data that is typically available for lower limb kinematics. *Gait & posture*, 31(4):517–521, 2010.

[VPCN17] Maryam Vaziri-Pashkam, Sarah Cormiea, and Ken Nakayama. Predicting actions from subtle preparatory movements. *Cognition*, 168:65–75, 2017.

[VYCL18] Ruben Villegas, Jimei Yang, Duygu Ceylan, and Honglak Lee. Neural kinematic networks for unsupervised motion retargetting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8639–8648, 2018.

[VZ83] V.N. Seluyanov V.M. Zatsiorsky. The mass and inertia characteristics of the main segments of the human body. *Biomechanics*, 1983.

[WB03] Jing Wang and Bobby Bodenheimer. An evaluation of a cost metric for selecting transitions between motion segments. In *Symposium on Computer Animation*, volume 2. Citeseer, 2003.

[WGH20] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. A scalable approach to control diverse behaviors for physically simulated characters. *ACM Transactions on Graphics (TOG)*, 39(4):33–1, 2020.

[WGH21] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. Control strategies for physically simulated characters performing two-player competitive sports. *ACM Transactions on Graphics (TOG)*, 40(4):1–11, 2021.

[WGH22] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. Physics-based character controllers using conditional vaes. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022.

[WGP+21] Markus Wirth, Stefan Gradl, Georg Prosinger, Felix Kluge, Daniel Roth, and Bjoern M Eskofier. The impact of avatar appearance, perspective and context on gait variability and user experience in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 326–335. IEEE, 2021.

[WHB+07] Werner Wirth, Tilo Hartmann, Saskia Böcking, Peter Vorderer, Christoph Klimmt, Holger Schramm, Timo Saari, Jari Laarni, Niklas Ravaja, Feliz Ribeiro Gouveia, et al. A process model of the formation of spatial presence experiences. *Media psychology*, 9(3):493–525, 2007.

[Whi14] Michael W Whittle. *Gait analysis: an introduction*. Butterworth-Heinemann, 2014.

[Win95] David A Winter. Human balance and posture control during standing and walking. *Gait & posture*, 3(4):193–214, 1995.

[WK88] Andrew Witkin and Michael Kass. Spacetime constraints. *ACM Siggraph Computer Graphics*, 22(4):159–168, 1988.

[WL19] Jungdam Won and Jehee Lee. Learning body shape variation in physics-based characters. *ACM Transactions on Graphics (TOG)*, 38(6):1–12, 2019.

[WS98] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[WS12] Cara M Wall-Scheffler. Size and shape: Morphology's impact on human speed and mobility. *Journal of Anthropology*, 2012(1):340493, 2012.

[YCS+83] Joseph W Young, Richard F Chandler, Clyde C Snow, Kathleen M Robinette, Gregory F Zehner, Maureen S Loftberg, et al. Anthropometric and mass distribution characteristics of the adult female. Technical report, Civil Aerospace Medical Institute, 1983.

[YHC+22] Tairan Yin, Ludovic Hoyet, Marc Christie, Marie-Paule Cani, and Julien Pettré. The one-man-crowd: Single user generation of crowd motions using virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2245–2255, 2022.

[YLVdP07] KangKang Yin, Kevin Loken, and Michiel Van de Panne. Simbicon: Simple biped locomotion control. *ACM Transactions on Graphics (TOG)*, 26(3):105–es, 2007.

[YON24] Goksu Yamac, Carol O'Sullivan, and Michael Neff. Understanding the impact of visual and kinematic information on the perception of physicality errors. *ACM Transactions on Applied Perception*, 22(1):1–30, 2024.

[YTK+21] Andy Wai Kan Yeung, Anela Tosevska, Elisabeth Klager, Fabian Eibensteiner, Daniel Laxar, Jivko Stoyanov, Marija Glisic, Sebastian Zeiner, Stefan Tino Kulnik, Rik Crutzen, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. *Journal of medical internet research*, 23(2):e25499, 2021.

[YWS+21] Ye Yuan, Shih-En Wei, Tomas Simon, Kris Kitani, and Jason Saragih. Simpoe: Simulated character control for 3d human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7159–7169, 2021.

[YY] Goh Jing Ying and KangKang Yin. SFU motion capture database. `.http://mocap.cs.sfu.ca/`.

[ZH02] Victor Brian Zordan and Jessica K Hodgins. Motion capture-driven simulations that hit and react. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 89–96, 2002.

[ZHT+22] Jingjing Zhang, Mengjie Huang, Xiaohang Tang, Yiqi Wang, and Rui Yang. How virtual body continuity with different hand representations influence on user perceptions and task performance. In *2022 15th International Conference on Human System Interaction (HSI)*, pages 1–5. IEEE, 2022.

[ZMCF05] Victor Brian Zordan, Anna Majkowska, Bill Chiu, and Matthew Fast. Dynamic response for motion capture animation. *ACM Transactions on Graphics (TOG)*, 24(3):697–701, 2005.