Kinematics-Driven Motion Analysis for 3D Hand Trajectory Prediction in Virtual Reality

NISAL MENUKA KANWEL GAMAGE DON

Supervisor: A/Prof. Anusha Withana
Co-Supervisor: Prof. Judy Kay

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

This research reported in this thesis was supported by the award of a
Research Training Program scholarship to the PhD Candidate.

School of Computer Science
Faculty of Engineering
The University of Sydney
Australia

7 June 2025

Abstract

Hand motion plays a fundamental role in how people interact with their environment, making hand trajectory prediction and analysis vital across numerous domains, including Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI). Accurate hand motion modeling and prediction are particularly beneficial in Virtual Reality (VR) applications, where they can reduce system latency and enable the development of novel and immersive experiences. However, despite prior work on various statistical and deep learning approaches, challenges persist in developing accurate, efficient, and generalizable models for deployment in real-time applications on devices with limited resources.

This thesis addresses these challenges by introducing novel techniques for analyzing and predicting hand motion, integrating empirical data and kinematics-based approaches to enhance existing mathematical models and statistical methods. First, we conducted a user study with 20 participants involving structured and unstructured VR movements to address the limitations of existing datasets. Using the collected data, we developed user-specific predictive models tailored to individual motion patterns, achieving high accuracy in hand motion prediction. Building on these personalized models, we developed generalized models that maintain comparable performance levels. These generalized models are designed to adapt across a broader user base without requiring individual customization, balancing accuracy and generalizability. Finally, we tested these models on new users, activities, and datasets, demonstrating their robustness and adaptability.

This thesis advances the fields of HCI and HRI by deepening the understanding of human hand motion and presenting techniques for accurate and efficient motion prediction. It offers a scalable approach that balances customization with broader applicability by introducing user-specific and generalized models. The future of this research lies in expanding it to encompass the entire human body, with the potential to significantly impact areas such as VR and collaborative human-robot environments.

Authorship Attribution Statement

Chapters 3 and 4 of this thesis are based on publications and works in progress, of which I am the lead author and have made significant contributions to each. According to university guidelines, included papers that have been published or are under review are presented here in this thesis without changes.

My supervisor's (A/Prof. Anusha Withana) work was supported by the Australian Research Council Discovery Early Career Award by the Australian Government under Grant DE200100479. My research was supported in part by the Engineering and Information Technologies Research Scholarship (EITRS) and in part by the Research Training Program (RTP) scholarship. Furthermore, all studies presented in these chapters received ethics approval under project number 2019/553 from the Human Research Ethics Committee (HREC) of the University of Sydney.

Chapter 3:

Gamage, Nisal Menuka, Deepana Ishtaweera, Martin Weigel, and Anusha Withana. "So predictable! Continuous 3D Hand Trajectory Prediction in Virtual Reality." In *The 34th Annual ACM Symposium on User Interface Software and Technology, pp. 332-343. 2021.*

The paper has been published at the ACM Symposium on User Interface Software and Technology (UIST) conference. My contributions included conceptual development, designing and conducting the user study, developing the predictive models, and writing the paper.

Chapter 4:

Gamage, Nisal Menuka and Anusha Withana. "Empirically Modified Minimum Jerk Models for 3D Trajectory Prediction for Reaching Movements."

The paper is under review at Special Interest Group on Computer Graphics and Interactive Techniques, Asia (SIGGRAPH Asia) journal track. My contributions included conceptual development, developing the predictive models, evaluation of the models, and writing the paper.

Additionally, during my PhD, I contributed to the following paper:

Lin, Stephen Shiao-ru, **Nisal Menuka Gamage**, Kithmini Herath, and Anusha Withana. "MyoSpring: 3D printing mechanomyographic sensors for subtle finger gesture recognition." In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 1-13. 2022.*

Nisal Menuka Kanwel Gamage Don

A/Prof. Anusha Withana

7th June 2025

Acknowledgements

I would like to express my sincere gratitude to my supervisor, A/Prof. Anusha Withana, for his supervision, guidance, and research insights throughout my doctoral research. My gratitude extends to my co-supervisor, Prof. Judy Kay, for her constant support and encouragement. Their support has been instrumental in the completion of this thesis.

I extend my thanks to my fellow lab members, including Stephen Lin, Deepana Ishtaweera, Thejan Elankayer, and Dr. Philip Mehrgardt, for their collaboration and support. I am deeply grateful to Dr. Martin Weigel for his expertise and valuable feedback, and Dr. Jiakun Yu, for his enduring friendship and support throughout the years.

Additionally, I am thankful to my close friends, Namitha Liyanage and Dr. Tharindu Ranasinghe, for their encouragement and insightful discussions throughout this journey. Their support gave me the strength and the motivation needed to overcome many challenges.

Most importantly, I would like to express my heartfelt gratitude to my loving parents, Lalith Dharamarathne and Thejani Perera, for their unconditional love and unwavering support throughout my life. They have been the supporting pillars behind my success.

Finally, I would like to express my heartfelt thanks to my lovely wife, Dr. Namalka Liyanage, whose encouragement, understanding, and unwavering support have been monumental in this journey. Her constant presence and support made this journey possible.

Declaration

I, Nisal Menuka Kanwel Gamage Don, hereby declare that the thesis titled " Kinematics-Driven Motion Analysis for 3D Hand Trajectory Prediction in Virtual Reality", submitted to the University of Sydney, is in fulfillment of the requirements for the Degree of Doctor of Philosophy (Ph.D.). This work is original and entirely my own, except where I have duly cited or acknowledged other sources. I also confirm that this work has not been previously submitted, either in its entirety or in part, to this or any other academic institution for any degree, diploma, or other qualification.

Nisal Menuka Kanwel Gamage Don

7th June 2025

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Understanding human hand motion is essential across various domains, as the hand serves as a primary interface for human interaction with the environment. These interactions are diverse, ranging from simple gestures like reaching to grab an object, to fast and powerful actions like executing a punch in boxing, and to coordinated hand movements involved in tasks such as cooking or assembling objects. At a fundamental level, motion analysis seeks to uncover the underlying patterns of these diverse movements, while motion prediction focuses on anticipating future actions. Together, these approaches enable a more holistic understanding of human motion.

This understanding plays a crucial role in areas such as Human-Computer Interaction (HCI) [7, 180] and Human-Robot Interaction (HRI) [54, 19], where precise motion modeling enhances usability, responsiveness, and safety. In virtual reality (VR) environments, noticeable delays in visual, auditory, or haptic feedback can disrupt the user experience and break immersion. Therefore, precise hand motion prediction is essential for creating immersive, interactive experiences that respond seamlessly to user movements by mitigating system delays through predictive actions [134].

For example, consider a scenario in a VR environment where a user punches a water-filled balloon. The complex physical interactions involved, such as the motion of water when the balloon bursts, present a significant challenge for rendering. This is particularly difficult in resource-constrained VR headsets, where achieving real-time rendering of such scenarios is demanding. However, if the trajectory of the collision is predicted in advance, the system can initiate pre-rendering of the relevant graphics, enabling timely visual feedback to the user. Similarly, in human-robot collaborative environments, the ability to predict a worker's hand trajectory can prevent collisions and optimize cooperative tasks [190].

Due to these clear advantages, various techniques have been proposed for human motion analysis and prediction, ranging from statistical models [16, 67, 179] to deep learning approaches [109, 98, 93]. The study of human motion has historically been a multidisciplinary endeavor, involving fields such as biomechanics, robotics, neuroscience, and computer science. Traditional models like the Minimum Jerk Model (MJM) [42] have provided foundational frameworks for predicting smooth reaching trajectories by minimizing the rate of change of acceleration. These models are valued for their simplicity and computational efficiency, making them suitable for real-time applications. However, they often fall short of accurately capturing the complexity of human movements, exhibiting significant discrepancies when compared to empirical data.

With the advent of deep learning, more advanced models have emerged, offering greater precision in representing human motion dynamics. These advancements have enabled applications in rehabilitation systems [34, 95], virtual avatar motion planning [145, 132], and collaborative human-robot environments [100, 105]. Furthermore, deep learning approaches have been employed to predict hand motion in virtual reality (VR) [27]. Nonetheless, these models typically demand substantial computational resources, which may not be feasible for devices with limited capabilities, such as standalone VR headsets. Additionally, they often lack interpretability, making it difficult to understand the underlying decision-making processes or to diagnose potential errors.

Despite these advancements, challenges remain in developing models that are accurate, interpretable, and efficient for real-time applications. Consider a VR dancing application where users perform a wide range of motions, from slow and smooth movements, such as graceful ballet gestures, to sudden and rapid actions, like those seen in salsa dancing. In such scenarios, the system must track and predict these motions in real time. While deep learning methods might handle the varied movements with greater accuracy, they could also impose significant computational demands that may exceed the capabilities of mobile VR headsets with limited processing power. Additionally, a classical model may run efficiently on a standalone VR headset but struggle to handle the diverse range of movements, including accommodating multiple users.

Our goal is to address the major gaps in current hand motion prediction models that have limited their broader adoption in VR. These gaps include: (1) the need for the models to perform consistently, (2) the computational limitations of resource-constrained devices, (3) the lack of interpretability associated with deep learning methods. Motivated by these challenges, we aim to develop approaches that bridge the gap between the simplicity and interpretability of traditional models and the accuracy and generalizability of advanced deep learning methods.

This thesis focuses on creating novel methods for human motion prediction and analysis that balance accuracy, computational efficiency, generalizability, and interpretability. Specifically, we investigate how to combine empirical observations and theoretical understanding of kinematics with classical statistical methods to develop robust predictive frameworks. For instance, we explore the development of kinematics-based regression models and the use of empirical data to create a user- and activity-independent continuous hand trajectory prediction model. Additionally, we propose modifications to established kinematics models, demonstrating how empirical observations can be incorporated as constraints. This focus on mathematically interpretable approaches and empirical calibrations is a key feature of this thesis, ensuring that the resulting models remain sufficiently flexible for diverse users and activities.

Ultimately, this thesis addresses the limitations of existing approaches, advancing human motion modeling and prediction for VR and related applications. It establishes a foundation for richer and more responsive interactions in HCI while enabling safer and more intuitive collaborations in HRI. We believe it addresses the challenges in hand motion prediction, thereby improving its applicability across a wide range of interactive systems.

1.2 Research Questions and Contributions

This thesis aims to answer the following research questions based on the identified research gaps.

> **RQ1:** How can the limitations and challenges of existing hand motion prediction and analysis approaches inform the development of effective solutions for VR applications?
> **RQ2:** How can effective prediction frameworks be developed to address challenges such as accuracy, computational efficiency, and interpretability?
> **RQ3:** How can generalized hand motion models be developed to perform comparably to personalized models?

Overall, this thesis expands the state of the art in hand motion modeling for VR by introducing methods that improve both accuracy and computational efficiency. Specific contributions from each chapter are outlined below.

First, we conducted a user study to gather human hand motion data across a range of structured and unstructured VR activities, addressing the limitations of existing VR datasets. This data collection utilized multiple sensors, including a motion capture system, VR headset sensors, and Inertial Measurement Units (IMUs), resulting in a comprehensive dataset that supports both current and future research. Building on this foundation, we developed a kinematics-based approach for continuous 3D hand prediction within VR. Additionally, we also developed a generalized model that performs comparably to personalized models, yet does not require additional training phases. The model is then thoroughly evaluated through cross-validation and further validated in a secondary study with new participants and activities. This is discussed elaborately in Chapter 3, including the user study, development of the prediction models and results.

Subsequently, we proposed a methodology to extend traditional mathematical models for generating 3D hand trajectories in reaching movements. Using the empirical data collected in Chapter 3, we first demonstrate that existing models fall short in accurately capturing realistic human hand motion. To address this, we frame trajectory generation as an optimization problem, constructing user-specific models tailored to individual motion patterns. Additionally, we develop generalized models, evaluating their performance comprehensively, as discussed in detail in Chapter 4.

For RQ1, we reviewed and discussed the main techniques used for hand motion prediction in Chapters 2 and 3, as well as approaches for hand trajectory generation in Chapters 2 and 4. Across these chapters, we highlighted the limitations and challenges of existing methods, identifying the gaps in current hand motion prediction and trajectory generation techniques. RQ2 (Develop effective models) and RQ3 (Develop generalized models) are both thoroughly addressed in Chapters 3 and 4, with each chapter emphasizing distinct aspects of hand motion prediction.

Chapter 5 revisits the research questions and provides a comprehensive discussion of the thesis contributions, their implications and potential applications. It further outlines the limitations, explores directions for future work, and presents the concluding remarks.

1.3 Thesis Outline

The specific contributions of each chapter are outlined as follows.

Chapter 2 examines prior research efforts in the context of this work. It reviews motion prediction techniques used in VR, various hand motion prediction methods across different domains, and publicly available motion datasets in VR.

Chapter 3 presents a novel hybrid classical-regressive kinematics model, which is a user- and activity-independent prediction model for continuous 3D hand trajectory prediction for VR activities. It also presents a comprehensive user study aimed at gathering VR hand motion data from 20 participants. Data was collected over a one-hour period for each user, covering four different VR activities involving extensive hand movements.

Chapter 4 presents the development of empirically modified minium jerk models for hand trajectory prediction for reaching activities in VR. It examines the limitations of the classical Minimum Jerk Model and introduces novel user- and gesture-specific predictive models where the trajectory prediction problem is formulated as an optimization problem.

Chapter 5 examines the research findings in relation to the thesis research questions, presenting the implications and potential applications of the work, evaluating its contributions and limitations, proposing future directions for human hand motion prediction in VR, and concluding remarks.

CHAPTER 2

Literature Review

This section examines prior work on human motion prediction in virtual reality (VR), highlighting three key areas. First, we focus on the importance of motion prediction within VR environments and explore techniques used, emphasizing hand motion prediction. Next, we explore the kinematics of hand movements, focusing on theories and models for hand trajectory prediction. Finally, we review the current techniques employed for human motion prediction, including classical methods and modern approaches such as deep learning models. This chapter aims to highlight the progress made in the field and identify existing challenges by critically evaluating these areas.

2.1 Motion Prediction in Virtual Reality

In VR environments, users engage with the environment in various ways depending on the activity. These activities include hand gestures [143, 29], such as reaching [27, 47], grasping [151, 149, 150] or throwing [193, 51]. Users may also participate in VR games, such as puzzles [99] or adventure games [46], or engage in exercises [184, 186], or other custom games [183, 192]. Additionally, VR can be used to simulate real-world activities like playing table tennis [173] or skiing [123, 124], offering immersive learning experiences. It is also applied in specialized fields, such as practicing surgeries in medicine [156] and other training applications [185, 113], providing a safe and controlled environment for skill development.

High latency in delivering visual, auditory, or haptic feedback for these interactions disrupts synchronization between user actions and system responses, reducing the immersiveness and causing potential user discomfort. Motion prediction mitigates this issue by allowing the system to anticipate user actions and implement measures such as pre-rendering graphics to reduce delays [72, 138]. By anticipating user movements, VR applications can reduce latency, improve system responsiveness, and create a more immersive and seamless interaction with the virtual environment.

In addition to minimizing latency, motion prediction provides several benefits that enhance the overall VR experience. It can facilitate collaboration, improving interactions between multiple users in shared virtual spaces. Furthermore, motion prediction can enable timely haptic feedback, especially in situations where haptic motors have significant activation latency. In addition, it can improve safety, particularly in augmented or mixed reality applications where the user operates in potentially unsafe environments where physical objects are present. Different types of motion prediction address various aspects of user interaction, including hand, head, and eye movements. While all are significant, this section focuses on hand motion prediction due to its relevance to this thesis.

2.1.1 Hand Motion Prediction

In VR environments, the hand is the primary method by which users interact with their surroundings. Hand gestures can be fast and complex, resulting in rich and dynamic interactions with the environment. Prior work has explored various methods to predict hand movements in VR, aiming to enhance responsiveness and user experience. One common approach involves template matching techniques, where the system predicts future movements by comparing the current motion sequence with a library of pre-recorded motion templates. Hahn et al. [64] employed this method for long-term prediction of hand motions in a working environment, achieving predictions over several tenths of a second. While this technique can be effective for repetitive or well-defined tasks, its applicability is limited when dealing with arbitrary or novel hand motions that are not represented in the template library.

Vu et al. [165] employed a Hidden Markov Model (HMM)-based gesture recognition framework to predict hand gestures using wrist-worn devices, specifically a smartwatch. Their approach aims to identify gestures before their completion, leveraging features derived from

HMM state probabilities. The model's performance was evaluated in the context of a virtual table tennis game. The findings reveal that motion prediction accuracy depends on user expertise. Expert users, with consistent strokes, enable higher precision, while non-expert users' variability leads to reduced accuracy. Additionally, the framework requires gestures to be prefix-free, meaning no gesture can be a sub-category of another, which limits its generalizability to a wider range of applications. Furthermore, the model needs to be retrained to adapt to different applications, adding to its limitations in broader usability.

Deep learning models have also been utilized to improve hand motion prediction in VR. Clarence et al. [27] explored the use of hand motion prediction for the novel application of haptic retargeting in virtual reality (VR). Haptic retargeting enables the decoupling of real and virtual hand movements, allowing the reuse of physical props. By analyzing the initial portion of a user's movement, the model anticipates the target object, allowing the system to respond more proactively. The authors proposed using a Long Short-Term Memory (LSTM) model for predicting intended targets, achieving an accuracy of 81.1% at approximately 65% of the movement. However, the model's applicability is limited, as it is restricted to specific setups and reaching movements.

2.1.2 Other Types of Motion Prediction

Other forms of motion prediction also play important roles in VR systems. Head and eye motion prediction is essential for maintaining visual stability and coherence within the virtual environment. By forecasting the user's head and eye movements, the system can adjust the virtual camera viewpoint accordingly, ensuring that the environment aligns accurately with the user's perspective. This reduces the *motion-to-photon* latency, the delay between the user's head movement and the VR device's display reflecting that movement. It is generally recommended to keep *motion-to-photon* latency below 20 ms, as higher latency can lead to motion sickness [133, 135].

To achieve this low latency, head motion prediction techniques have been extensively explored, focusing on anticipating user head movements to enable real-time adjustments to the virtual environment [12, 60, 59, 141]. Similarly, eye motion prediction methods have been developed to forecast gaze direction, ensuring accurate alignment of the visual display with the user's perspective [8, 56, 114]. In addition, commercial VR headsets are equipped with built-in prediction mechanisms to address the above latency. For instance, the Oculus Rift-S incorporates head pose prediction to estimate the user's head position for the next frame [139].

Henrikson et al. [67] developed a technique that combines hand and head motion data to predict the landing position of a ray pointer in VR environments. By considering the correlation between where a user looks and where they point, the system improves prediction accuracy. However, this approach still relies on template matching and requires a comprehensive library of motion templates. The necessity of such a library poses challenges for predicting arbitrary hand movements, as it is impractical to catalog all possible motion variations.

2.1.3 Limitations and Challenges

The existing methods for hand motion prediction in VR exhibit several limitations that hinder their effectiveness in diverse applications. Techniques based on template matching are constrained by the requirement of extensive motion libraries, which may not encompass the full range of possible hand movements. This reliance restricts the system's ability to predict novel or uncommon gestures, reducing its flexibility in dynamic VR environments.

Deep learning approaches, while powerful in modeling complex patterns, are highly computationally expensive, making them difficult to implement on resource-constrained devices such as standalone VR headsets. Their significant computational and power requirements pose challenges for deployment in such devices, where efficiency is a critical factor.

The current landscape of hand motion prediction in VR demonstrates a range of techniques, each with its strengths and weaknesses. While progress has been made in improving system responsiveness and user experience, significant challenges remain. There is a need for prediction techniques that not only achieve high accuracy but also operate efficiently on resource-constrained devices, ensuring seamless integration into standalone VR headsets. These techniques must be capable of adapting to dynamic VR environments, and should cater to the diverse behaviors and movement patterns exhibited by different users. Addressing these requirements is essential to enhance the scalability, usability, and immersive quality of VR systems across various applications.

2.2 Kinematics of Hand Movements

Understanding the kinematics of hand motion is important for analysing and predicting hand movements. Mathematical models based on control theory and polynomial trajectories have been extensively used for decades due to their simplicity and efficiency. These models aim to accurately represent the human hand's motion, facilitating more natural and responsive interactions within virtual environments.

2.2.1 Classical Kinematic Models of Hand Movement

Fitts' Law [41], proposed in 1954, is one of the earliest models in the field of human movement prediction. It states that the time required to move to a target (T) is proportional to its distance (D) and inversely proportional to its width (W), which can be expressed as $T \propto (2D/W)$ Fitts' Law has been foundational in understanding the speed-accuracy trade-off in human motor control. Although initially developed for one-dimensional pointing tasks, it was later extended to various conditions, including 2D [102] and 3D movements [116]. It was also extended to different environments such as VR [35, 30, 71].

Meyer et al. [111] expanded on Fitts' Law by introducing the stochastic optimized-submovement model, which explains the law through the dynamics of rapid, spatially constrained movements. Their work highlighted that hand movements consist of an initial rapid phase, covering most of the distance, followed by slower corrective sub-movements to ensure accuracy. Through experiments, they validated this decomposition, offering deeper insights into the mechanics of human motion and the interplay between speed and precision.

Another significant contribution to the field is the *Two-Thirds Power Law* which was proposed by Lacquaniti et al. This law states that the velocity of hand movement (v) is proportional to the radius of the curvature of the path (R) raised to the one-third power which can be expressed as $v \propto R^{1/3}$. This means that humans tend to slow down when moving along curved paths and speed up along straighter trajectories. They also demonstrate through experiments that this relationship is valid for a variety of trajectories, including learned and extemporaneous movements, and in the presence of external constraints. While Fitts' Law and the Two-Thirds Power Law offer valuable insights into movement patterns, they are limited in applicability in dynamic and complex environments such as VR due to challenges in generalizability.

2.2.2 Models for Point-to-Point Movements

Another area of research in hand motion is in point-to-point movements, where the hand moves from a stationary initial position to a specific target location. These movements are common in daily activities such as reaching, pointing and grasping. One common approach to modeling point-to-point movements is based on the idea that the human motor system optimizes certain criteria to produce smooth and efficient motions. This perspective suggests that the coordination of muscles and joints is not arbitrary but instead governed by principles aimed at minimizing certain costs, such as energy expenditure, time, or abrupt changes in motion. By focusing on these criteria, various mathematical models have been proposed to model how humans execute hand movements.

The *Minimum Jerk Model (MJM)* proposed by Flash and Hogan [42], suggests that that human movements are planned to minimize the jerk, which is the rate of change of acceleration. Mathematically, this involves finding the movement trajectory where jerk is minimized and this can be expressed as minimizing $\int_{t_0}^{t_f} \left(\frac{d^3x}{dt^3}\right)^2 dt$ over the complete trajectory. Their experiments showed that minimum jerk trajectories closely follow experimental observations with approximately straight trajectories with bell-shaped velocity profiles. This model has been widely adopted in various human-computer interaction (HCI) and human-robot interaction (HRI) tasks, particularly for predicting trajectories in constrained movements like point-to-point tasks. For instance, Bratt et al. [16] utilized the MJM to predict the intended target location in virtual reality ball-catching tasks after the participant had completed the first half of the movement. By applying the MJM, they could anticipate the remaining trajectory of the hand, enhancing the system's responsiveness and the user's interaction experience.

Uno et al. [163] proposed the Minimum Torque-Change Model, which suggests that human movements are planned to minimize the squared rate of change of torque throughout the trajectory. Unlike the Minimum Jerk Model, this approach accounts for parameters such as arm length, payload, and torque, rather than relying only on the initial and final positions of the hand. Uno et al. demonstrated the applicability of this model by comparing it with experimental trajectories of two-joint arm movements. However, this model is more computationally intensive, and requires additional data such as torques, which are not readily available, making it not widely adopted as MJM.

Minimum-Variance Theory, which was proposed by Harris et al. [65] is a common theory for both eye and hand movements, which makes the assumption that neural controls are affected by noise whose variance is proportional to the size of the control signal. This model states that trajectory is planned to minimize the variance of the final eye or the arm position. Moreover, the authors show that minimum variance theory is capable of explaining both the speed-accuracy trade-off in Fitt's law and the two-thirds-power-law. However, due to the dependence on noise parameters, which may vary between individuals and the complexity, this model is also not widely adopted as MJM.

Recent studies have investigated the optimal-control framework from a broader perspective [157, 159, 158]. The key idea is that earlier models such as MJM disregarded sensory feedback for trajectory generation. This framework attempts to incorporate these framework linking high-level goals with real-time sensorimotor strategies for achieving them.

These theories suggest that humans rely on optimization criteria to determine hand trajectories. This implies that identifying the specific optimization criterion should enable accurate prediction of motion trajectories.

2.2.3 Modifications to the Minimum Jerk Model

Despite the widespread adoption of the MJM, subsequent research has identified limitations in its ability to accurately represent real human reaching movements. Gong et al. [54] demonstrated that the traditional MJM does not always align closely with observed human motion trajectories. One key observation was that individual motion patterns vary significantly, suggesting that a single, universal model may not capture the nuances of every person's movements. To address these shortcomings, they introduced the Modified Minimum Jerk Model (MMJM). This model incorporates an error modification term based on a second-order Fourier series to better fit actual human movement data. While the MMJM improved accuracy for specific motions, the variability of the error modification term across different movements made it challenging to generalize the model to all types of reaching actions.

Another approach to refining the MJM was proposed by Todorov et al. [160], who developed a constrained version of the model by integrating principles from the Two-Thirds Power Law. By setting the jerk along the normal (perpendicular) direction to the movement path to zero, they generated velocity profiles that more closely resembled those predicted by the Two-Thirds Power Law. This modification was particularly effective for modeling complex arm movements along curved paths, providing a better representation of natural human motion in such scenarios.

Further studies have explored the limitations of the MJM in dynamic tasks. Svinin et al. [154] found that relaxing the acceleration constraints within the MJM led to a better fit with experimental data during dynamic activities. By allowing more flexibility in acceleration, the model could more accurately reflect the variations observed in actual human movements, particularly in tasks that involve rapid changes in motion.

Wiegner et al. [172] evaluated the applicability of the MJM for single-joint, goal-directed fast movements. Their findings indicated that the Minimum Snap Model (MSM), which minimizes the snap (the fourth derivative of position with respect to time), provided a better fit to experimental data than the MJM. The MSM accounts for muscle and limb dynamics more effectively, especially in modeling fast movements where higher-order derivatives play a significant role.

In addition to these modifications, researchers have explored techniques that utilize early motion data to predict movement outcomes. Ohmura et al. [122] demonstrated that in pointing tasks, the final position of the trajectory could be predicted by detecting the first acceleration peak, typically occurring within the first 21% of the entire gesture. This suggests that early phases of movement contain valuable information for trajectory prediction, offering potential for more efficient and timely predictions.

Lank et al. [86] proposed an endpoint prediction technique for ballistic point-to-point movements based on the MJM. They derived an equation for instantaneous speed over distance and employed polynomial fitting to extrapolate the future endpoint of the movement. To enhance the accuracy of their predictions, they introduced a pre-calculated correction coefficient that was adjusted based on the percentage of the gesture completed. While effective in certain contexts, this method was primarily applicable to specific tasks, such as two-dimensional pointing movements.

These studies highlight the ongoing efforts to refine classical kinematic models to better capture the complexities of human hand motion. While foundational models like the MJM have significantly advanced our understanding, their limitations in accurately predicting diverse and dynamic human movements have prompted the development of modified models and new approaches. Incorporating individual variability, relaxing model constraints, and integrating additional principles from human motor control research have all contributed to creating more precise and generalizable models for hand motion prediction.

2.3 Current Techniques for Hand Motion Predictions

Human motion prediction is a widely studied area due to its applications in various domains, such as computer vision, human-computer interaction, human-robot interaction, and computer graphics. The state of the art goes beyond hand motion prediction, and trying to predict entire human body poses, trajectories and different actions. The main challenge of human motion prediction is the non-linearity, time-varying characteristics and context dependence of human movements. Due to these complexities, researchers have developed a number of techniques to predict human motion, ranging from statistical methods to very complex deep learning models. This section gives a brief overview of these approaches, highlighting their core principles in human motion prediction.

2.3.1 Statistical Methods

Statistical methods form a foundational approach in human motion prediction, using parametric or probabilistic models that incorporate assumptions about movement dynamics. These methods are often grounded in theoretical principles and observations from experimental data. While they have limitations in handling the complexities of non-linear, full-body motion, they have proven effective in specific scenarios, such as predicting hand, head, or eye movements.

2.3.1.1 Kinematics Models

Kinematics models like the Minimum Jerk Model (MJM), characterize human movement as a smooth, natural trajectory resulting from a certain optimization principle. MJM is mainly used for human arm prediction and has been utilized in a number of HCI [16, 44, 7] and HRI [190, 120, 146] applications. Please refer to Section 2.2.2 for discussion on similar kinematics models. These kinematic models, grounded in biomechanics and human motor control theory, provide valuable insights into how humans plan and execute movements. While they are generally simple and computationally efficient, they are less suitable for capturing highly complex motion patterns.

2.3.1.2 Template Matching Techniques

Template matching techniques predict future movements by comparing observed sequences to a library of representative motion sequences. They identify the closest match from the library and extrapolate forward to make the prediction. Template matching performs well in controlled environments with curated databases, such as for gait patterns, leveraging their simplicity and accuracy when motions align with known templates. Template matching techniques are used for predicting trajectory endpoint [67, 53] and gesture recognition [108, 178]. However, reliance on template libraries limits the effectiveness of template matching techniques, especially in scenarios with novel or varied movements, reducing their general applicability.

2.3.1.3 Hidden Markov Models

Hidden Markov Models (HMMs), characterize human motion by representing movements as sequences of hidden states that generate observed states over time. In these models, the hidden states correspond to unobservable movement stages, while the observed states are represented in measurable motion data. During training, existing datasets are used to calculate the transition and emission probabilities that define state dynamics. Motion is then predicted by inferring the most likely sequence of future states based on past observations. HMMs are commonly used in activity recognition tasks in different contexts, such as interactive applications [165, 77], sign language detection [82, 189], fall detection [179, 162, 187], and others [142].

2.3.1.4 Probabilistic Filtering Methods

Probabilistic filtering methods, such as the Kalman filter, treat human motion as a dynamical system defined by a motion model and noisy measurements. The filter has two steps, the predictive step which estimates the next state based on the motion model, and an update step, which corrects the estimate based on the measurements. By iteratively performing these two steps, this technique is widely used for both smoothing noisy trajectory data and short-term prediction. Probabilistic filtering methods, most notably the Kalman filter, are widely used for smoothing noisy trajectory data and short-term prediction. Its non-linear variants, such as the Extended and Unscented Kalman Filters, relax the Kalman filter's linearity and Gaussian assumptions by approximating the state distribution more flexibly, thus providing better accuracy for complex, non-linear movement patterns. Due to their computational efficiency and robustness, Kalman filters are widely used in human motion prediction, including head [80, 61], eye [14, 81] and hand [153, 127] motion. However, the use of these models is also limited due to their simplified assumptions about the motion model.

2.3.1.5 Regression-based approaches

Regression-based approaches predict human motion by learning a direct mapping from input features, such as past motion data, to the future trajectory. These models rely heavily on training datasets and commonly use techniques like polynomial regression [86], support vector regression [92], or more advanced methods such as Gaussian-based [176, 94] or tensor-based regression models [57] for predicting human motion. Once trained, these models can be directly applied to predict future motion, making them straightforward and computationally efficient for continuous trajectory prediction. However, similar to other models described in this section, regression-based models lack generalizability and are heavily dependent on the dataset used for training.

2.3.2 Deep Neural Networks

Deep Neural Networks (DNNs) have emerged as one of the most powerful tools in computer science in the past decade. They are used in various tasks, including human motion prediction, where they can learn the complex relationships in motion data, enabling them to provide accurate predictions. While these methods often achieve higher accuracy and outperform statistical approaches, they require large datasets for training. Additionally, their high computational cost makes it challenging to deploy them on resource-constrained devices, such as VR headsets.

Recurrent Neural Networks (RNNs) [109, 32, 107], including variants like Long Short-Term Memory (LSTM) [52, 20] and Gated Recurrent Units (GRUs) [55, 63], have been widely used to model the sequential nature of human movements. By processing motion data frame-by-frame and maintaining internal states that encode temporal context, these models excel at predicting time series data such as human motion.

Convolutional Neural Networks (CNNs) [98, 32] and Graph Neural Networks (GNNs) [93, 91] approach human motion prediction from a spatial perspective. CNNs, adapted from image analysis, treat spatiotemporal motion data as structured grids, allowing them to capture both local and global movement patterns. GNNs represent the human body as a graph of interconnected joints, making it possible to model complex inter-joint dependencies that influence human motion and gestures. Both CNNs and GNNs have demonstrated strong performance in applications like multi-joint motion forecasting.

Recently, transformer-based architectures have gained attention for their ability to predict motion data [110, 2]. Unlike RNN-based methods, transformers operate on entire motion sequences, leveraging a global perspective to effectively learn the spatial and temporal relationships inherent in human movements.

Autoencoders are a generative technique widely used as an unsupervised learning method for motion prediction. They can learn efficient representations of motion data with dimensionality reduction and have demonstrated effectiveness in handling high-dimensional motion data [5, 18]. Generative Adversarial Networks (GANs) are another generative technique used in motion prediction, employing a probabilistic approach [15, 83]. GANs consist of a generator, which produces future motion data, and a discriminator, which evaluates the likelihood of the generated outcomes.

Transformer-based architectures have recently gained prominence for their ability to handle long-range dependencies in motion data. Unlike RNNs or LSTMs, which process data sequentially, Transformers operate on entire sequences simultaneously using self-attention mechanisms. This global perspective enables them to model intricate temporal relationships effectively, making them well-suited for capturing subtle variations in human motion.

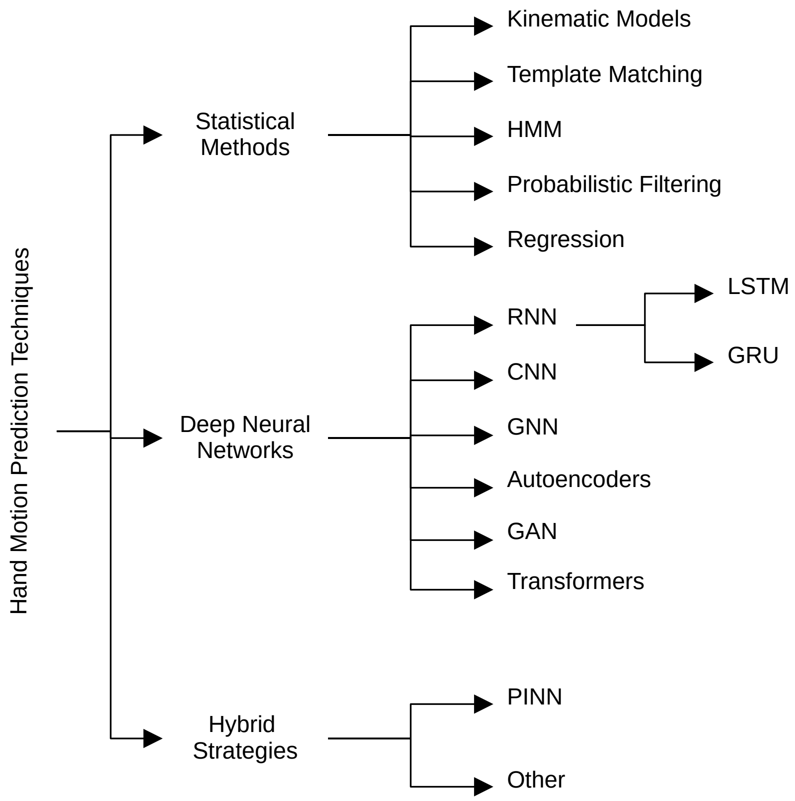


FIGURE 2.1: Overview of the hand motion prediction techniques.

2.3.3 Hybrid strategies

Recently, hybrid approaches that combine DNNs with statistical methods or theoretical frameworks have been proposed for human motion prediction. These models aim to combine purely data-driven DNNs with known theoretical knowledge to combine the advantages of both methods. One prominent example of this hybrid approach is Physics-Informed Neural Networks (PINNs). PINNs [188, 181, 155] incorporate the fundamental principles of mechanics and biomechanics into deep learning models. Instead of relying solely on data patterns, PINNs encode physical laws into the neural network enabling them to produce more realistic and consistent motion predictions.

Other hybrid strategies include combining DNNs with Kalman filter [76], probabilistic models [177]. and human motor control principles such as minimum jerk model [85].

An overview of the hand motion predictions is shown in Figure 2.1

2.4 Motion Datasets in Virtual Reality

High-quality human motion datasets are essential for motion prediction, enabling researchers to analyze and model human movement effectively through analytical and predictive frameworks. These datasets are typically collected using motion capture systems such as OptiTrack [115] and Vicon [10]. Note that this section does not focus on video-only datasets, as they fall outside the scope of this thesis. Notable existing full-body human motion datasets include Human3.6M [75], which includes data from 11 subjects with 15 everyday motions such as walking and posing. The CMU Graphics Lab Motion Capture Database [40] is another well-known database that includes wider range of activities including walking, dancing and sport activities for over 140 subjects. Other large motion datasets include AMASS [104], HumanEva [152], KIT [106] and others [147]. These datasets have made significant contributions in developing human motion analysis and prediction techniques.

While numerous generic motion datasets are available, those specifically collected in VR environments, particularly involving full-body or upper-body movements, remain uncommon. Existing VR motion datasets often have a limited scope, focusing on specific activities such as hand-reaching tasks [28, 89], 360 degree video head movements [31], hand gestures [126], and specific activities such as martial arts [24]. Additionally, datasets centered on motion sickness during gameplay [171] and eye-tracking data [79, 49] are available. However, comprehensive datasets encompassing a wide range of VR activities with full-body or upper-body movements are still lacking.

Clarence et al. [27] collected data from 12 participants, capturing hand and head movements using a VR headset for reaching tasks involving two target configurations in a virtual environment. Participants were seated, and the study utilized a desktop setup. Lento et al. [89] conducted a user study with 20 participants, recording head, gaze, and arm movements along with trunk, shoulder, and arm joint data during reaching activities in a VR environment. However, both datasets are limited to reaching movements, and comprehensive datasets encompassing a wider range of VR activities with full-body or upper-body movements remain unavailable.

For motion analysis in VR, it is beneficial to include explicit VR activities covering a wide range of movements, from structured actions like point-to-point movements to unstructured, free-form activities such as VR gaming. Unlike traditional motion capture scenarios, VR environments introduce additional complexities, including head-mounted displays, handheld controllers, and dynamic virtual interactions, all of which uniquely influence user motion. Consequently, VR motion datasets must not only capture full-body motion data but also include sensor data recorded through the VR headset to fully represent the interaction dynamics. To the best of our knowledge, no such dataset exists at the time of this writing.

2.5 Summary

This chapter provided an overview of prior research related to hand motion prediction in VR. We observed that hand motion prediction offers several benefits to VR systems. While it is primarily used for latency reduction, it also enables novel applications such as haptic retargeting.

As we discussed, hand motion prediction in the VR context is often studied for a limited set of hand movements, with a particular focus on point-to-point actions such as reaching gestures or specific activities. The existing kinematics-based hand motion models and statistical methods offer simple and computationally efficient approaches for hand motion prediction. However, they are unable to handle the wide range of motions involved in VR activities and tend to exhibit lower accuracy when tested with empirical data. Additionally, the widespread adoption of deep learning prediction frameworks in other domains is less apparent in VR. This limitation is largely due to the resource-constrained nature of VR headsets, which are restricted by computational power and energy efficiency.

To develop and evaluate better prediction models, comprehensive VR hand motion datasets that capture a wide range of hand movements are essential. However, there is a noticeable lack of such datasets. These datasets should include not only point-to-point motions but also activities involving greater freedom of hand movement, such as those observed in VR games.

In summary, we identify two key requirements for hand motion prediction models in VR. The first is accuracy, which is essential for effective predictions, and the second is computational efficiency, ensuring compatibility with resource-constrained hardware. Additionally, we recognize two further requirements, generalizability and interpretability, which are important for broader applicability and clearer understanding. Figure 2.2 illustrates the scope and focus of the thesis, in relation to prior research, as conceptualized by the authors in relation to the requirements identified. The predictive models presented in this thesis can be categorized as hybrid approaches that combine statistical methods with empirical data and kinematics techniques. These models retain the computational efficiency and interpretability of traditional methods while achieving improved accuracy and generalizability.

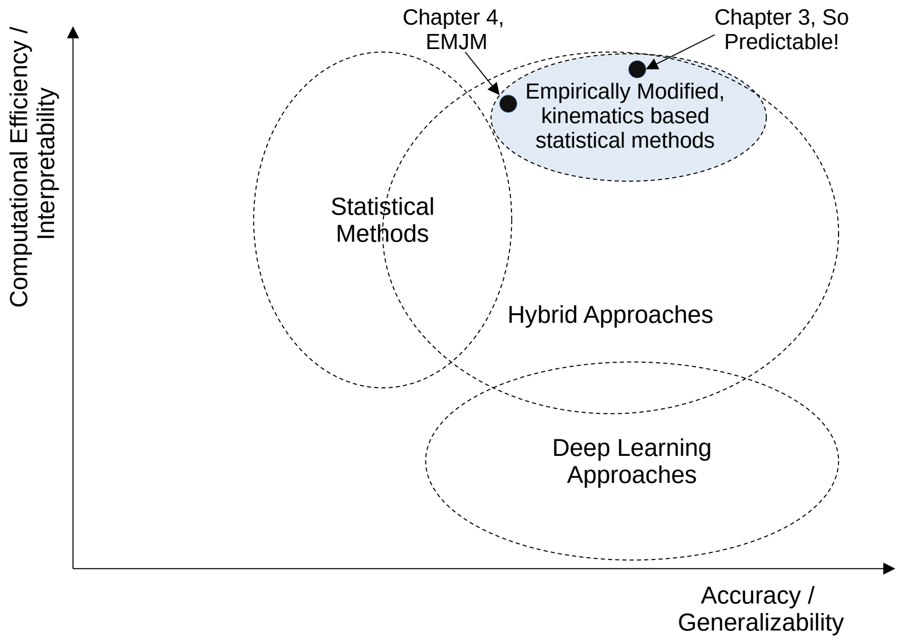


FIGURE 2.2: Focus of the thesis in relation to prior research, as conceptualized by the authors.

To develop predictive models that fulfill these requirements, we design and create two distinct predictive frameworks for hand motion prediction in VR, presented in Chapters 3 and 4. Both frameworks integrate classical statistical methods with kinematics-based techniques and empirical observations. In Chapter 3, we develop a user- and activity-independent 3D hand prediction model capable of continuously predicting a user's hand trajectory. Chapter 4 focuses on developing empirically modified Minimum Jerk Models designed to generate trajectories for reaching movements. Both predictive frameworks are expressed as simple mathematical equations, making them highly computationally efficient. In these chapters, we analyze and interpret these equations, providing insights into how their coefficients influence predictions. Additionally, we thoroughly evaluate the models to demonstrate their accuracy and validate them with new users, activities, and datasets, showcasing their generalizability. Chapter 3 also presents a user study involving 20 participants, where motion data was collected for various VR activities using multiple sensor systems, which was conducted to address the lack of comprehensive hand motion datasets in VR.

CHAPTER 3

So Predictable! Continuous 3-D Hand Trajectory Prediction in Virtual Reality

---

This chapter presents the paper titled "So predictable! Continuous 3D Hand Trajectory Prediction in Virtual Reality", which was published in *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST), pp. 332-343. 2021*. The paper is included in this thesis without any modifications, in accordance with the University of Sydney thesis submission policy.

We contribute a novel user- and activity-independent kinematics-based regressive model for continuously predicting ballistic hand movements in virtual reality (VR). Compared to prior work on end-point prediction, continuous hand trajectory prediction in VR enables an early estimation of future events such as collisions between the user's hand and virtual objects such as UI widgets. We developed and validated our prediction model through a user study with 20 participants. The study collected hand motion data with a 3D pointing task and a gaming task with three popular VR games. Results show that our model can achieve a low Root Mean Square Error (RMSE) of $0.80\,\mathrm{cm}$, $0.85\,\mathrm{cm}$ and $3.15\,\mathrm{cm}$ from future hand positions ahead of $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$ respectively across all the users and activities. In pointing tasks, our predictive model achieves an average angular error of 4.0° and 1.5° from the true landing position when 50% and 70% of the way through the movement. A follow-up study showed that the model can be applied to new users and new activities without further training.

3.1 Introduction

Accurate and timely user interaction tracking is essential in virtual reality (VR) to deliver an immersive experience with high-quality graphics and physics simulations. Despite recent hardware improvements, sensing and computations are still very time and energy consuming, particular for standalone VR headsets [38, 191]. Moreover, a significant delay in feedback such as visual, auditory or haptic would lead the user to notice the asynchrony and break the immersion of the virtual environment [134].

Anticipating future interactions can compensate for these issues. Predicting user activities has been shown to reduce delays and improve the experience in interactive applications [21, 119, 118]. In VR, predictive models using eye gaze tracking [8] and head motion prediction [74, 3] can enable pre-rendering of complex graphics scenarios to reduce latency [138]. Commercial VR headsets such as the Oculus Rift-S employ head motion prediction to estimate the head position for the next frame[1]. The most frequently used input method in VR being arm and

[1] https://developer.oculus.com/documentation/native/pc/dg-render/ (Accessed on 2021-03-26).

a)

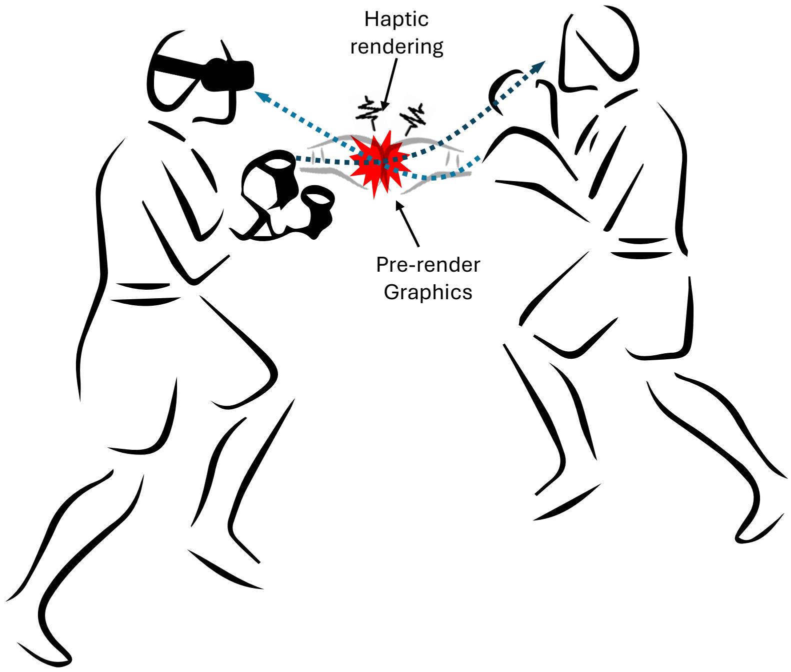


FIGURE 3.1: We present a *hybrid classical-regressive kinematics model* to predict hand motion trajectories in virtual reality. a) Future trajectory can be used to forecast events such as hand collision with other users or non-player characters, enabling pre-rendering of graphics or haptic feedback.

hand movements, recent works by Henrikson et al. [67] and Clarence et al. [27] develop predictive models for pointing and reaching tasks using hands.

Existing models often predict a particular event such as landing on a target [67, 28, 8] or the collective movement [165]. However, in immersive applications, not only the end of the movement, predicting a continuous movement trajectory is important to identify intermediate events. For instance, Figure 3.1a shows a user playing boxing with a virtual character where the location and time of collision between the user's glove and the virtual character is not an endpoint, but a location along the trajectory of the movement. By anticipating the trajectory, such events can be predicted to pre-render rich graphics, calculate complex physics, support real-time multi-modal feedback such as haptics and sounds, and even recover short term

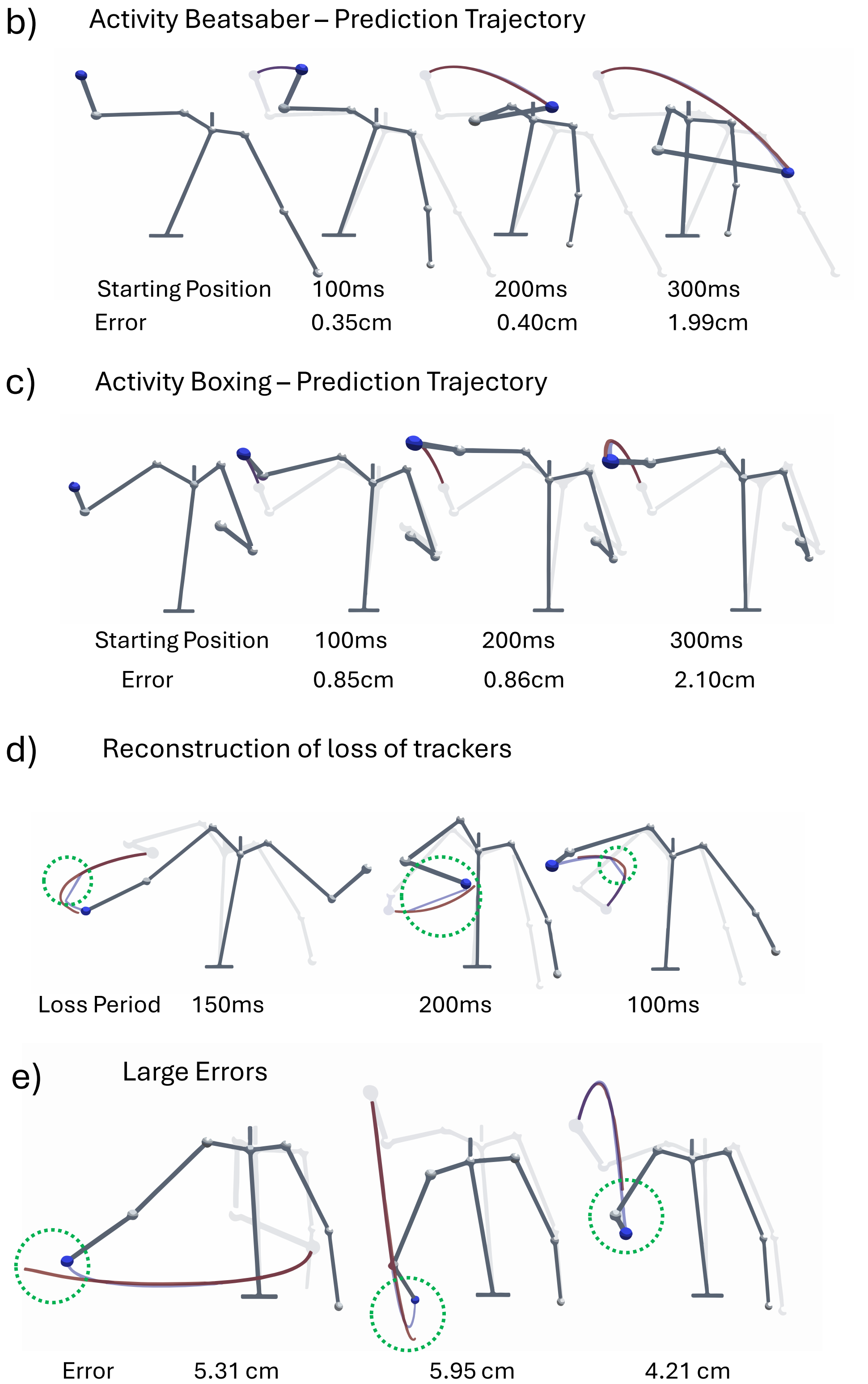


FIGURE 3.1: We present a *hybrid classical-regressive kinematics model* to predict hand motion trajectories in virtual reality. b,c) Comparison of the predicted (red) and the real (blue) trajectories for different prediction intervals (PI) for *BeatSaber* and *FitXR-Box* games, showing average error; d) Prediction model can reconstruct the trajectory when tracking fails; e) Example cases of high errors with sudden changes to movement directions (PI = 300 ms).

tracking errors (Figure 3.1d). Despite the advantages, to the best of our knowledge, little work has applied continuous hand movement trajectory prediction in VR.

This paper contributes a novel *hybrid classical-regressive kinematics model* for continuous 3D hand trajectory prediction for ballistic movements in virtual reality. Our approach uses a coefficient interpolation method between multiple regressions to estimate a unified kinematic model independent of prediction times. We developed and validated this model using data from a user study with 20 participants. Our study collected hand motion data through a structured 3D pointing task and an unstructured gaming task in which the participants played three popular VR games. Both tasks included aimed hand movements, mainly consist of voluntary ballistic movements.

Our findings shows that, a *user- and activity-independent* model performs comparable to personalized and specialized models, and does not require additional training phases. We show that our model can achieve a low average Root Mean Square Error (RMSE) of $0.80\,\mathrm{cm}$ (SD=$0.12\,\mathrm{cm}$), $0.85\,\mathrm{cm}$ (SD=$0.14\,\mathrm{cm}$) and $3.15\,\mathrm{cm}$ (SD= $0.38\,\mathrm{cm}$) from future hand positions ahead of $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$ respectively across all the users and activities. In pointing tasks, our predictive model achieves an average angular accuracy of prediction $4.0°$ (SD =$1.6°$) at 50% of the way and $1.5°$ (SD =$0.6°$) at 70% of the way of the movement. Figure 3.1b and c show a reconstruction of example 3D continuous hand movements for different activities predicted by our model (red) at different prediction intervals (PI) compared to the real movements (blue).

One major challenge of using prediction for continuous hand movements is that prediction models can produce significant errors on some occasions, e.g., during abrupt movements [87]. Error distribution of our model show that such unexpected large errors are minimal in our model with 90% of the errors that occurred are less than $0.6\,\mathrm{cm}$, $0.8\,\mathrm{cm}$ and $3.4\,\mathrm{cm}$ for $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$ across all users and activities.

In summary, this paper makes three main contributions:

(1) A kinematics-based prediction approach for *structured and unstructured ballistic 3D hand movements* in VR activities.

(2) A *user- and activity-independent model* with similar performance to personalized and specialized models without the need of additional training phases.
(3) Evaluation of the model through a cross-validation and a secondary study with new participants and new activities.

3.2 Related Work

This section presents prior work on human motion prediction, its applications in VR and the kinematics of hand motion.

3.2.1 Human Motion Prediction Techniques

The primary goal of human motion prediction is to predict future positions, poses or trajectories of the human body given past motion data. This is a challenging task due to the non-linear dynamics and time-varying behaviour of the movements. Prior work explored various statistical methods [64, 165, 86] and deep-learning methods [109, 90, 68] to tackle the challenges of human motion prediction.

Template matching techniques, where the movement is compared to a library of known template movements [64, 67] is used in human motion prediction. As template matching techniques require building a motion template library first, it cannot be applied for predicting arbitrary movements, which is a major limitation. Hidden Markov Models (HMM) are also leveraged for human motion prediction [165, 169, 96] in the literature. However, similar to template matching techniques, HMMs also require to be trained onset of seed sequences, limiting its usability for predicting arbitrary movements. Regression models allow capturing the important relationships between the predicted values and the predictor variables, which is a major advantage when compared to other prediction methods. In contrast to classification models, regression outputs a continuous value making it better suited for trajectory prediction without requiring a template library or seed sequences. Prior work explores on various regression methods including end point prediction with polynomial regression [86], Electromyography (EMG) based motion prediction [25].

One class of commonly used deep-learning methods for motion prediction are Recurrent Neural Networks (RNN) due to their capability in modelling sequence-to-sequence learning problems [109, 4, 59, 129, 130, 167]. Other classes of deep learning techniques include Convolutional Neural Networks (CNN) [90], graph neural networks [107] and Generative Adversarial Networks (GAN) [68]. However, deep learning methods require large training data sets [74] and have a higher computational overhead, which is not well suited for standalone VR systems with limited computation power.

3.2.2 Motion Prediction for VR

Motion prediction is a key latency reduction technique used in VR, which allows pre-rendering graphics [138, 88]. However, predictive models also enable novel applications such as foveated rendering [8] and haptic retargeting [28]. Commercial VR headsets such as the Oculus Rift-S predict the head pose for the next frame [2]. Predicting the head motion is beneficial for the VR system as it allows the estimation future focus points of the eyes to pre-render future frames. Head motion prediction for VR is further explored in [12, 60, 61, 141]. Saccadic landing point prediction, which estimates the landing position of the fast eye movements [8, 56, 114] is another technique used for pre-rendering.

Hand motion prediction is important for VR as the hand is the primary method of user interaction with the VR environment. Hahn et al. [64] used template matching for long-term prediction (some tenth of a second) of hand motions in a working environment. Clarence et al. [27] proposed a deep learning model to predict the intended target for reaching activities in VR. Henrikson et al. [67] proposed a template matching technique to predict the ray landing position in a VR environment by integrating the head motion into the predictive model. However, due to the requirement of having a template library, this technique is limited when applying for arbitrary hand motions. Vu et al. [165] specifically focused on predicting hand gestures for VR applications and evaluate their performance in a table tennis game. However,

[2] https://developer.oculus.com/documentation/native/pc/dg-render/ (Accessed on 2021-03-26).

they indicate that prediction heavily relies on the expertise of the user to perform the table tennis strokes accurately and show that accuracy drops with non-expert users.

Our method is related to the kinematics-based regression model for endpoint prediction for stylus targeting tasks by Lank et al. [86]. Assuming the start and end velocities to be zero for the pointing tasks, the authors develop a model of speed over the distance that permits extrapolation. However, their technique is limited to pointing tasks in 2D space. In contrast, we utilize a kinematics-based regressive model for continuous motion prediction for *arbitrary hand movements in 3D space*.

3.2.3 Kinematics of Hand Movements

Understanding the kinematics of hand movements is an important aspect of hand motion prediction. Prior work explored dynamic end-effector models for hand movements. *Plamondon's Kinematics Theory* and *Vector Integration to Endpoint* [17] have been proposed to explain the dynamics of hand motion. The *Minimum Jerk Model*, which was proposed by Hogan et al. [69], develops a mathematical model to describe voluntary movements of primates. It was later validated for human arm movements [43]. It states that our nervous system tries to make the smoothest movement possible when performing voluntary movements by reducing accelerative transients. Dynamic models have been proposed for specific activities such as mouse pointing [9]. Recently, Bachynskyi et al. [13] proposed a dynamic hand model integrating third-order lag model for modelling mid-air movements for pointing tasks. Compared to both models [9, 13], we focus on modelling hand motion for arbitrary activities, including structured movements such as pointing *and* unstructured movements in VR games.

3.3 Design Goals

This section outlines important requirements for the implementation of our prediction model. It also discusses the novelty of our system and its advantages for VR applications. The following five goals guide our design and implementation:

3.3.1 Continuous Prediction

Our first goal is to create a model for continuous predictions that is able to estimate the user's hand position at arbitrary time points in the future. In contrast to discreet models, continuous models can be used to predict continuous trajectories of the user's movement. This is important to predict events such as a collision between the user's hand and a virtual object or character as shown in Figure 3.1a.

3.3.2 Structured and Unstructured Motion

Most prior work on hand movement prediction studies movements with a limited and controlled set of user actions, e.g., pointing [67]. However, it remains unclear how such models transfer to movements in more generic VR applications such as VR games where user movements are unstructured and less restrictive. Our goal is to create a single model which is applicable for both structured and unstructured tasks in VR.

3.3.3 User-independent

While the personalization of prediction models can help to improve their accuracy, personalization requires a training phase for each new user before the prediction can be used. This is problematic for settings with many users who might want to use the system for a short time, e.g., museum exhibitions and public displays. Our goal is to create a user-independent model that works for all users. We compare our user-independent model with a personalized kinematics-based model and show it provides a similar prediction performance, without requiring new training data for each user.

3.3.4 Activity-independent

VR applications are used for a wide variety of activities in many domains, including entertainment [170], industry [50, 164], health care [182, 144], sketching [39], and education [45, 26]. The user interactions and activities in different applications can be drastically different. For

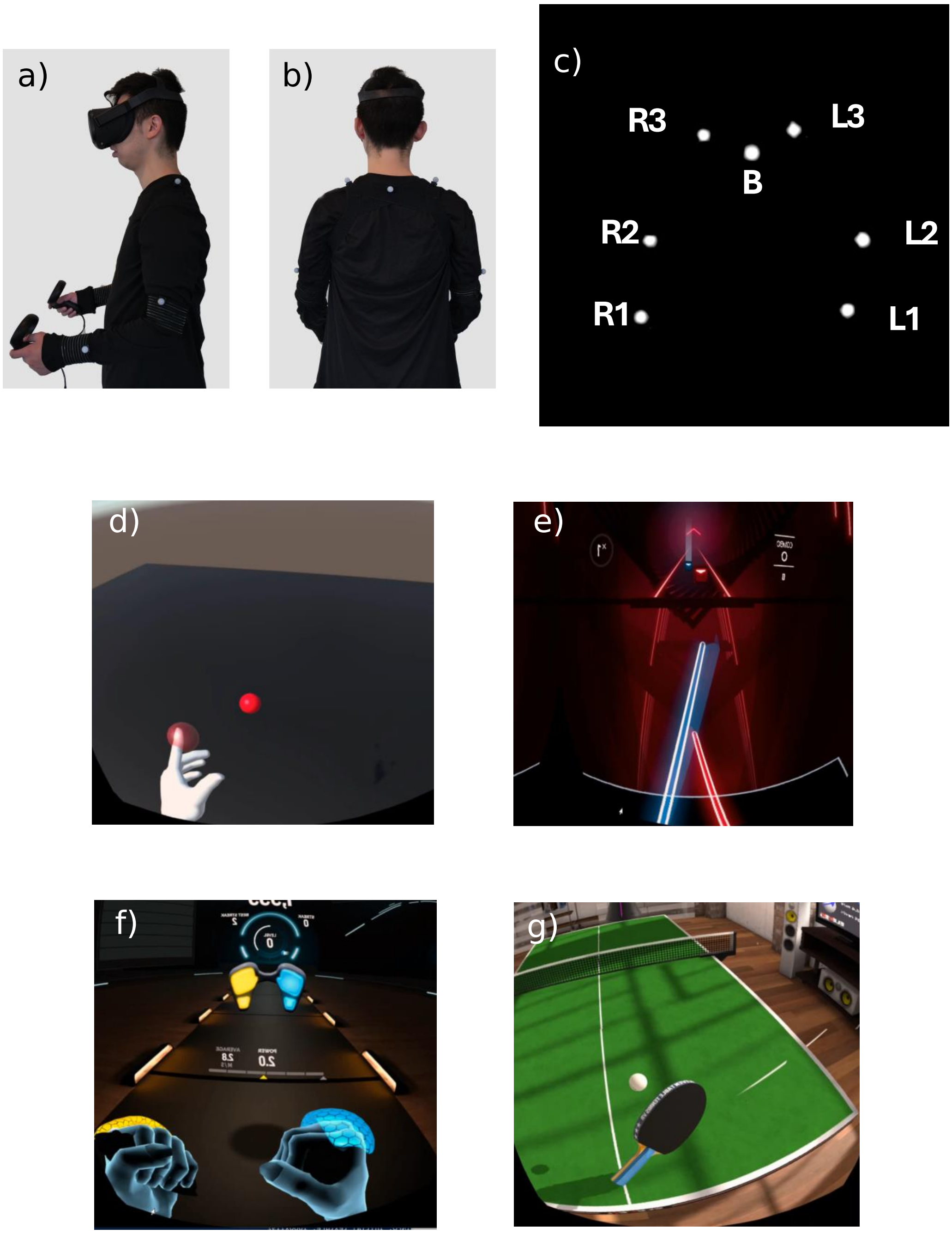


FIGURE 3.2: User study setup: a) side and b) back view of a participant with trackers attached. c) Tracker positions and labels. VR tasks in the user study: d) Custom VR application used in Task 1. VR games of Task 2: e) ©Beat Saber, f) ©FitXR and g) ©Eleven.

example, they can be fast or slow, small or large, and precise or vague. A model specialized in a single activity would require new training sets for each application. Hence, application developers would need to acquire training data and validate it before they could use the model. We believe a prediction model should be generalizable to different activities to allow for simple integration into different fields. In this paper, we develop such an activity-independent model and show that it performs equally well compared to specialized kinematics-based models.

3.3.5 Explainable Prediction

Our final goal states that our prediction model should be explainable to researchers and practitioners. This means that the used methods for the prediction should be transparent and relate to movement parameters. This goal contrasts with typical deep learning techniques such as Long-Short Term Memory (LSTM) neural networks. Although explainable deep learning is under active research [175], most current deep learning methods lead to black-box models. Instead, we contribute a kinematics-based regression model. Such regression models have the advantage that their inner working is fully transparent and all parameters map to movement parameters, such as velocity, acceleration and jerk.

To achieve these design goals, we contribute a novel kinematics-based regression model for predicting arm movements. The model is trained and validated with structured and unstructured hand movement data, which we collected in a user study.

3.4 User Study

We conducted a controlled experiment in a lab environment using a series of VR applications to validate our strategy for continuous hand movement prediction. To show our model would hold for a wide range of hand movements, we gathered hand motion data on applications that focused on structured hand movements (reaching and pointing task) and unstructured hand movements (games).

3.4.1 Participants

We recruited 20 healthy participants (7 female, 13 male; mean age 22.4y SD=5.1y). 19 participants were right-handed and 1 participant was left-handed. Participants who require glasses were allowed to wear them during the study under the head-mounted display. Each participant was given a long sleeves t-shirt to wear, and trackers were mounted on the t-shirt during the study preparation.

The study was conducted according to COVID-19 safety guidelines and the study received ethics clearance from the Human Research Ethics Committee (HREC) of the University of Sydney (Application number: 2019/553). All apparatus was cleaned after each study adhering to the Australian Government regulations.

3.4.2 Apparatus

The hand motion data were recorded with the OptiTrack motion capture system (version 1.10.2) with eight cameras mounted on the ceiling. The participant wore seven trackers on the upper body, which included trackers in participant's wrist, elbow and shoulder on each arm, including a tracker on the back as shown in Figure 3.2a and b. The tracking data was recorded at $100\,\mathrm{Hz}$. Trackers were labeled as shown in Figure 3.2c.

We used an Oculus Quest as the VR headset and Oculus Touch handheld controllers. To collect the structured hand movements, a custom application was developed with Unity3D (Figure 3.2d) where the controller position and orientation were recorded at a rate of $72\,\mathrm{Hz}$ in addition to the data from the OptiTrack system. For the VR games, no motion data from the Oculus was recorded as our 3rd party applications could not access the sensor data.

In addition to the sensor data, the VR screen was mirrored to a computer for screen recording. The whole user study session was video recorded with a camera.

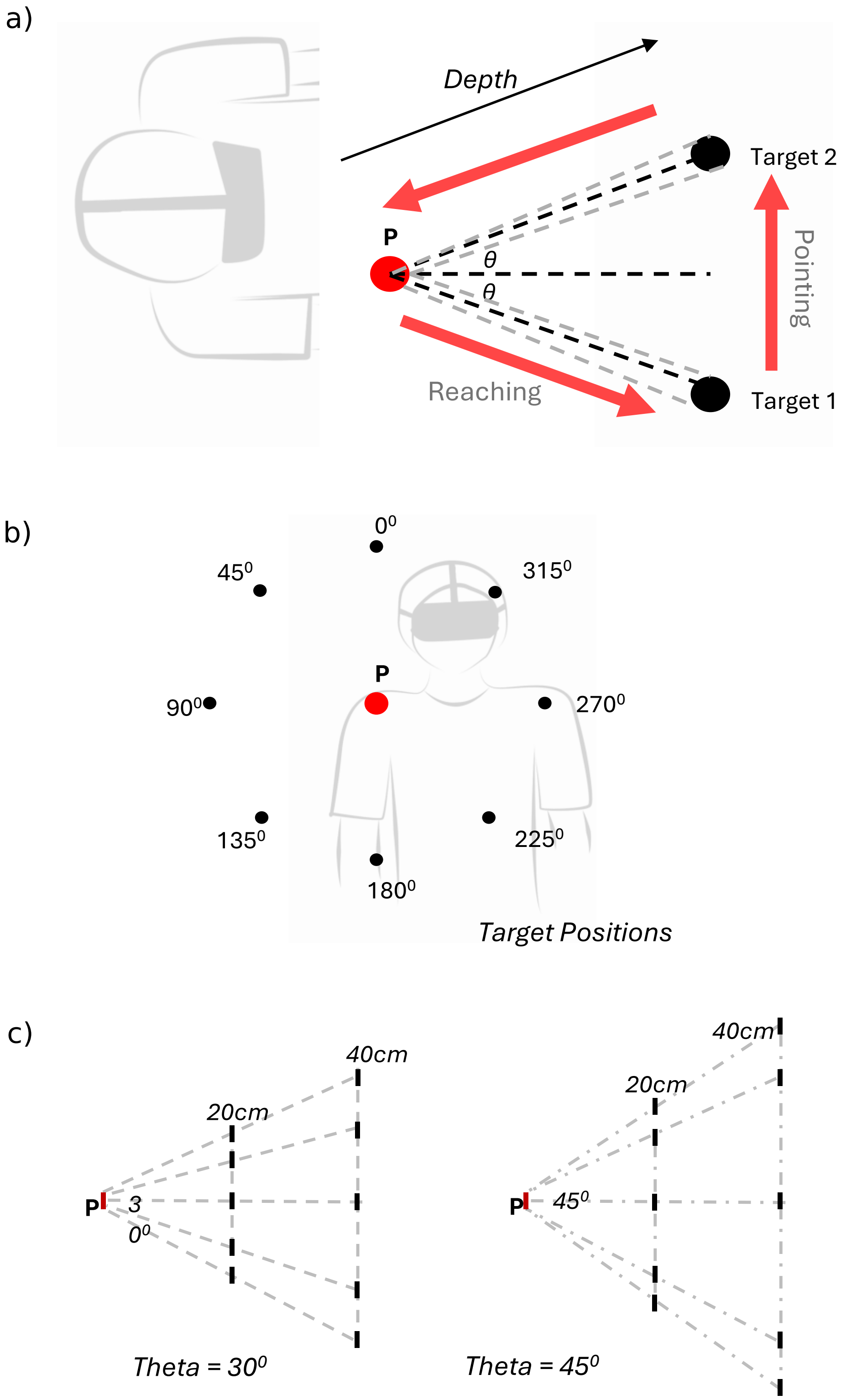


FIGURE 3.3: a) Target positions for a single iteration of the reaching and pointing task (T1). The participant started each trial at starting position P, reached for the first target before pointing to the second target. b) Front-view and c) side-view of the target arrangement.

3.4.3 Study Design

The study was divided into two tasks. Each task collected hand movement data for different activities. The first task (T1) samples *structured* hand movements between indicated points in a three-dimensional pointing study. The second task (T2) samples *unstructured* hand movements from three VR games to increase the external validity of our data set. Both tasks include aimed movements in VR, which contains an initial voluntary ballistic movement followed by a corrective movement [97]. For both tasks, participants were in a standing posture and were instructed not to move their legs during the tasks. However, movements such as twisting, bending the body and leaning to the sides without moving their feet were allowed.

The study was conducted in a single session taking approximately 1 hour per participant. We followed a within-subjects design for the study, where task order was counterbalanced. Participants were allowed to practice until they felt comfortable with the task and to take breaks during the study to prevent fatigue.

T1: Structured Movement via 3D Pointing

The first task followed a repeated-measures, within subject design and collected data from hand movements in all three dimensions. This study opted for a structured approach, where the participant was asked to move their hand towards virtual point targets. The targets were represented as 3D spheres with a diameter of $5\,\mathrm{cm}$. For each participant, the starting hand position was initialized before the experiment with the participant placing their hand in-line with the shoulder while making an approximately $90°$ angle between their forearm and upper arm similar to the study conducted by Cha et al. [22]. At the start of the study, the participant places the virtual index finger on the initial position as seen from the VR headset (red circle in Figure 3.3a and b). When the first target appears, the participant moves their finger from the initial target to the first target (Figure 3.3a-*reaching*). A change of colour indicated to the participant that they successfully reached the target. Afterwards, a second target on the opposite side of the same circle appears and the participant moves their hand from the first target to the second (Figure 3.3a-*pointing*). Finally, the participant moves their finger back to the initial target which completed one iteration of the task. The participant was asked to do all the movements as quickly and accurately as possible.

The targets were equally distributed in four circles in front of the user in $45°$ increments (Figure 3.3b). The circles differed in their distance to the start position ($Depth$) and their angular deviation ($Theta$) as shown in Figure 3.3c. $Depth$ was measured from the starting position of the hand. Our initial experiments indicated that when an angular deviation of $60°$ is used, it is difficult to spot all targets due to the limitations of the viewing angles of the VR headset. To limit the duration of the study, we evaluated only two distances ($20\,\mathrm{cm}$ and $40\,\mathrm{cm}$) and two angular deviations (30° and $45°$). The study contained five iterations of 32 movement blocks (2 depths $\times$ 2 angles $\times$ 8 positions). Participants could take a break after each iteration to prevent fatigue. The movement order was randomized for each iteration to avoid biases. In total, we collected 160 trials (32 movement blocks $\times$ 5 iterations) for each participant.

T2: Unstructured Movement via VR Gameplay

Arm movements in virtual reality applications can be complex. They combine various properties, such as direction, curvature, distance, and speed. In Task 2, we collected arm movement during VR gameplay to ensure our prediction is working in realistic VR scenarios. To cover a wide variety of movements, we asked participants to play three popular VR games (see Figure 3.2e–g).

- *BeatSaber*[3] is a rhythm game where the user needs to slash small cubes with two sabers on both hands. The game contains fast directional slashing movements from both hands.
- *FitXR-Box*[4] is a rhythm game where the user needs to hit small targets using both hands. The game contains fast and powerful forward movements of both hands which closely resembles boxing.

[3] https://beatsaber.com (Accessed on 2021-03-26).
[4] https://fitxr.com (Accessed on 2021-03-26).

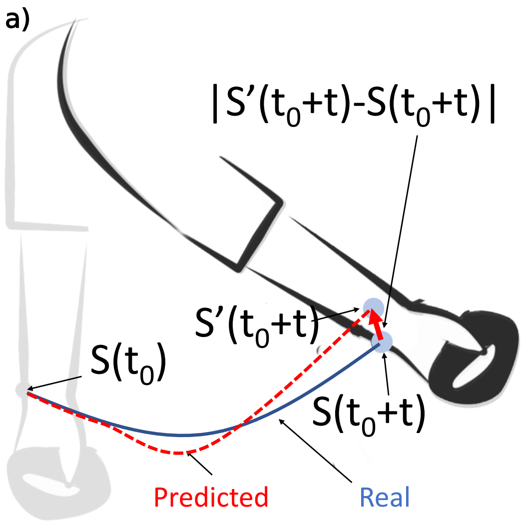


FIGURE 3.4: (a) Hand trajectory prediction at a given time t_0, from the position $S(t_0)$ to position $S(t_0 + t)$ at time t

- *Eleven*[5] closely resembles real-world table tennis strokes with the dominant hand.

With an initial study, we observed that each game had different move dynamics. Due to the fast and wide slashing movements in *BeatSaber*, it had the highest average speed ($0.72\,\mathrm{m/s}$) and highest spans in horizontal and vertical directions ($0.85\,\mathrm{m}$, $0.95\,\mathrm{m}$). *FitXR-Box* had a much lower horizontal span ($0.57\,\mathrm{m}$) and the highest span in frontal direction ($0.75\,\mathrm{m}$) as expected from the boxing movements. Meanwhile, move dynamics of *Eleven* varied greatly among users, as individuals have different styles for playing table tennis.

Each game was played for approximately three minutes. Before recording the data, participants could get familiar with the game by following the in-game tutorials and playing the game for 1 minute. Participants could take breaks between games.

3.4.4 Data Preparation and Presentation

We used data from the Optitrack motion capture system for training and testing of our prediction models. To reduce the noise introduced from the trackers, we apply a Gaussian

[5] https://linktr.ee/elevenvr (Accessed on 2021-03-26).

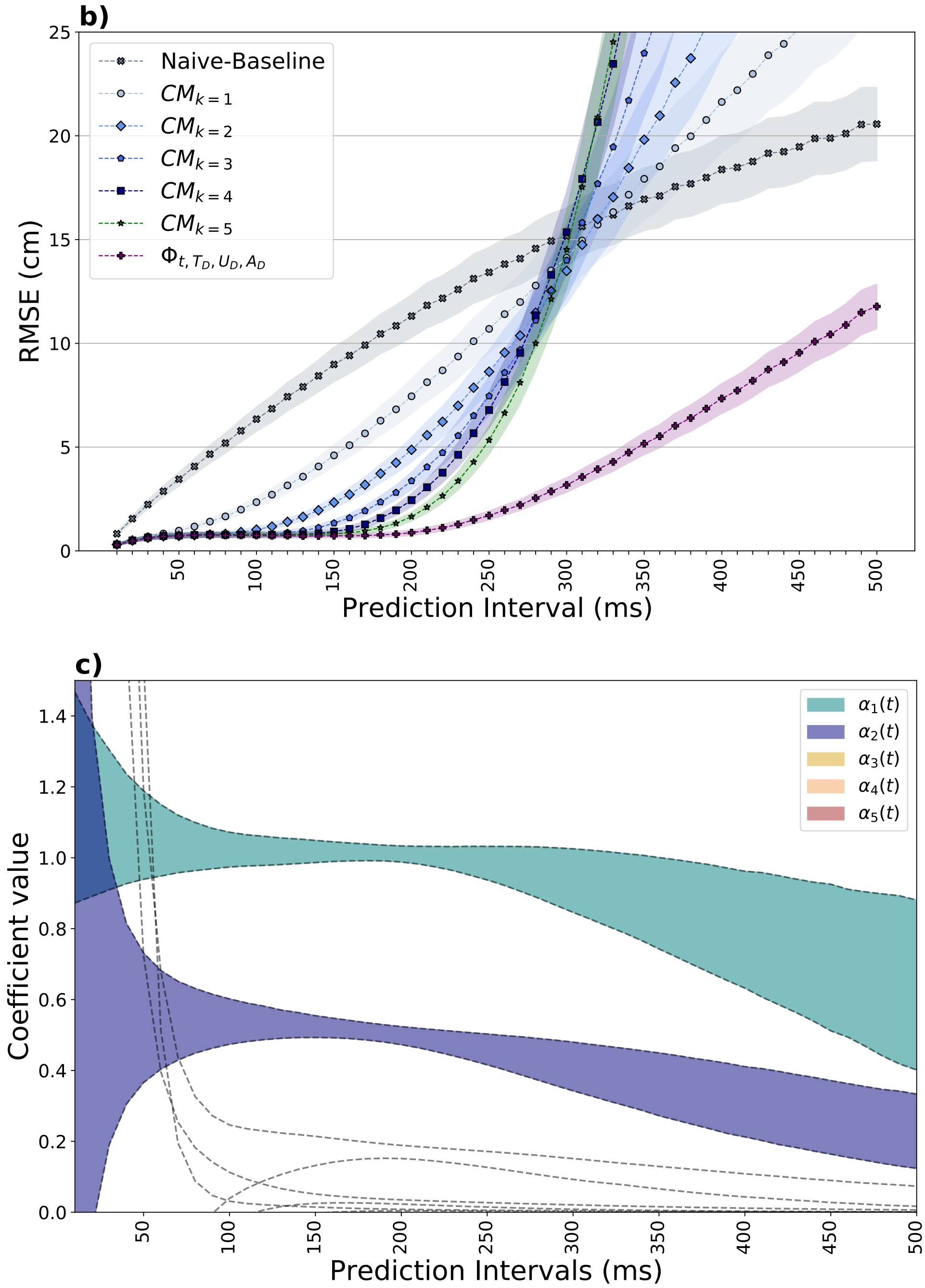


FIGURE 3.4: (b) Average RMSE (in cm) from classical kinematic models showing improvement with each added derivative of motion compared to a series of prediction-time dependent regressive kinematic models Φ_{t,T_D,U_D,A_D} developed and tested for each task, user and axis; (c) Distribution of 95% confidence interval of five α_ns from Φ_{t,T_D,U_D,A_D} models.

Filter to smooth the trajectory similar to prior work [67]. For the hand trajectory we used the position of the marker R1 or L1, $3.5\,\mathrm{cm}$ above the wrist (Figure 3.2e) relative to the body frame with reference to centre of the shoulder plane, i.e., marker B shown in Figure 3.2e.

3.4.4.1 Training – Testing Data Split:

The data collected was split to training and testing portions upon collection. We used $30\,\mathrm{s}$ (<10% of data from T1 and 15% of data per game from T2) as the training set. The rest was allocated for testing. Therefore, all the testing we conducted throughout this paper was conducted on data independent from the training data.

3.5 Hybrid Kinematic Regressive Model

Our predictive model uses classical kinematic equations as the base of a multi-layer regressive model. This section explains the process used to develop the model with an overview of metrics used to evaluate prediction accuracy, classical kinematics, prediction-time dependent kinematics regression and inferred prediction-time independent kinematic modelling.

3.5.1 Metrics for Prediction Accuracy

Before developing the model, it is important to identify metrics to compare the performance of the prediction accuracy. In this section, we explain our definition of the prediction error per each instance and the aggregated accuracy measures we used to compare the developed models.

Figure 3.4a depicts the prediction error $|S'(t_0 + t) - S(t_0 + t)|$, the distance between the predicted ($S'(t_0 + t)$) and the actual ($S(t_0 + t)$) hand locations where t_0 and t are initial time and the prediction time interval respectively. To add a relative comparison, it is common in predictive models to use the initial point as a *naive prediction* (i.e. $S'(t_0 + t) = S(t_0)$) [67]. In other words, *naive prediction* assumes the hand does not move during the predicted time period. As $S(t_0 + t)$ is the ground truth at time $t_0 + t$, *naive prediction* error can be expressed as $|S(t_0+t)-S(t_0)|$, which is the actual displacement of the hand during the predicted time period. Therefore, we take *naive prediction* as a *naive baseline* to compare our prediction models with the actual movement.

The prediction errors defined above apply to a single point in the trajectory. Since our aim is to predict continuous hand motion trajectory, we need to aggregate the per point prediction errors to a single metric. Root Mean Square Error (RMSE), or RMSE of $|S'(t_0+t)-S(t_0+t)|$, is a commonly used aggregated metric to evaluate predictive trajectories [36]. Another commonly used metric is the Mean Absolute Error (MAE) [78]. We chose RMSE as the primary metric since it gives a higher weight for larger errors in prediction due to the squared terms. Therefore, RMSE is better suited when higher errors are particularly undesirable, which is important in trajectory prediction.

Furthermore, Nancel et al. [119] showed that while RMSE provides an overall measure for the accuracy, it does not capture side effects of latency compensation methods from a user's perspective. They identified seven spatial accuracy metrics to capture the side effects for touch location prediction in 2D touchscreen. We extended their metrics Lateness (slow to react to the actual movement), Over-anticipation (over-react to the actual movement) and Wrong Orientation (not going in the same direction as the motion) to 3D space. We also use these additional metrics to compare our final model with the baselines.

3.5.2 Classical Kinematics of Motion

Classical kinematics can be used to model behavior of moving bodies with respect to a frame of reference. As shown in Figure 3.4a, given the three-dimensional hand position vectors $S(t_0)$ and $S(t_0+t)$ at times t_0 and t respectively, $S(t_0+t)$ can be expressed as:

$$S(t_0+t)=\sum_{n=0}^{k}\frac{1}{n!}\frac{\mathrm{d}^n S(t_0)}{\mathrm{d}t^n}t^n \tag{3.1}$$

This equation assumes that $\frac{\mathrm{d}^k S(t_0)}{\mathrm{d}t^k}$ to be constant. For instance, movements with constant acceleration, where the second derivative is constant. We can set $k=2$ to get the equation for displacement $D(t)$ (i.e., $D(t)=S(t_0+t)-S(t_0)$), in the familiar form $D(t)=ut+\frac{1}{2}at^2$, where $u = \frac{\mathrm{d}S(t_0)}{\mathrm{d}t}$ and $a = \frac{\mathrm{d}^2S(t_0)}{\mathrm{d}t^2}$ are velocity and acceleration. However, acceleration of the hand movements are not constant and changes with time with the forces exerted by muscles. Therefore, it is important to identify a k, which is low enough to make the model plausible and high enough to accommodate real hand movements. Hogan et. al. show that voluntary hand movements in mammals follow a *minimum jerk law*, where *jerk* (j) is the 3rd derivative of the motion. The law states that the nervous system tries to make the smoothest movement possible by reducing accelerative transients. The law further states that pointing movements would have a constant *crackle* (c), which is the 5th derivative of the motion [86, 140]. Therefore, we tested the classical models (CM_k) for $k \in [1, 5]$, assuming a constant *crackle*. For instance, $CM_{k=5}$ results in $S'(t_0 + t) = S(t_0) + vt + \frac{1}{2}at^2 + \frac{1}{6}jt^3 + \frac{1}{24}st^4 + \frac{1}{120}ct^5$, where s is the 4^{th} derivative named *snap*.

We used the data collected in our experiment to calculate future positions of the trajectory using these classical models (CM_k) at 50 prediction intervals at $10ms$ steps from $10ms$ to $500ms$. Figure 3.4b shows the average RMSE across users and activities under each k values, where RMSE values are calculated for the prediction error $|S'(t_0+t)-S(t_0+t)|$. Figure 3.4b also shows a comparison with the *naive baseline* where $S'(t_0 + t) = S(t_0)$, which is the classical model with $k = 0$ $(CM_{k=0})$.

The classical model with $k = 5$, $CM_{k=5}$, generated low average RMSEs until $t = 0.16s$ $(mean = 0.5cm, SD = 0.07cm)$. However, after $t = 0.18s$, data shows that the time varying nature of real hand movements is difficult to capture in these equations and it exponentially overestimates the movements. In Figure 3.4b, it is important to note each added derivative contributes to better predictions at smaller prediction intervals, but at larger prediction intervals increasingly contributes to the error. Since classical models with $k = 5$, $CM_{k=5}$, is the best performing classical model, we use it as a *non-naive baseline* for comparison with our models.

3.5.3 Prediction-Time Dependent Kinematics Regression

In Figure 3.4b, it is evident that the motion characteristics such as u, a, j, s and c are essential to estimate future locations; with increasing prediction intervals, their contributions with constant weights in the classical kinematics (e.g., $\frac{1}{2}, \frac{1}{6}, \frac{1}{24}, ...$) leads to an exponential error. This can be explained by the accumulation of errors in the integrative nature of the equation (i.e. $u = \int_{t_0}^{t} a$). Therefore, to counter future changes to higher order derivatives, even beyond *crackle* (c), prediction-time (t) dependent weights for each derivative in Equation 3.1 are needed. Essentially, this can be expressed as:

$$D(t) = \begin{bmatrix} vt & at^2 & jt^3 & st^4 & ct^5 \end{bmatrix} \times \begin{bmatrix} \alpha_1(t) \\ \alpha_2(t) \\ \alpha_3(t) \\ \alpha_4(t) \\ \alpha_5(t) \end{bmatrix} \tag{3.2}$$

Where each $\alpha_{n(t)}$ represents a three dimensional variable (for three axis of movement), specific to a given prediction interval t, and $D(t) = S(t_0 + t) - S(t_0)$. We considered identifying each $\alpha_{n(t)}$ as a regressive problem. To develop regressive models for each t, we used the training portion of the movement data collected in our user study. Our attempt to fit a model that takes prediction time as an independent variable failed with $R^2 < 0.3$ even for prediction times $t < 250ms$. Therefore, we considered creating *prediction time dependent* models where inputs to the regression was $\{v_{(t_0)}t, a_{(t_0)}t^2, j_{(t_0)}t^3, s_{(t_0)}t^4, c_{(t_0)}t^5\}$ where t_0 is current sample time and t is the prediction time interval.

For increased granularity across prediction time (t), we used 50 time steps in $[10ms, 500ms]$ range in par with our sampling interval of $10ms$. To accommodate the user, task and prediction time interval dependent nature of the movements, we regressed 4000 $(50 - timesteps \times 20 - users \times 4 - task)$ independent models resulting $5\ \alpha_n$s per model (i.e., 4000 - α_1, 4000 - α_2, etc.), each 1×3 vector representing three axes. We represent these three models as Φ_{t,T_D,U_D,A_D} representing, prediction **t**ime, **T**ask, **A**ctivity and **U**ser **D**ependent nature of the model.

To validate each model, we used the test data portion from the user study to calculate RMSE for each model with respect to the conditions the models created against, i.e. relevant to prediction interval (t), users (U) and tasks (T). Figure 3.4b shows the average RMSE across each model in comparison to the *Naive Baseline* and Classical models (CM_k). A Mann-Whitney test indicated that for higher prediction intervals of $t > 160ms$, the 5th order classical model has a prediction error ($median = 0.44$) significantly greater than for Φ_{t,T_D,U_D,A_D} ($median = 0.35$), $U = 2622$, $p = 0.024$, and shows a proportional increment of error with t. Therefore, Φ_{t,T_D,U_D,A_D} can be used as the best possible model, however, it is too specific (user, task) and the prediction time, and will be implausible to apply in a realistic scenario.

3.5.4 Inferred Prediction-Time Independent Kinematic Modeling

The major challenge of the Φ_{t,T_D,U_D,A_D} is that any practical system needs to maintain a series of models (4000 in our case) for each specific scenario. And each new factor will exponentially increase the number of models. Furthermore, any change in the coordinate system will need either coordinate translation or re-calibrations. For real-life applications, a general model is desired, where minimum or no additional training needed for each new scenario. However, Φ_{t,T_D,U_D,A_D} showed great promise in the regressive nature of the hand movements. A generalizable regressive model would be ideal for the fast prediction of future movements.

The fitted α_ns in the Φ_{t,T_D,U_D,A_D} models can be used as a basis to develop an *inferred regression model*, which can be generalizable and independent of the scenario. Figure 3.4c examines the distribution of α_n in each model in Φ_{t,T_D,U_D,A_D}, with the regions indicating the 95% confidence interval at each prediction time. The model shows a converging pattern towards higher prediction intervals (t), but at lower ts, it shows high variations. This is due to the regressive model trying to accommodate large variability of higher derivatives of movements (α_3, α_4 and α_5). This intern affects the regressed coefficient of lower derivatives. A straightforward approach is to create a general model of each α_n as a regression of the distribution in Figure 3.4c. However, this resulted in poor fitting with $R^2_{\alpha_1} = 0.77$, $R^2_{\alpha_2} = 0.38$, $R^2_{\alpha_3} = 0.07$, $R^2_{\alpha_4} = 0.07$, $R^2_{\alpha_5} = 0.08$ for each α_n. Also, we observed that, classical models perform equally well until t reaches $0.16s$ (first statistically difference as

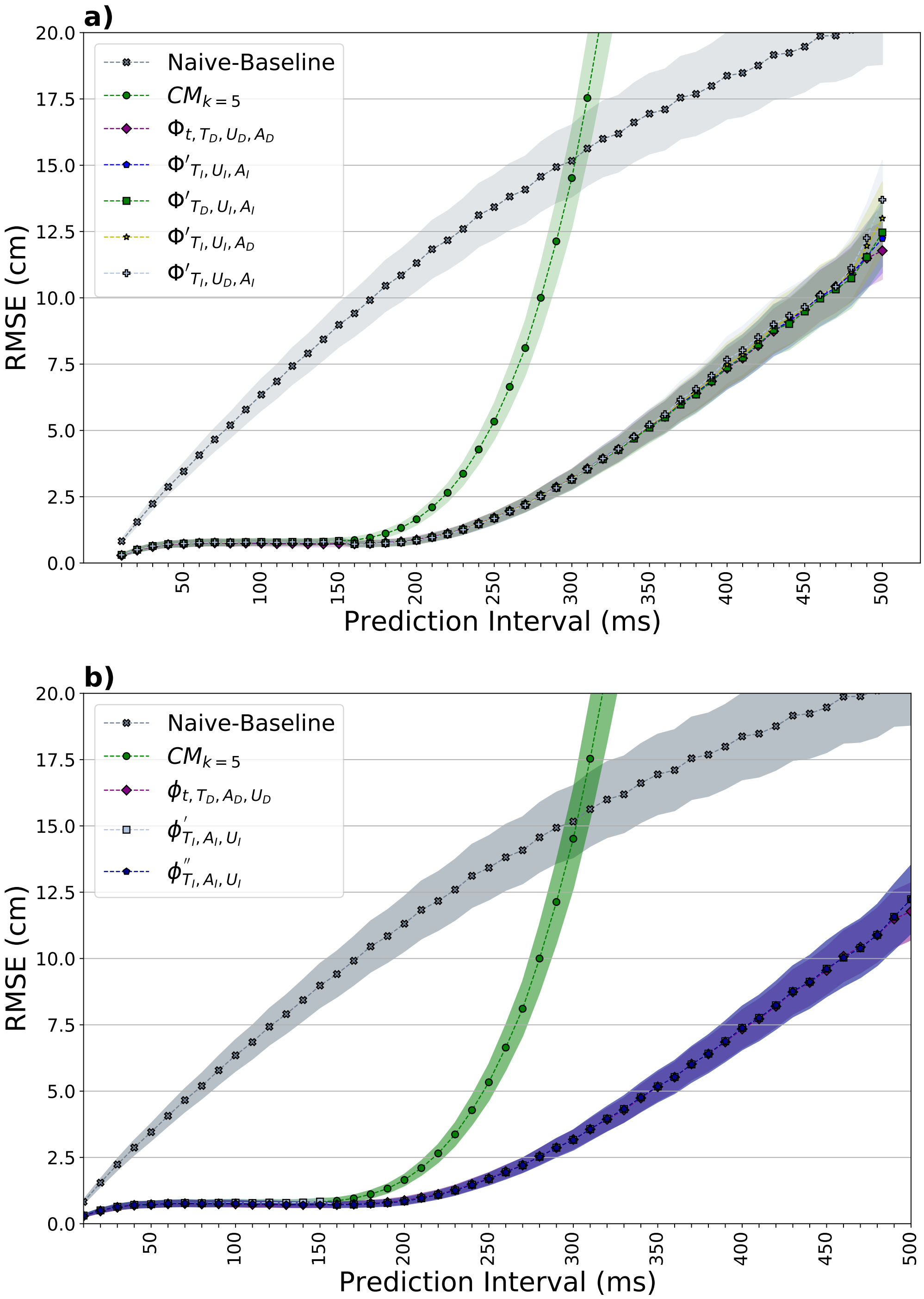


FIGURE 3.5: a) Average RMSE for models Φ'_{T_D,U_I,A_I}, Φ'_{T_I,U_D,A_I}, Φ'_{T_I,U_I,A_D} and Φ'_{T_I,U_I,A_I} compared to the baselines; b) Average RMSE of model Φ''_{T_I,U_I,A_I} compared to Φ_{t,T_D,U_D,A_D}, Φ'_{T_I,U_I,A_I} and the baselines;

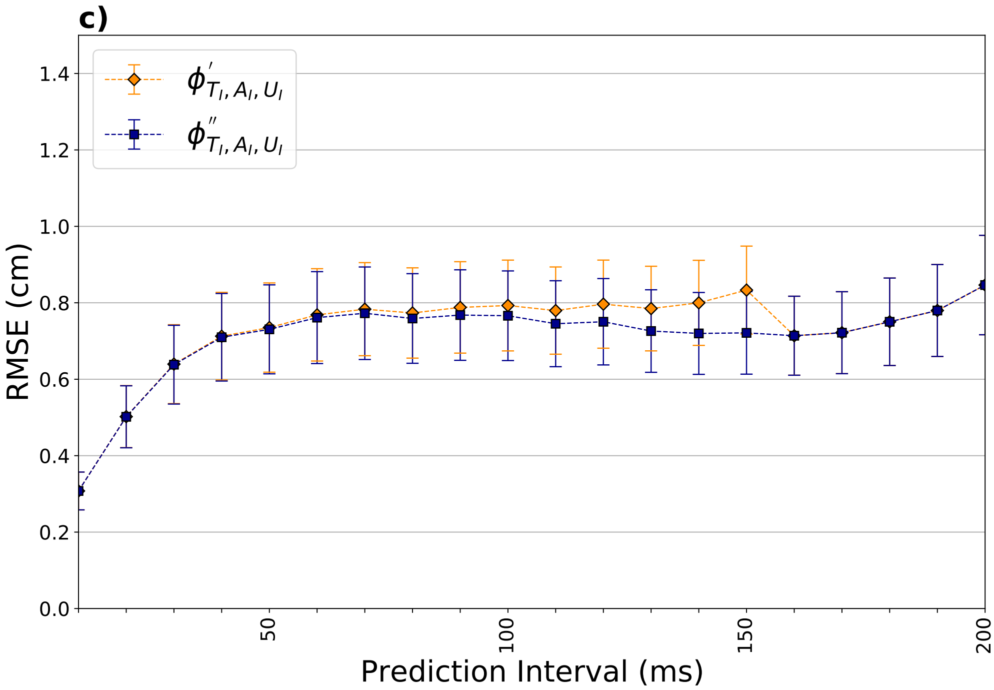


FIGURE 3.5: c) Comparison of the two models Φ'_{T_I,U_I,A_I} and Φ''_{T_I,U_I,A_I} in $t = [10, 200]$ region;

compared in Figure 3.4b). Therefore, a hybrid approach of classical and regressive models is necessary to capture the high performing parts of each method. We considered two piecewise approaches to create two hybrid models.

3.5.4.1 Direct Classical + Regressed Piecewise Models (Φ'):

In this approach, we directly replaced the first portion of the regressive model until the time interval (t), where RMSE reached a statistically significant advantage with the model Φ_{t,T_D,U_D,A_D}. Specifically, we created a piecewise split of the coefficient function at $t = 0.16s$.

$$\alpha'_n(t) = \begin{cases} \alpha_n^c & t < 0.16s \\ \alpha_{n(t)}^r & t \geq 0.16s \end{cases} \tag{3.3}$$

Where, $\alpha_n^c = \frac{1}{n!}$ are constant classical α values and $\alpha_{n(t)}^r$ represents the regressed and time dependent values, which can be expressed as a second order polynomial $\alpha_{n(t)}^r = \beta_0 + \beta_1 t + \beta_2 t^2$.

In order to compare the effect of task, user and axes dependency, we regressed $\alpha^r_{n(t)}$ with each factor dependent and a final model which is completely independent of all the factors. Each of these models is denoted by Φ'_{T_D,U_I,A_I}, Φ'_{T_I,U_D,A_I}, Φ'_{T_I,U_I,A_D} and Φ'_{T_I,U_I,A_I} where T, U, and A indicates task, user and axes and D or I indicates dependency or independency. The series of regressive models had an average R^2 of 0.85 with a SD=0.06.

Figure 3.5a shows the average RMSE against prediction interval for four models compared to the Φ_{t,T_D,U_D,A_D} model and the baselines. Surprisingly, we found no significant difference (Mann-Whitney test) of the task, user or axes dependency in the comparison. This is indicative of a task, user and axes independent model that can be developed for hand movement prediction without any personalized training. However, in the results, we noticed a sudden drop of RMSE at the piecewise junction ($t = 0.16s$). This is indicative of a gradual transition from classical to the regressive model may further increase the performance.

3.5.4.2 Interpolated Classical + Regressed Piecewise Models (Φ''):

In the second hybrid model, our goal was to implement a gradual transition from the classical model to a regressive model. Rather than setting fixed classical values when $t < 0.16s$, we included an interpolation between the classical model and the regression model, resulting in new piecewise definition of the function:

$$\alpha''_n(t) = \begin{cases} \alpha^c_n + \frac{\alpha^r_{n(0.16s)} - \alpha^c_n}{0.16} t & t < 0.16s \\ \alpha^r_{n(t)} & t \geq 0.16s \end{cases} \tag{3.4}$$

This model holds the same behaviour for Φ' for prediction intervals greater than 0.16 s. Since all independent approaches should hold for the most challenging prediction intervals, we tested RMSE for this model only for all independent regressive configuration, Φ''_{T_I,U_I,A_I}. Figure 3.5b shows the average RMSE against prediction interval for Φ''_{T_I,U_I,A_I} compared to the Φ_{t,T_D,U_D,A_D}, Φ'_{T_I,U_I,A_I} and the baselines. We did not observe any significant difference between the models with the Mann-Whitney test. Results show RMSEs of $0.80\,\mathrm{cm}$ (SD=$0.12\,\mathrm{cm}$), $0.85\,\mathrm{cm}$

(SD=$0.14\,\mathrm{cm}$) and $3.15\,\mathrm{cm}$ (SD=$0.38\,\mathrm{cm}$) from future hand positions ahead of $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$ respectively across all the users and activities.

Figure 3.5c shows a comparison of the two models Φ'_{T_D,U_I,A_I} and Φ''_{T_D,U_I,A_I} in $t = [10, 200]$ region where the two models differ. The figure shows that Φ''_{T_D,U_I,A_I} outperform Φ'_{T_D,U_I,A_I} in the region $t = [70, 150]$ and that average error of Φ''_{T_D,U_I,A_I} has improved in the prediction interval $t = [70, 150]$. However, we did not find a statistical significance with the Mann-Whitney test. The transition of the error from classical to regressive model has become smoother in the Φ''_{T_D,U_I,A_I}. Therefore, we conclude that the combination of the interpolated classical model and the task, user and axes independent regressed model, Φ''_{T_D,U_I,A_I}, resulted in the best outcomes for prediction.

3.5.5 Results

Our final predictive model (Φ''_{T_D,U_I,A_I}) achieves RMSEs of $0.80\,\mathrm{cm}$ (SD=$0.12\,\mathrm{cm}$), $0.85\,\mathrm{cm}$ (SD=$0.14\,\mathrm{cm}$) and $3.15\,\mathrm{cm}$ (SD=$0.38\,\mathrm{cm}$) from future hand positions ahead of $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$ respectively across all the users and activities. Compared to *naive baseline* and $CM_{k=5}$, our model reduces RMSE by 79.1% and 78.1% for 300, $500\,\mathrm{ms}$. Our model achieves MAEs of $0.28\,\mathrm{cm}$ (SD=$0.19\,\mathrm{cm}$), $0.33\,\mathrm{cm}$ (SD=$0.23\,\mathrm{cm}$) and $1.97\,\mathrm{cm}$ (SD=$1.10\,\mathrm{cm}$) for $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$ PIs, which is in the same order as the tracking errors of the commercial VR headsets (Oculus Quest 0.69cm, Samsung Galaxy S9 1.69cm) [128].

To compare our predictive model for pointing tasks, we predicted the landing position of each pointing task at 10% increments along the way from start to end similar to Henrikson et al. [67]. Figure 3.6a shows the average angular accuracy of models Φ''_{T_D,U_I,A_I} compared to the *naive baseline* and $CM_{k=5}$. New Φ''_{T_I,U_I,A_I} creates average angular accuracy of prediction $4.0°$ (SD =1.6°) at 50% of the way and $1.5°$ (SD =0.6°) at 70% of the way of the movement. At 50% of the way, Φ''_{T_I,U_I,A_I} model shows a reduction of angular error by 74.5% and 74.4% compared to the *naive baseline* and $CM_{k=5}$ respectively. Similarly, for 70% of the way, the error reduction of our model is 82.4% and 66.9%.

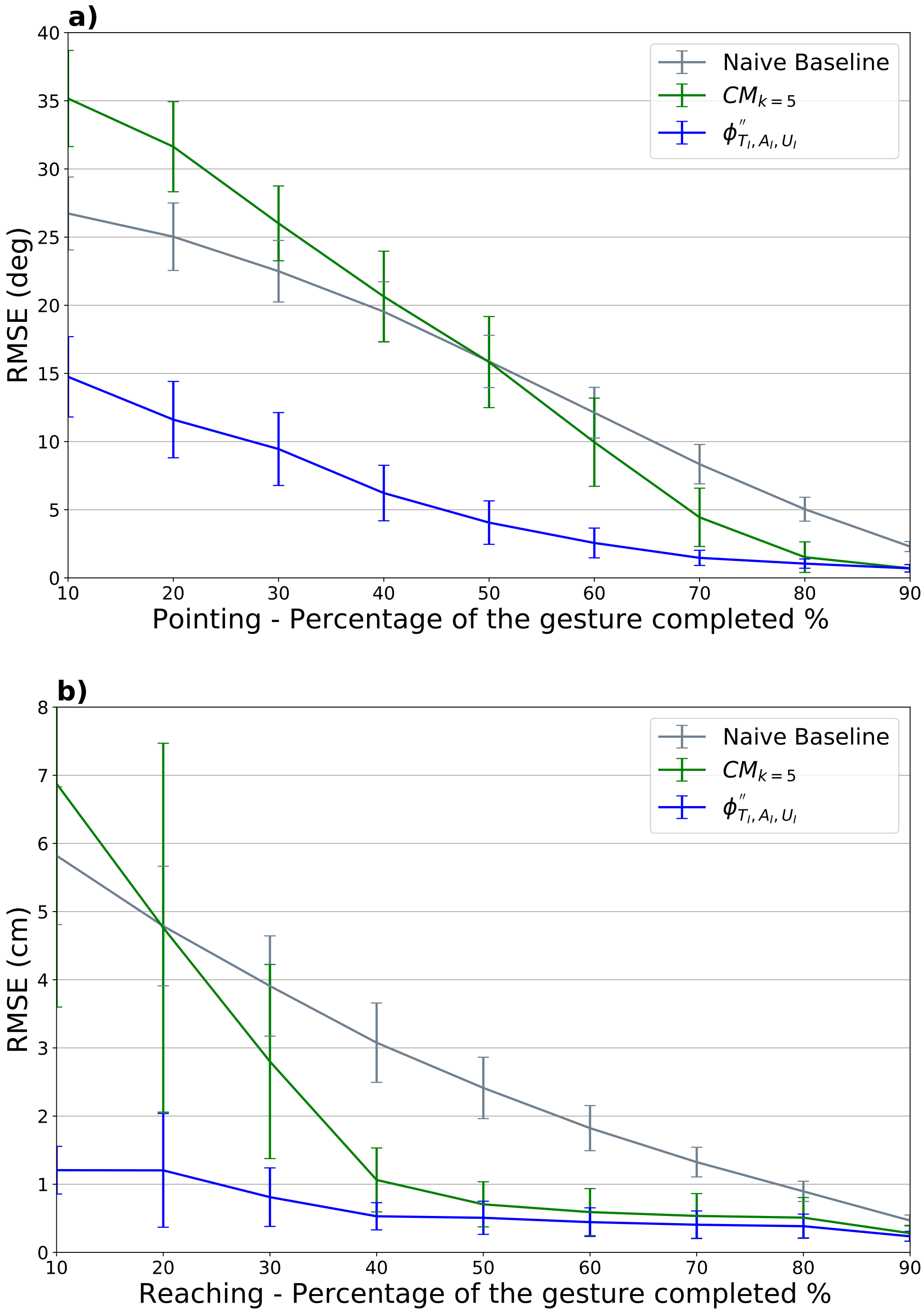


FIGURE 3.6: a) Average angular error for the pointing task at 10% increments along the way from the start for Φ''_{T_I,U_I,A_I}; b) Average distance error of the model Φ''_{T_I,U_I,A_I} for reaching task compared to the distance from the target as the baseline;

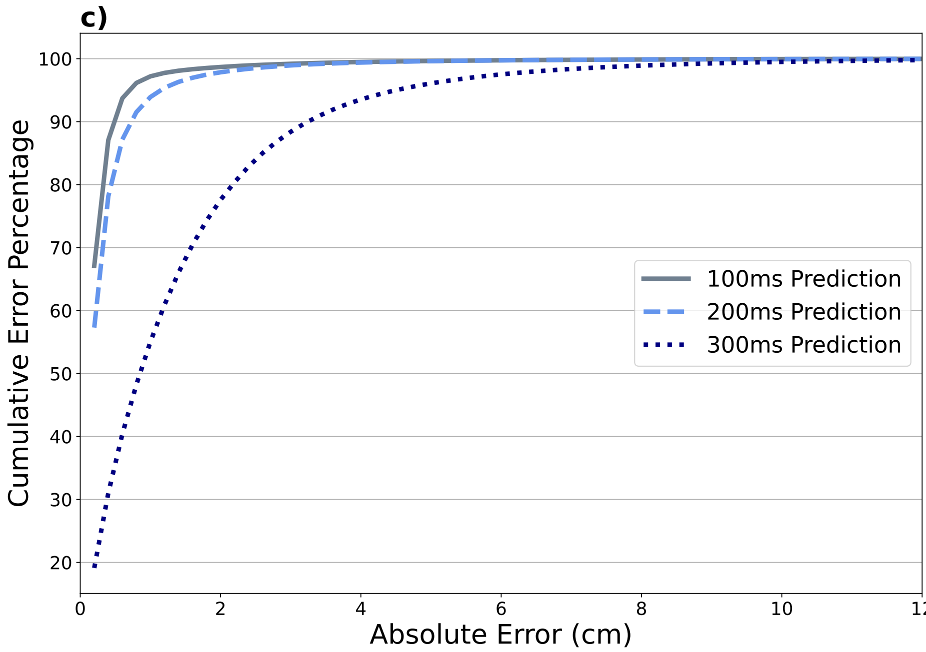


FIGURE 3.6: c) Cumulative error distribution for the Φ''_{T_I,U_I,A_I} model for different PIs.

We further studied how our models perform for reaching tasks, where the activity involves radial outward movements shown in Figure 3.3a. In this task, we report prediction accuracy as distance to target error at 10% increments along the way from the start. Figure 3.6b shows the average distance error of the model Φ''_{T_I,U_I,A_I} compared to distance from the target as the baseline. Φ''_{T_I,U_I,A_I} model achieves an average accuracy of prediction $0.51cm$ $(SD = 0.24cm)$ at 50% of the way and $0.41cm$ $(SD = 0.20cm)$ at 70% of the way of the movement. The average accuracy of the model reaches the pointing target's diameter $5cm$ at 47% of the way. Compared to baseline and the $CM_{k=5}$, Φ''_{T_I,U_I,A_I} model achieves a reduction of error by 78.9% and 28.0% at 50% of the way. Similarly, for 70% of the way, the error reduction of our model is 69.3% and 24.1%.

We also calculated the side effects introduced in [119] for our final model Φ''_{T_I,U_I,A_I}. For Lateness, Over-Anticipation and Wrong Orientation we report $1.21\,\mathrm{cm}$, $1.41\,\mathrm{cm}$, 27.63° for our model and $3.03\,\mathrm{cm}$, $2.77\,\mathrm{cm}$, 39.75° for 5th order classical model ($CM_{k=5}$). In comparison, our model reduces these side effects by 60%, 49% and 30% respectively compared to the 5th order classical model.

Furthermore, Figure 3.1b–c shows a comparison between the predicted trajectory by our model (red) and the real motion trajectory (blue) for activities *BeatSaber* and *FitXR-Box* for PI $[10, 300]$ ms, showing how the predicted trajectory closely follow the real one.

Since this model is axes independent, it is resilient for changes in the coordinate system given that it is with respect to the body. Also, $\alpha_n''(t)$ is a single axis function since all axes share the same coefficients. Therefore, $\alpha_n''(t)$ for displacement vector calculation using the Equation 3.2 can be expressed as:

$$\begin{bmatrix}\alpha_1(t)\\ \alpha_2(t)\\ \alpha_3(t)\\ \alpha_4(t)\\ \alpha_5(t)\end{bmatrix} = \begin{bmatrix}\begin{cases}1.0000 + 0.1693t & t < 0.16s\\ 1.0174 + 0.4547t - 2.4655t^2 & t \geq 0.16s\end{cases}\\ \begin{cases}0.5000 + 0.1837t & t < 0.16s\\ 0.6550 - 0.7458t - 0.2458t^2 & t \geq 0.16s\end{cases}\\ \begin{cases}0.1667 + 0.1151t & t < 0.16s\\ 0.2637 - 0.5122t + 0.1308t^2 & t \geq 0.16s\end{cases}\\ \begin{cases}0.0417 - 0.0343t & t < 0.16s\\ 0.0739 - 0.2809t + 0.2836t^2 & t \geq 0.16s\end{cases}\\ \begin{cases}0.0083 - 0.0064t & t < 0.16s\\ 0.0150 - 0.0569t + 0.0555t^2 & t \geq 0.16s\end{cases}\end{bmatrix} \tag{3.5}$$

3.5.6 Error Analysis

Figure 3.6c shows that our model often makes smaller errors and when large errors occur, they are less frequent. For instance, 90% of the errors that occurred are less than $0.6\,\mathrm{cm}$, $0.8\,\mathrm{cm}$ and $3.4\,\mathrm{cm}$ for $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$ across all users and activities. We considered the dominant hand of the participants for our analysis and the left-handed user was not given special treatment. Surprisingly, for the left-handed user, the model performed better than for the average of all users with RMSE of $2.34\,\mathrm{cm}$ of $300\,\mathrm{ms}$ prediction. We did not observe any impact on accuracy with the VR expertise of the participants. We also did not find a direct correlation between the movement kinematics and the prediction error. Anecdotally, we observed that large errors occur with large directional changes as shown in Figure 3.1e which need further investigation.

3.6 Verification of the Model

To assure the generalizability of our model, all the accuracy measures presented in section 3.5 is conducted on a test data set, which was not used to derive the model. For instance, of each task (pointing and games), less than 15% of the data is used for training, and the rest is used for testing. Especially with gradually progressing tasks such as VR games, it is fair to assume a large portion of the movements in the training data would differ from that of testing. However, our model uses a portion of data from each participant and each activity. Therefore, to investigate overfitting or selection biases, we conducted two further explorations: (1) *cross validation* of the methodology and (2) a new user study with *two new activities* to apply the model to a completely independent scenario.

3.6.1 Cross Validation

We conducted 4 folds of cross validation by separating 25% of the users as test data and calculated the average RMSE across all folds for all the activities. Figure 3.7a shows the results of the cross-fold evaluation compared to the collective model's RMSE. The average RMSE of cross validation was lower than that of the collective model at the higher end of the prediction, but we did not observe any significant differences. In addition, we built 20 models adding users 1 by 1 for training and tested them across the rest of the users progressively. After 13 users the model stabilizes with the change of error $0.99\,\mathrm{mm}$ for consecutive models. However, this may change with other user factors (e.g., age, injuries, etc.).

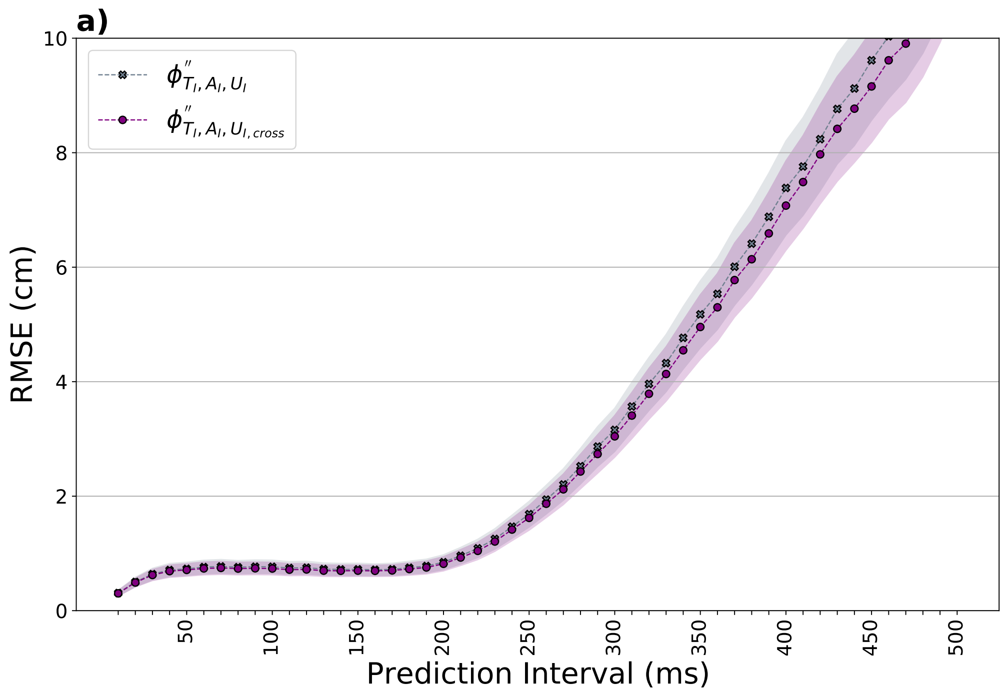


FIGURE 3.7: a) Average RMSE (in cm) of four-fold cross validation of the approach.

3.6.2 New Users and Activities

To verify the applicability of the model to a new user group and a new set of activities, we recruited 3 participants (age 22 to 30, one female), and asked them to perform two new tasks. *First task was performing free form sweeping movements* including flexion, extension, abduction and tracing a horizontal figure of 8 parallel to the body plane. We selected these movements since it covers a larger area of the space and they are dissimilar to the movements used in the first study. For the second task, we selected a dancing game *FitXR Dance Mode*, which also consists of movements dissimilar to that of the previous study. Participants performed each task for approximately 6 minutes and 3 minutes consecutively. Data collection and recording followed the same procedure as the first study. Figure 3.7b shows the average RMSE for the new tasks against the prediction time in comparison to the overall RMSE calculated for the original users. Results show a low average RMSEs of $0.55cm$ ($SD = 0.61cm$), $0.61cm$ ($SD = 0.59cm$) and $3.36cm$ ($SD = 1.37cm$) at prediction intervals $t = 100, 200, 300$ respectively. This shows even with new users and significantly different tasks, the model performs fairly well.

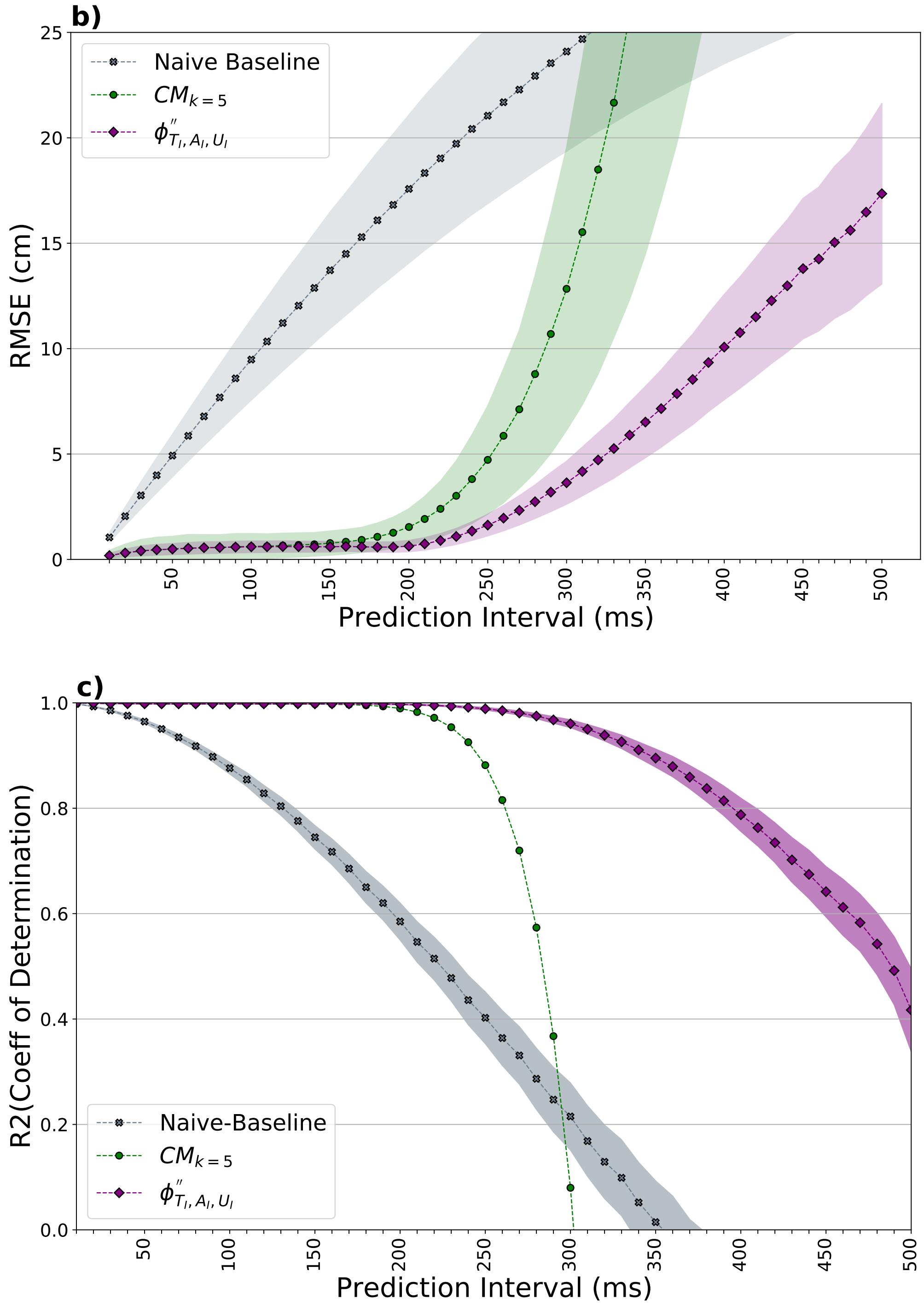


FIGURE 3.7: b) Average RMSE for three new users with two new tasks compared to the baseline. c) Average R^2 score for models Φ''_{T_I,U_I,A_I} compared to the baselines.

3.7 Discussion

In this paper, we presented a user- and activity-independent parametric kinematic model for 3D hand trajectory prediction in VR environments. Our results show that the model produces better performance compared to a non-naive baseline and needs little additional training for new users and activities. Furthermore, the simplicity of the model creates low computational overhead, which is an important factor for predictive systems. This section discusses the implications of the presented system and future perspectives.

Timely and high-quality multi-modal feedback is critical to implement realistic VR systems. Despite rapid progress in hardware developments, high-quality graphics and realistic physics simulations (e.g., interactions with fluids) are still very time-consuming in stand-alone VR systems. A predictive model can help the VR systems to forecast future events (e.g., Figure 3.1a collision) and pre-renders complex graphics in advance [138]. Offloading heavy computational tasks to remote clouds is another solution. However, offloading introduce communication delays ($40\,\mathrm{ms}$) and online computing delays ($100\,\mathrm{ms}$) [38]. A simple predictive model like ours can significantly contribute to counter these delays.

Another important area where a predictive model can be instrumental is to overcome asynchrony in multimodal feedback. For instance, delays as small as $50\,\mathrm{ms}$ in haptic feedback are noticeable to users in VR [37]. However, in commercial VR systems, delays in tracking ($22\,\mathrm{ms}$) [121], actuation of haptic systems ($33\,\mathrm{ms}$)[6] and other communication delays can easily exceed the required latency. Researchers also experiment with other types of sensory feedback such as thermal [131], wind [161, 136], smell and taste [137] where the onset delay in actuation is significant. These numerous delays lead to noticeable latency that breaks the immersion of the virtual environment [134]. A prediction model like ours can compensate for these delays by forecasting the future interactions of the users (Figure 3.1a) with minimum overhead.

[6] https://developer.oculus.com/documentation/native/pc/dg-input-touch-haptic/ (Accessed on 2021-03-26)

Predictive models have many other potential applications in VR beyond latency reduction. One such example is an error correction mechanism. Intermittent loss of tracking data is common in motion tracking systems due to occlusion or lighting issues. We observed such losses in the data we collected. Predictive data can be used to reconstruct missing hand trajectories as shown in Figure 3.1d, where circled areas demonstrate how missing path is redrawn using our model. Other areas where prediction can be used include haptic retargeting [28, 11] which enables the reuse of a single physical object to provide passive haptics for multiple virtual entities.

Our model performed surprisingly well without any further training for new users. We believe the success of the approach is due to commonalities of movement kinematics at short time intervals (e.g., $300\,\mathrm{ms}$). However, we further explored other possible factors that could degrade the performance of the model. The *generalizability* was an important concern and we evaluated our model against the dancing move set of CMU Motion Capture Dataset [40] to further explore the external validity. Our model achieved an RMSE of $3.86\,\mathrm{cm}$ for $300\,\mathrm{ms}$ prediction, which is an improvement of 83.6% compared to $CM_{k=5}$. It is important to notice that, unlike our dataset, this data includes hand motion data when the user is moving their feet. Another concern was the grounded and third-person perspective of the OptiTrack system we used and whether the model will apply to first-person wearable tracking systems used in most commercial VR headsets. We tested our prediction model for the Oculus Controller data we collected from the structured task (T1). This also gave us the opportunity to test if the model is completely independent of the tracking coordinate system and the sampling rate, where Oculus records data at $72\,\mathrm{Hz}$ in a coordinate system with respect to the participant's head. Furthermore, the marker we used for the hand was $3.5\,\mathrm{cm}$ above the wrist while Oculus tracks the controller held in hand, making the data truly of the hand location. With Oculus data, our model achieved a RMSE of $2.39\,\mathrm{cm}$ for $306\,\mathrm{ms}$ prediction time, which is a smaller error compared to OptiTrack prediction data at $300\,\mathrm{ms}$.

3.8 Limitations and Future Work

Our model was primarily trained and tested on aimed movements, which contained a majority of voluntary ballistic movements [97]. Further investigations are required on how the model perform for other types of movements (i.e., steering movements). However, the dancing task in the follow-up study (*FitXR-Dance*) was partly a steering task, where the participant copied the movements of a VR character simultaneously. For this task, error reduction is 76.6% at $300\,\mathrm{ms}$ with respect to $CM_{k=5}$ which is comparable to 78.1% error reduction in other tasks. We recommend our model is best used for predicting ballistic movements up to $340\,\mathrm{ms}$, as the model's R^2 score decreases below 0.9 (Figure 3.7c) beyond this prediction interval, indicating that the confidence of the model deteriorates beyond this interval.

All participants in our study were between 18 and 39 years old, which is currently the core demographics for VR applications[7]. The findings and movement models we discussed in this paper might differ for other age groups and people with injuries or disabilities.

While only the wrist trajectory is used in this study, we expect that the proposed approach can be adapted to motion trajectories for other body locations. For example, it would be possible to consider the motion trajectories for the elbow and shoulder to build a kinematics model of the whole arm. Moreover, our model does not take *hand orientation* and *wrist flexions* into account. Their possible effects on the prediction need to be explored in future work.

Although developed and evaluated within a 3D VR environment, our model is not fundamentally limited to predictions in VR applications. Since we use 3D motion trajectory, this work can be expanded to non-VR motion prediction such as hand movements in the real world or daily activities. It would be important to investigate if the generalized model changes for non-VR activities.

[7] See https://www.nielsen.com/us/en/insights/report/2017/us-games-360-report-2017/ (Accessed on 2021-03-26).

3.9 Conclusion

This paper contributed a novel user- and activity-independent *hybrid classical-regressive kinematics model* for continuous 3D hand trajectory prediction for ballistic movements in VR. Through a user study with 20 participants, we show our model performs comparably to personalized and specialized models for both structured and unstructured ballistic hand motions. Across all the users and activities, our model achieves a low Root Mean Square Error (RMSE) of $0.80\,\mathrm{cm}$, $0.85\,\mathrm{cm}$ and $3.15\,\mathrm{cm}$ for future hand positions of $100\,\mathrm{ms}$, $200\,\mathrm{ms}$ and $300\,\mathrm{ms}$. Finally, we evaluate our model through cross-validation and a follow-up study with new participants and activities. To the best of our knowledge, this is the first attempt to develop a generalized hand motion prediction model across different users and activities for ballistic movements. Our prediction model can be used in VR to pre-render graphics, calculate complex physics, support real-time multi-modal feedback, and even recover short-term tracking errors. While this paper focuses on VR, we believe there are benefits in extending our work to other domains in the future.

CHAPTER 4

Empirically Modified Minimum Jerk Models for 3D Trajectory Prediction for Reaching Movements

This chapter presents the paper titled "Empirically Modified Minimum Jerk Models for 3D Trajectory Prediction for Reaching Movements", which has been submitted for review at *Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH Asia)* journal track. The paper is included in this thesis without any modifications, in accordance with the University of Sydney thesis submission policy.

Reaching movements play a central role in human interaction with the environment, and reliably modeling them is vital for a range of applications, including predicting human hand motion in interactive systems for pre-rendering realistic computer graphics, synthesizing hand trajectories for human avatars and robotic platforms, and enabling safe and efficient human-robot interactions. Delivering these capabilities in practice requires accurate and lightweight models that can operate in real time on energy-constrained mobile and wearable devices, enabling low-latency interaction with minimal battery impact and supporting broad deployment across diverse interactive applications. In this work, we present novel user- and gesture-specific empirically modified minimum jerk models, which combines mathematical modelling with empirical observations for predicting and synthesizing 3D hand trajectories for reaching movements. These extremely lightweight models are designed to account for individual and gesture-level variability, and to generate naturalistic motion suitable for interactive and graphics-driven environments. We develop and evaluate our models using a dataset of 20 users performing reaching movements and further validate their generalizability on a public dataset. Finally, we demonstrate the ability of our models to synthesize user-specific hand trajectories with natural variation across repetitions, even for unseen users.

4.1 Introduction

Reaching movements are fundamental to human interaction with the physical world, playing a central role in various tasks, from everyday activities to complex manipulative actions. Accurately predicting and synthesizing these movements is crucial for many interactive applications, including computer graphics, Human-Robot Interaction (HRI), and Human-Computer Interaction (HCI) systems. In computer graphics, modeling these hand movements enables virtual characters to perform realistic and physically plausible actions, supporting applications such as hand motion synthesis for human avatars [180] and the pre-rendering of complex, anticipatory animations for interactive environments [73]. In collaborative human-robot environments, a reliable model of human hand motion is vital for enabling robots to anticipate human actions and adapt their behavior accordingly, playing a critical role in ensuring safe and efficient interactions between humans and robots [54, 19, 190]. This

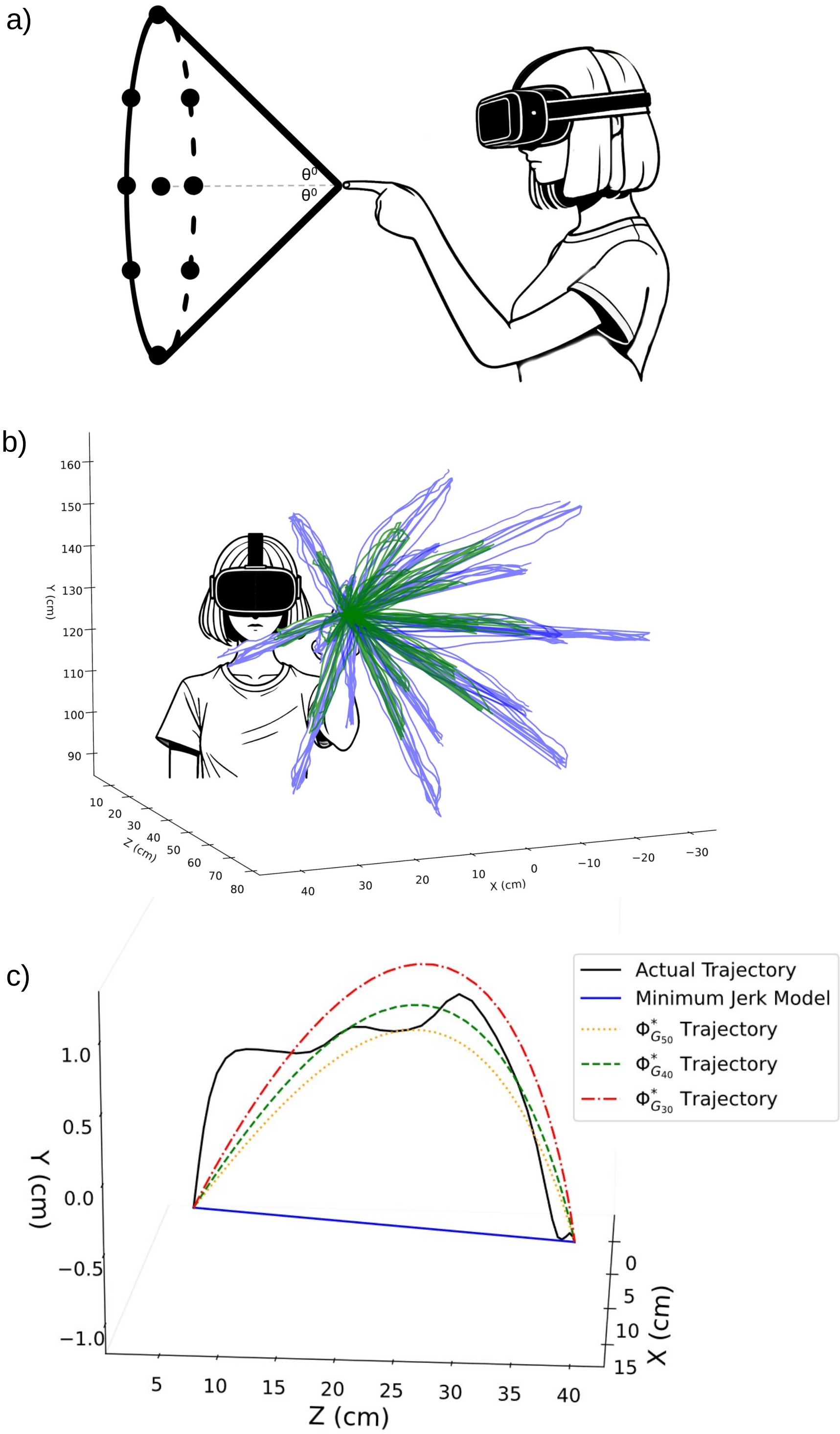


FIGURE 4.1: a) User study setup for reaching movements in the dataset, virtual targets are placed at varying distances of 20cm and 40 cm and angular deviation of 30 ° and 45 °, b) Reaching hand trajectories recorded for a specific user, highlighting the variance in gesture execution. Trajectories for the distances of 20 cm and 40 cm are shown in blue and green, respectively, c) Predicted trajectories for a reaching movement, showing that our developed gesture-specific models outperform the traditional Minimum Jerk Model.

is equally important in HCI contexts, including rehabilitation applications [7], and reach trajectory planning in Virtual Reality (VR) [27].

Given the significance of reaching gestures in many domains, substantial research efforts have focused on modeling and predicting their trajectories. Recent advancements have introduced deep learning [27, 148] approaches to capture the complexity of human motion. While these models have demonstrated relatively high accuracy in trajectory prediction, they are often complex, computationally expensive, and challenging to deploy on devices with limited resources. In contrast, purely mathematical models like the Minimum Jerk Model (MJM) [42] offer advantages in simplicity, predictability, and interpretability, making them attractive for real-time applications. The MJM has been widely used for decades to predict human reaching movements due to its ability to generate smooth trajectories that align with human motor behavior. However, it has been shown that the MJM is not sufficiently accurate when compared with empirical data, exhibiting significant errors between the predicted trajectories and actual human motion [54, 172]. This limitation highlights the need for enhanced mathematical models that can more precisely represent human reaching movements while preserving computational efficiency. Such models would enable real-time execution on modest processors with the limited energy budgets of smartphones, smartwatches, and headsets, offering low-latency responses and minimal impact on battery life. This would broaden deployment across various mobile and wearable platforms, supporting truly ubiquitous HRI and HCI applications.

Previous research has focused on refining the Minimum Jerk Model (MJM) to reduce its errors and improve its alignment with observed motion data[54, 154, 190, 160]. In a different approach, Wiegner et al.[172] proposed a higher-order Minimum Snap Model (MSM), which demonstrated a better fit to experimental data compared to the MJM. Additionally, more alternative models have been proposed [117, 65]. Despite these advancements, existing models still struggle to account for user-specific differences and the varying characteristics of individual gestures.

In this paper, we analyze over 2000 reaching gestures performed by 20 users in VR, as captured in the study by Gamage et al. [48]. We first compare these gestures with the MJM and MSM models, finding that although the MJM outperformed the MSM, it still exhibited notable errors. Furthermore, we demonstrate that users display distinctive characteristics in their reaching gestures. To capture these individual variations, we develop a user-specific empirically modified Minimum Jerk Model. Additionally, to address gesture-specific variations, we introduce a gesture-specific version of the empirically modified Minimum Jerk Model, which can effectively model the trajectory without requiring user-specific information once 20% of the gesture is completed. Building upon the MJM, we develop our models by framing them as an optimization problem with enhanced constraints to improve predictive accuracy. Our unique approach incorporates the total percentage of distance covered within a given percentage of time, derived from empirical data, directly into the optimization parameters. This method effectively captures the features of the trajectories and is particularly advantageous for vision-based hand-tracking devices such as VR headsets. It avoids the need for constraints based on velocity or acceleration, which depend on derivatives of position information. The proposed models offer simplicity, interpretability, and computational efficiency due to their purely mathematical formulation, making them more desirable compared to machine learning-based approaches. Finally, we show how our technique can synthesize user-specific hand trajectories, yielding multiple distinct realizations of the same reach so that even repetitive movements exhibit slight differences. By adjusting model parameters, we further construct plausible movement patterns for entirely novel user profiles.

Our findings show that the user-specific empirically modified model achieves an average improvement of 83.8% in approximating user-specific trajectories compared to the MJM and reduces the Root Mean Square Error (RMSE) by 16.9% across all users and gestures. The gesture-specific empirically modified models reduce the average RMSE across all users and gestures by 62.7% and 78.2% when predicting the trajectory after 30% and 50% of the gesture is completed, respectively, compared to the MJM. Additionally, we validated the models using another publicly available dataset for reaching gestures, demonstrating similar performance and confirming their generalizability. Figure 4.2a and b show simplified diagrams of our methodology for prediction and synthesis, respectively.

In summary, this paper makes the following main contributions:

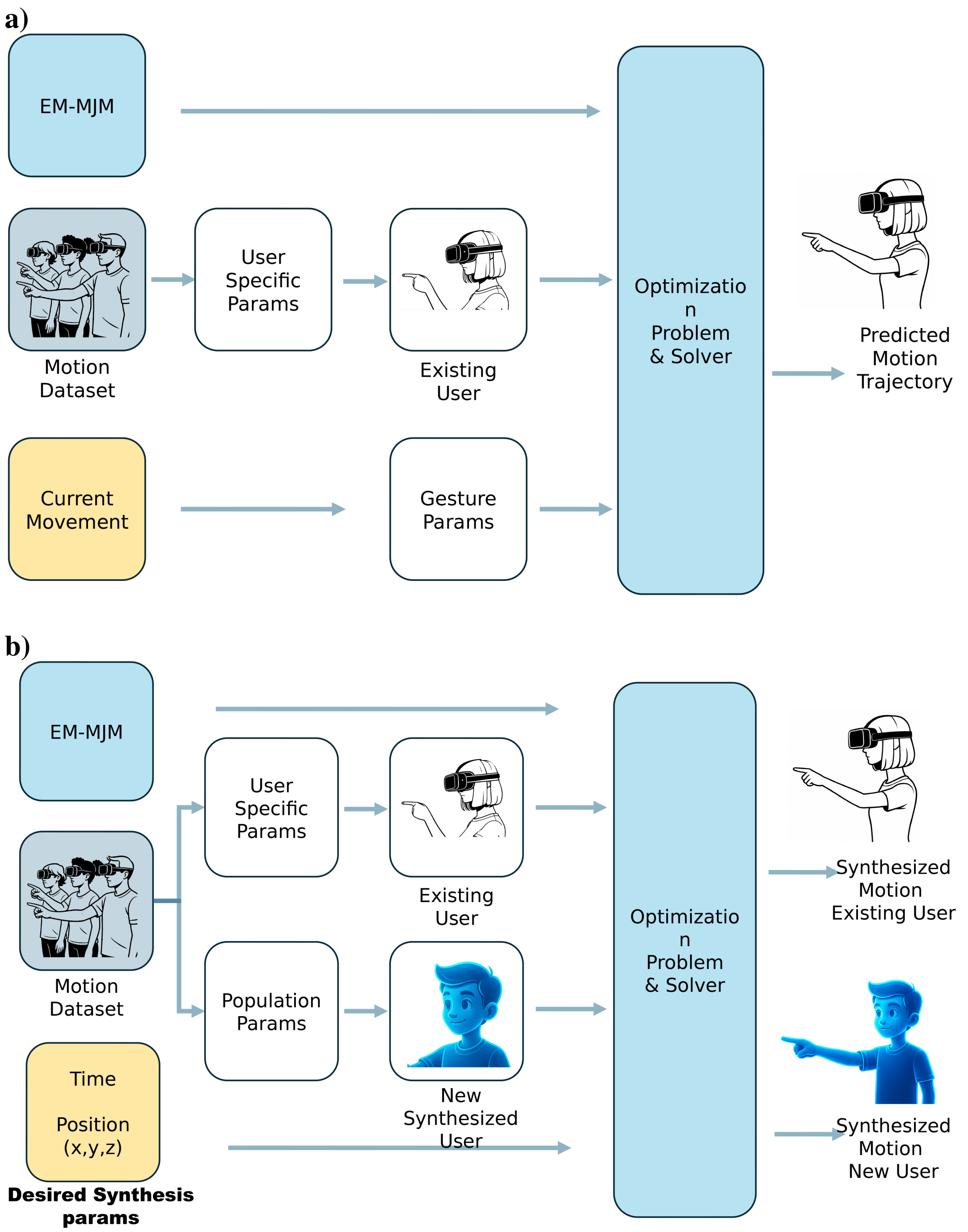


FIGURE 4.2: Simplified pipeline of the empirically modified Minimum Jerk Model (MJM) for a) prediction and b) synthesis of hand reaching movements.

(1) A comparison of MJM and MSM models for reaching gestures in VR, demonstrating that while MJM performs better, it still exhibits significant errors.

(2) A method for generating mathematical models for 3D trajectories of human reaching movements, consisting of:

   (a) A user-specific empirically modified Minimum Jerk Model to capture individual variations in reaching movements.
   (b) A gesture-specific empirically modified Minimum Jerk Model to address variations in specific gesture trajectories.

(3) A quantitative evaluation of the proposed models, including validation with a third-party dataset.

(4) A synthesis method that produces user-specific hand trajectories with natural variation across repetitions for existing and unseen users.

4.2 Related Work

This section reviews prior work on predicting human and robotic hand motion, with a particular focus on mathematical models for human motion prediction.

4.2.1 Mathematical Modelling of Hand Motion

Mathematical models based on principles such as control theory, polynomial trajectories, and splines have been widely used for modelling hand motion for decades due to their simplicity and efficiency. One of the earliest predictive models of human movement is Fitts' Law [41], which states that the time required to move to a target is a function of the ratio between the distance to the target and its size in one-dimensional (1D) tasks. Meyer et al. [111] further refined this by demonstrating that movements can be divided into a fast initial phase followed by slower corrective submovements, providing a deeper understanding of the speed-accuracy trade-off. Subsequent work expanded Fitts' Law beyond 1D tasks to broadening its applicability to many tasks and environments [101, 1, 33, 58]. Another similar model is the two-thirds power law by Lacquaniti et al. [84] states that the velocity of human hand motion is proportional to the curvature of the path raised to the power of one-third, meaning that movements slow down on curved paths and speed up on straighter ones. While these studies provide valuable insights into human motion, they cannot be directly used to compute human motion trajectories.

Another branch of research was on investigating hand motion trajectories for point-to-point movements, assuming compliance with a global minimum-cost condition [70, 66]. In this context, Flash and Hogan introduced the Minimum Jerk Model (MJM) [42], which suggests that human motion is optimized by minimizing jerk, or the rate of change of acceleration. This model has played a key role and is extensively applied in different HCI [103, 16, 122] and HRI [168, 103, 6, 125, 112, 7] tasks. Specifically, the MJM is widely used for predicting human motion trajectories in constrained movements, such as point-to-point tasks. Bratt et al. [16] utilised the MJM to predict the intended target location in ball-catching tasks in virtual reality after the first half of the movement was completed.

Ohmura et al. [122] demonstrated that for pointing tasks, the final position of the trajectory could be predicted by detecting the first acceleration peak, which occurs within the first 21% of the entire gesture. Lank et al. [86] proposed an endpoint prediction technique for ballistic point-to-point movements, based on the minimum jerk model (MJM). They derived an equation for instantaneous speed over distance using the MJM and employed polynomial fitting to extrapolate the future endpoint. To improve accuracy, they used a pre-calculated correction coefficient, based on the percentage of the gesture completed. Furthermore, Gong et al. [54] showed that traditional MJM trajectories are not accurate enough to model real human arm-reaching movements. To address this, they introduced the Modified Minimum Jerk Model (MMJM), incorporating an error modification term based on a second-order Fourier series. Although this improved accuracy, the error modification term varied with different motions, making it difficult to develop a generalised model for all reaching movements. Todorov et al. [160] proposed a constrained MJM model by incorporating both MJM and 2/3 power law. They set the jerk along the normal to the path to zero, which generates velocity profiles similar to the expected by the 2/3 power law focusing on complex arm movements with curved paths.

However, subsequent studies have found the MJM to be less accurate in certain scenarios. Gong et al. [54] demonstrated that the classical MJM exhibits inaccuracies when compared to actual human reaching movements. One key observation they made was that motion trajectories differ between individuals due to their unique motion patterns. Additionally, Svinin et al. [154] observed that relaxing the acceleration constraints of the MJM resulted in a better fit to experimental data for dynamic tasks. Wiegner et al. [172] also evaluated the applicability of the MJM for single-joint, goal-directed fast movements. They found that the higher-order Minimum Snap Model (MSM) provided a better fit for the experimental data than the MJM, concluding that muscle and limb dynamics account for the MSM's success in modeling fast movements.

4.2.2 Human Hand Motion Prediction

Due to the non-linearity and the dynamic nature of hand movements, predicting it is a challenging task. It involves either trying to predict either the endpoint, or the complete trajectory of the hand.

End-point prediction is often studied in the context of point-to-point movements such as reaching in constrained environments. Prior work has explored various mathematical models, including the Minimum Jerk Model (MJM) [86, 16, 122, 16], statistical methods and classical machine learning. Recent studies have explored deep learning approaches, including Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) networks [27], to address this problem. However, deep learning methods are computationally complex and resource-intensive, which poses significant challenges for deployment in resource-constrained environments.

Hand trajectory prediction focuses on forecasting the entire trajectory of human motion. This task is typically more challenging, and a variety of approaches, including statistical methods [16, 67, 179] and deep learning techniques [148, 109, 98], are employed to address it. Recently, Gamage et al. [48] proposed a kinematics-based regressive model for continuously predicting ballistic 3D hand movements in VR activities. They proposed that arm motion trajectories can be represented as a polynomial function of time, incorporating up to the fifth derivative of position. Their model achieved high prediction accuracy, with errors of 0.6 cm and 3.4 cm for prediction intervals of 100 ms and 300 ms, respectively, across both structured and unstructured movements. However, prediction accuracy declines for prediction times exceeding 300 ms. We aim to predict hand trajectories for reaching movements with minimal overhead, building upon the Minimum Jerk Model (MJM) to generate trajectories that closely resemble real human movements than existing models.

4.2.3 Hand Trajectory Prediction for Robotics

Predicting hand trajectory is crucial in human-robot interactions, as it enables systems to accurately mimic human motion, anticipate actions, and coordinate movements with humans.

In robot motion planning, mathematical models like the Minimum Jerk Model (MJM) are often used to generate the hand trajectory, which is then applied in inverse kinematics models to solve for joint locations [174]. Prior work has particularly focused on human-like motion planning for robots during reaching movements. Xie et al. [174] used the MJM to generate the finger path that the robot needs to trace. While the MJM is widely employed, it is noted that it does not always accurately reflect real human hand motions [23]. For a more detailed and comprehensive review of human like motion generation in HRI, readers are referred to [62].

4.3 Methodology

In this study, we aim to develop a mathematical model to describe human hand motion trajectories for point-to-point movements, utilizing an existing dataset for model development and evaluation. This section first outlines the dataset and then describes the data preparation process which is used for the model development.

4.3.1 Dataset

We utilized a dataset from a previous study by Gamage et al. [48], which investigated three-dimensional hand movements using both structured and unstructured activities in VR for 20 participants. For this study, we focused on the structured 3D pointing task, captured using an Oculus Quest VR headset at 72 Hz.

The structured task employed a repeated-measures, within-subject design to collect comprehensive data on hand trajectories in all spatial dimensions. At the start of each trial, participants positioned their virtual index finger on a designated starting point visible through the VR headset. A target then appeared, prompting the participant to reach from the initial position to the target, with a color change indicating successful acquisition. A second target would then appear on the opposite side of the same circle, requiring the participant to move from the first target to the second. The trial concluded when the participant returned their finger to the initial starting point.

Targets were evenly distributed across four circles in front of the participant, varying in distance (20 cm and 40 cm) and angular deviation (30° and 45°) from the starting position as shown in Figure 4.1(a). This arrangement ensured a wide range of movement trajectories for analysis. Participants were instructed to execute all movements as quickly and accurately as possible.

4.3.2 Data Preprocessing

To prepare the continuous motion capture data for analysis, several preprocessing steps were performed to segment individual gestures, smooth the data, calculate velocities, adjust gesture boundaries, remove outliers, and normalize data. Since the data were recorded continuously, it was necessary to segment the dataset into individual point-to-point gestures corresponding to each movement task. Figure 4.1(b) visualizes all segmented trajectories for a single user across all performed reaching movements.

We applied a Savitzky-Golay filter to smooth the positional data and compute velocities similar to prior work [190, 13, 85]. This filter was chosen because it preserves key features of the original signal, such as peak heights and widths, while reducing noise through polynomial fitting. This is particularly useful when differentiating positional data to calculate velocities, as differentiation can amplify high-frequency noise. The filter ensures more accurate velocity calculations without distorting the underlying movement patterns. Using the computed velocities, we refined the segmentation process to clearly define the start and end points of each gesture. A challenge encountered was that minor involuntary hand movements and sensor noise could generate small velocity values even when the hand was stationary. To address this, we further segmented the data by ensuring a strictly increasing displacement magnitude, effectively isolating the meaningful movement phases from the noise. To improve data quality and ensure robust model fitting, we identified and removed outliers using both statistical and manual methods. Lastly, to facilitate comparison across gestures of varying durations, we normalized both time and displacement, allowing for consistent analysis across different gestures.

4.3.3 Minimum Jerk Model

To establish a theoretical baseline for our analysis, we first present the mathematical formulation of the standard Minimum Jerk Model(MJM) proposed by Flash and Hogan [42], which will be used as a reference in evaluating the arm motion trajectories in our dataset.

According to minimum jerk theory, human motion trajectories are planned to minimize the magnitude of the *jerk* of the hand trajectory. For a 3D movement, where $\mathbf{r} = (x, y, z)$ are the coordinates and $[0, T]$ is the time interval, minimum jerk condition can be expressed;

$$\arg\min_{\mathbf{r}} \int_0^T \left(\left(\frac{d^3x(t)}{dt^3} \right)^2 + \left(\frac{d^3y(t)}{dt^3} \right)^2 + \left(\frac{d^3z(t)}{dt^3} \right)^2 \right) dt \tag{4.1}$$

Mathematically, it can be shown that Equation 4.1 is satisfied for trajectories with constant crackle (*5th derivative of displacement*) or zero pop (*6th derivative of displacement*). Therefore, $x(t)$, $y(t)$ and $z(t)$ can be expressed as a quintic polynomial function of time. They share the same form and it can be expressed as;

$$r(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5 \tag{4.2}$$

For point-to-point reaching movements, minimum jerk trajectory can be derived by imposing the following boundary conditions at $t = 0$ and $t = T$:

$$r(0) = r_0, \quad \dot{r}(0) = 0, \quad \ddot{r}(0) = 0, \tag{4.3}$$

$$r(T) = r_f, \quad \dot{r}(T) = 0, \quad \ddot{r}(T) = 0 \tag{4.4}$$

Thus, the classical equation that describes the minimum jerk trajectory for point-to-point movements can be expressed as follows where $\tau = t/T$;

$$r(\tau) = r_0 + (r_f - r_0)\left(10\tau^3 - 15\tau^4 + 6\tau^5\right) \tag{4.5}$$

By replacing $\hat{r}(t) = r(t)/r(f)$ and $r_0 = 0$, this can be simplified to use normalized time and displacement values.

$$\hat{r}(\tau) = 10\tau^3 - 15\tau^4 + 6\tau^5 \tag{4.6}$$

4.3.4 Minimum Snap Model

The Minimum Snap Model assumes that human motion trajectories are planned to minimize the magnitude of the *snap* of the trajectory. Similar to above, the minimum snap condition can be expressed as;

$$\arg\min_{\mathbf{r}} \int_0^T \left(\left(\frac{d^4x(t)}{dt^4} \right)^2 + \left(\frac{d^4y(t)}{dt^4} \right)^2 + \left(\frac{d^4z(t)}{dt^4} \right)^2 \right) dt \tag{4.7}$$

Mathematically, a minimum snap trajectory can be expressed as a septic polynomial equation.

$$r(t) = a_0 + a_1t + a_2t^2 + a_3t^3 + a_4t^4 + a_5t^5 + a_6t^6 + a_7t^7 \tag{4.8}$$

MSM model can be derived by solving 4.8 with conditions in Equations 4.3, 4.4 and additional constraints where;

$$\ddddot{r}(0) = 0, \quad \ddddot{r}(T) = 0 \tag{4.9}$$

Therefore, the classical equation that describes the minimum snap trajectory for point-to-point movements can be expressed similarly to Equation 4.6 as follows;

$$\hat{r}(\tau) = 35\tau^4 - 84\tau^5 + 70\tau^6 - 20\tau^7 \tag{4.10}$$

While the MJM and MSM models are described using different polynomial expressions, the trajectories they generate exhibit similar characteristics, as shown in Figure 4.3. Both models notably cover 50% of the total distance by the halfway point, with the velocity reaching its peak at this same time. However, the MSM model's acceleration peak occurs after the MJM model's peak.

4.3.5 Empirical Trajectory Comparisons

We began by comparing the empirical trajectories from our dataset with the predictions of the MJM and MSM models. To evaluate the model's applicability to our data, we generated predicted trajectories using equations 4.6 and 4.10 and converted them for the same initial

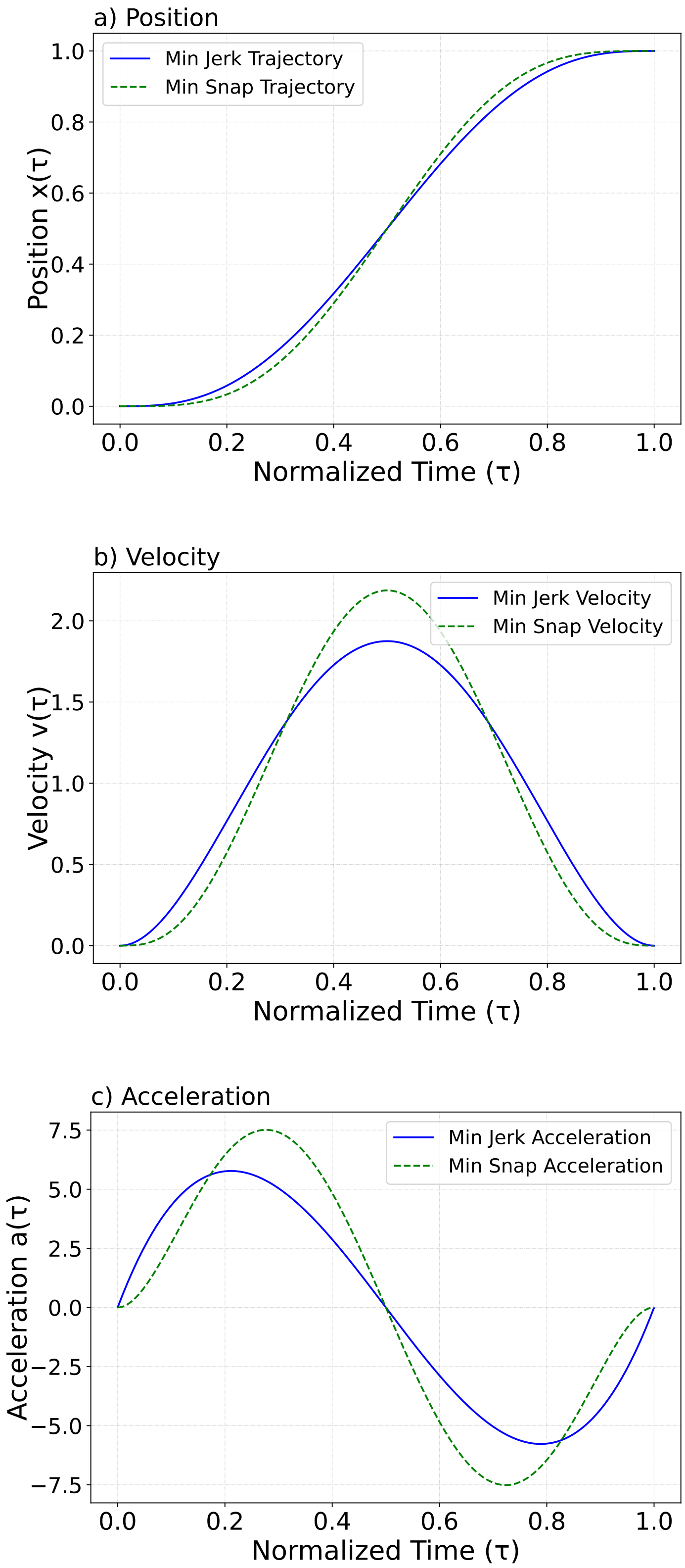


FIGURE 4.3: Comparison of Minimum Jerk Model and Minimum Snap Model trajectories showing a) displacement b) velocity c) acceleration profiles

and final positions and movement durations as in our recorded gestures. We quantified the differences between the model predictions and the empirical data by calculating the root mean square error (RMSE) for displacement profiles. As shown in Figure 4.4a, the MJM model performs better than the MSM model for reaching activities when considering all gestures by all users.

We also examined the errors at quarterly intervals along the trajectory (25%, 50%, and 75%) to better understand the performance of both models. As shown in Figure 4.4b, both the MJM and MSM models exhibit similar high errors at the 50% mark, with errors of 4.8 cm and 4.9 cm respectively. However, at the 25% and 75% points of the trajectory, the MSM shows larger errors compared to the MJM. At 25%, the MJM shows errors of 3.9 cm, while the MSM records 4.9 cm. Similarly, at 75%, the MJM has errors of 2.3 cm, whereas the MSM has a higher error of 3.1 cm. In addition, we also observe that maximum error occurs between 35% and 45% for both models.

Our analysis suggests that while the MJM performs better than the MSM in generating movement trajectories, it still shows considerable errors. Despite its improved accuracy, the MJM falls short in fully capturing the exact nature of the motion, leaving room for further improvement in trajectory modeling.

4.3.5.1 Inter-user Variations

We examined the RMSE values for each user, as shown in Figure 4.4c, which highlights the varying degrees of deviation from the MJM and MSM models across different participants. This variation is more pronounced when analyzing the average trajectories for each user, as shown in Figure 4.4d. The figure reveals distinct displacement profiles, indicating that while the overall task is the same, users tend to follow different movement paths.

We also calculated the distance traversed at quarterly intervals, 25%, 50%, and 75% of the total movement time, for each user and found significant differences between the models and the empirical data. This difference is particularly prominent at the 50% mark, where both the MJM and MSM models predict that the movement covers exactly 50% of the total distance at

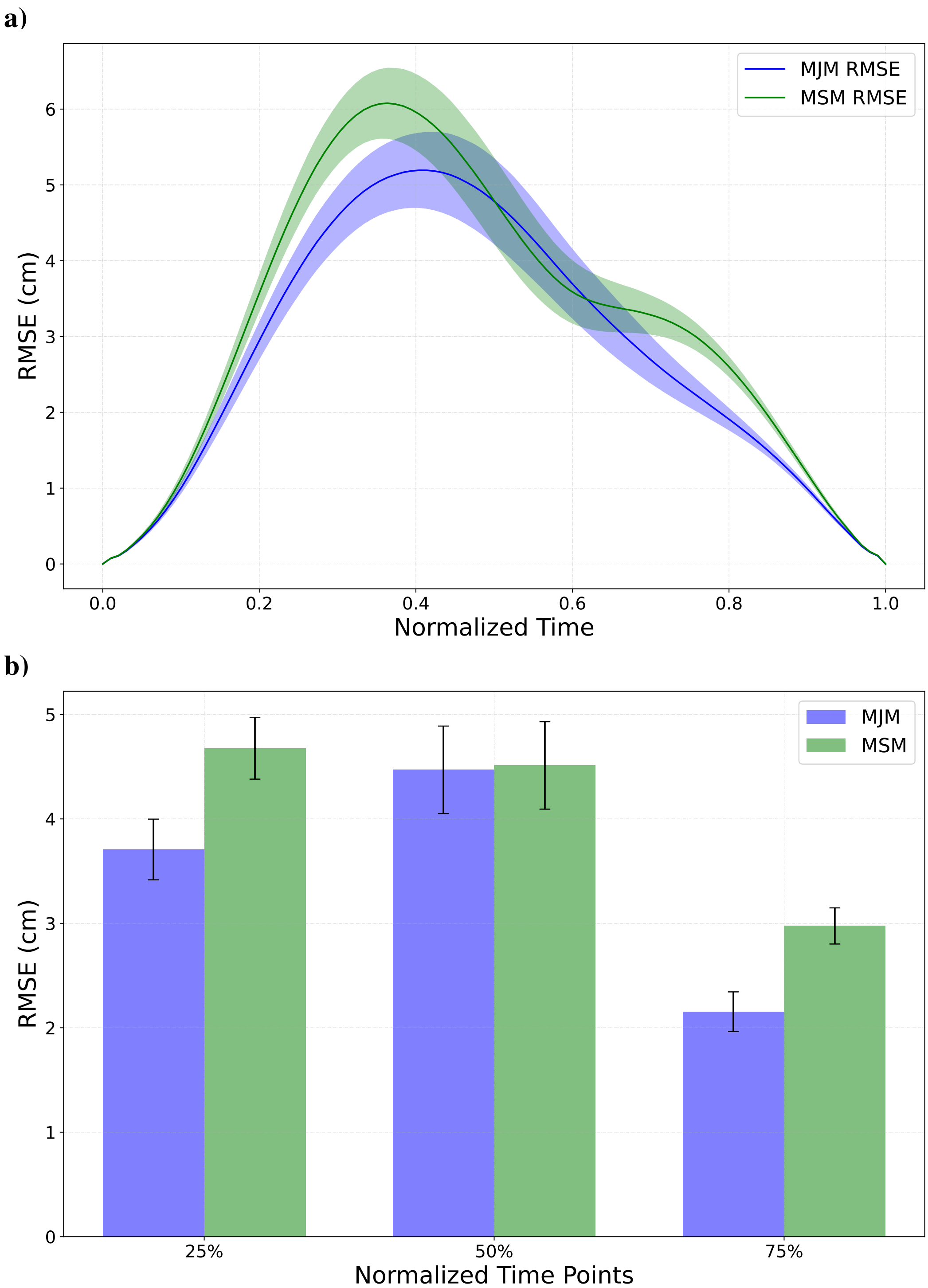
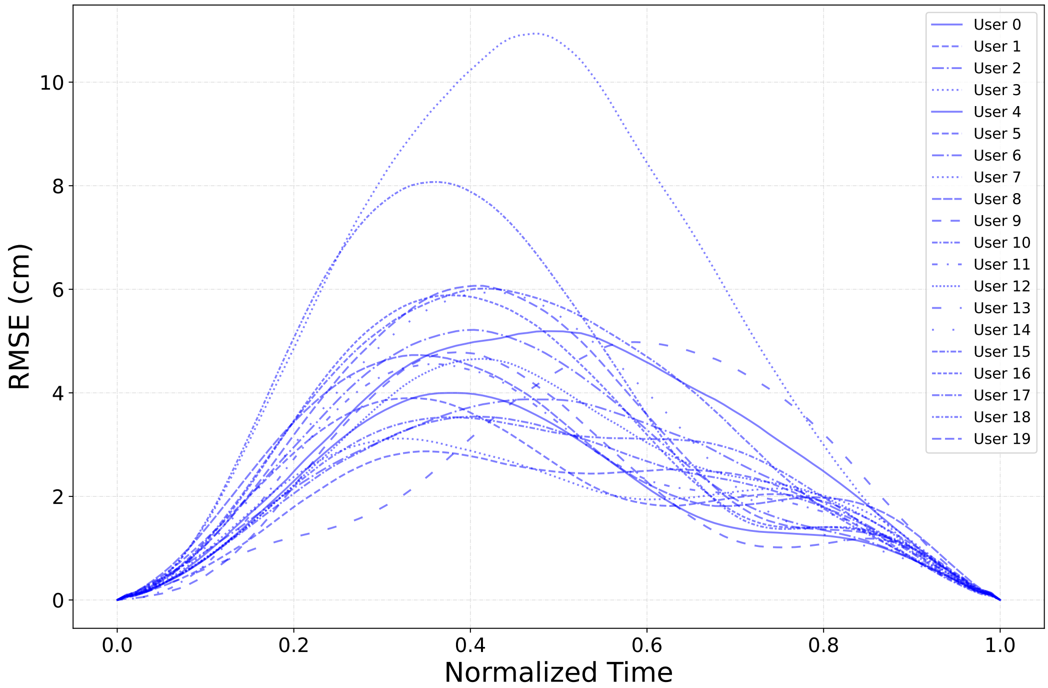


FIGURE 4.4: a) Comparison of Minimum Jerk Model (MJM) and Minimum Snap Model (MSM) trajectories showing displacement RMSE for normalized time. b) RMSE for 25%, 50%, and 75% of the MJM and MSM trajectories.

c)



d)

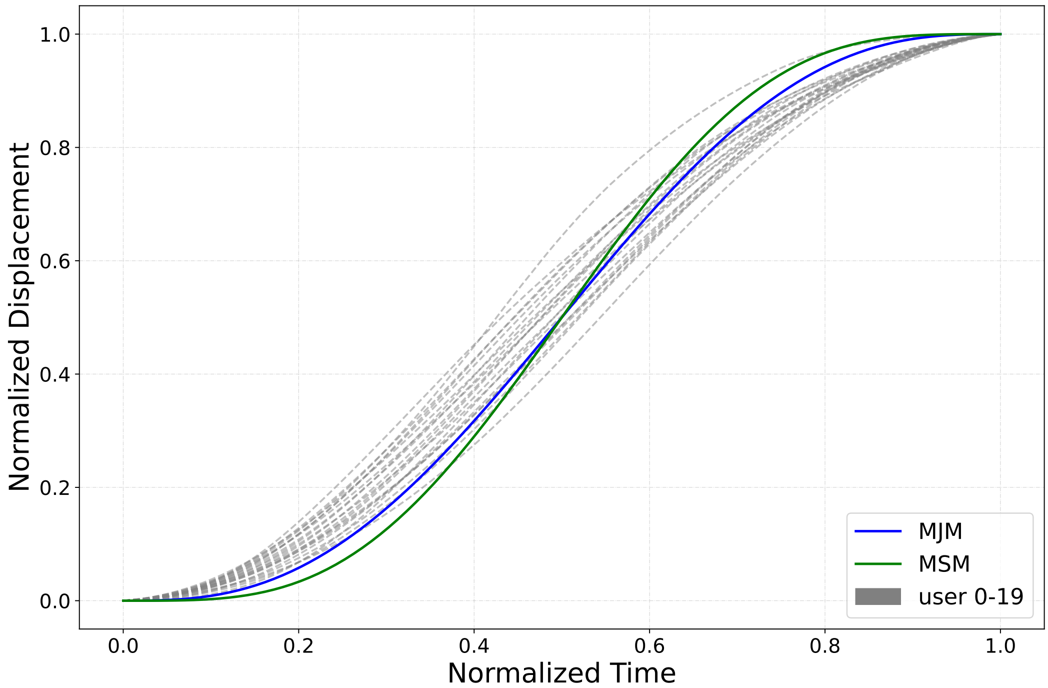


FIGURE 4.4: c) MJM trajectories showing displacement RMSE for each user for normalized time. d) Normalized averaged trajectories of each user showing along with MJM and MSM model trajectories.

half the total time due to their symmetric nature. However, our empirical data showed that this midpoint distance varied among users.

These user-specific variations suggest that the MJM and MSM models are unable to capture the uniqueness of individual movement patterns, as they assume uniformity in behavior. This indicates that a more individualized or adaptive model may be necessary to better account for these distinct user-specific movement characteristics.

4.3.5.2 Intra-user Variations

We observed that movement trajectories vary not only across different users but also for the same user when performing different gestures or even the same gesture repeatedly. This suggests that a user-specific model, while an improvement over the standard MJM, may still fall short in fully capturing the intricacies of individual movement patterns.

These variations may result from factors such as fatigue, gesture complexity, or slight shifts in user technique. Therefore, while a user-specific model may offer better predictions than the MJM, incorporating a more gesture-specific approach could yield even greater accuracy by accounting for the unique characteristics of each movement.

In the next two sections, we address the observed discrepancies between the empirical movement trajectories and the predictions of the MJM and MSM by proposing two new approaches. One approach focuses on capturing the unique variations of each individual user, offering a user-specific model for more accurate trajectory generation. The other approach focuses on accounting for variations across different gestures, aiming to predict the trajectory early in the process, specifically after the partial completion of a gesture.

4.4 User-specific Empirically Modified Minimum Jerk Model

We derive our models based on the MJM, as it performs better than the MSM for our empirical data. Accordingly, we express the displacement as a quintic polynomial, as shown in Equation 4.2. However, we adopt a similar approach to Svinin et al. [154], by relaxing the acceleration constraints. Similar to Svinin et al., we observe that the initial and final acceleration values for gestures in our empirical data tend to be non-zero as well. When only the initial conditions, $r(0) = r_0$ and $\dot{r}(0) = 0$ are applied to Equation 4.2, it can be written as;

$$r(t) = a_2t^2 + a_3t^3 + a_4t^4 + a_5t^5 \tag{4.11}$$

Similarly, $\hat{r}(\tau)$ can be written as follows, where b_2, b_3, b_4, b_5 are user-specific variables that need to be determined.

$$\hat{r}(\tau) = b_2\tau^2 + b_3\tau^3 + b_4\tau^4 + b_5\tau^5 \tag{4.12}$$

Our aim is to minimize the sum of squared deviations between the predicted displacements and the observed displacements at specific normalized time points τ_k such that $r(\tau_k) = d_{\tau_k}$ where $k = 1, 2, ..., N$ and N represents the number of chosen time points. d_{τ_k} is calculated from the empirical data where it is the mean of normalized displacement across all gestures for a specific user. We include constraints for the final displacement and velocity, as well as a constraint to prevent the trajectory from overshooting. Combining the objective function and constraints, the optimization problem is formulated as:

$$\begin{aligned} \underset{b_2,b_3,b_4,b_5}{\arg\min} \quad & \sum_{k=1}^{N} \left(r(\tau_k) - d_{\tau_k}\right)^2, \\ \text{subject to} \quad & r(1) = 1, \\ & \dot{r}(1) = 0, \\ & r(t) \leq 1, \quad \forall\, t \in [0, 1]. \end{aligned} \tag{4.13}$$

We solve the optimization problem using Python's *cvxpy*[1] library's default solver for all users to determine their specific b_2, b_3, b_4, b_5 coefficients. To provide the solver with some flexibility,

[1] https://www.cvxpy.org/

we slightly modify the overshoot constraint to $(t) \leq 1, \quad \forall\, t + \epsilon \in [0, 1]$ where ϵ is a small positive tolerance. We calculate d_{τ_k} based on the total displacement ($r(\tau)$) and used calculated coefficients on x, y and z axes to calculate the final trajectory. We call the resultant model the User-specific Empirical Minimum Jerk Model for Hand and denote it with the symbol Φ_U.

Note that for $k = 1$, we can try to relax only the final acceleration constraint, with the idea of making $b_2 = 0$ to transform Equation 4.12 to $\hat{r}(\tau) = b_3\tau^3 + b_4\tau^4 + b_5\tau^5$ which is similar to MJM equation as in Equation 4.4. However, this results in no valid solutions that satisfy the constraints for some trajectories. Removing the constraint $r(t) \leq 1, \quad \forall\, t \in [0, 1]$ does yield solutions, but these significantly deviate from the observed trajectories with large overshoots.

4.4.1 Model Selection

To select the optimal model that closely approximates each user's average normalized trajectory, we computed the Root Mean Square Error (RMSE) of the normalized trajectories. We explored models with the number of constraints $k = 1, 2, 3$. As illustrated in Figure 4.5 (for $k = 1, 2$), we compared the performance of the user-specific models Φ_U for different values of k and various time points τ. The results indicate that as the number of constraints k increases, the models generally achieve lower RMSE values, demonstrating improved accuracy in approximating the users' trajectories. For $k = 1$, the best-performing model was $\Phi_{U,\tau_{50}}$, which uses a single constraint at $\tau = 50\%$ of the total movement time. This model achieved a mean RMSE of 1.15 cm. Increasing k to 2 significantly reduced the RMSE values. The model $\Phi_{U,\tau_{40,70}}$, incorporating constraints at $\tau = 40\%$ and $\tau = 70\%$, achieved a mean RMSE of 0.66 cm. This substantial improvement indicates that adding an additional constraint allows the model to better align with users' movement patterns by capturing finer variations in the trajectory.

For $k = 3$, the models demonstrated marginal improvements over the $k = 2$ models. The model $\Phi_{U,\tau_{30,50,70}}$, which includes constraints at $\tau = 30\%, 60\%$, and 80%, yielded the lowest mean RMSE of 0.62 cm. However, the improvement compared to the $k = 2$ models is minimal, suggesting diminishing returns with further increasing k. In summary, while models

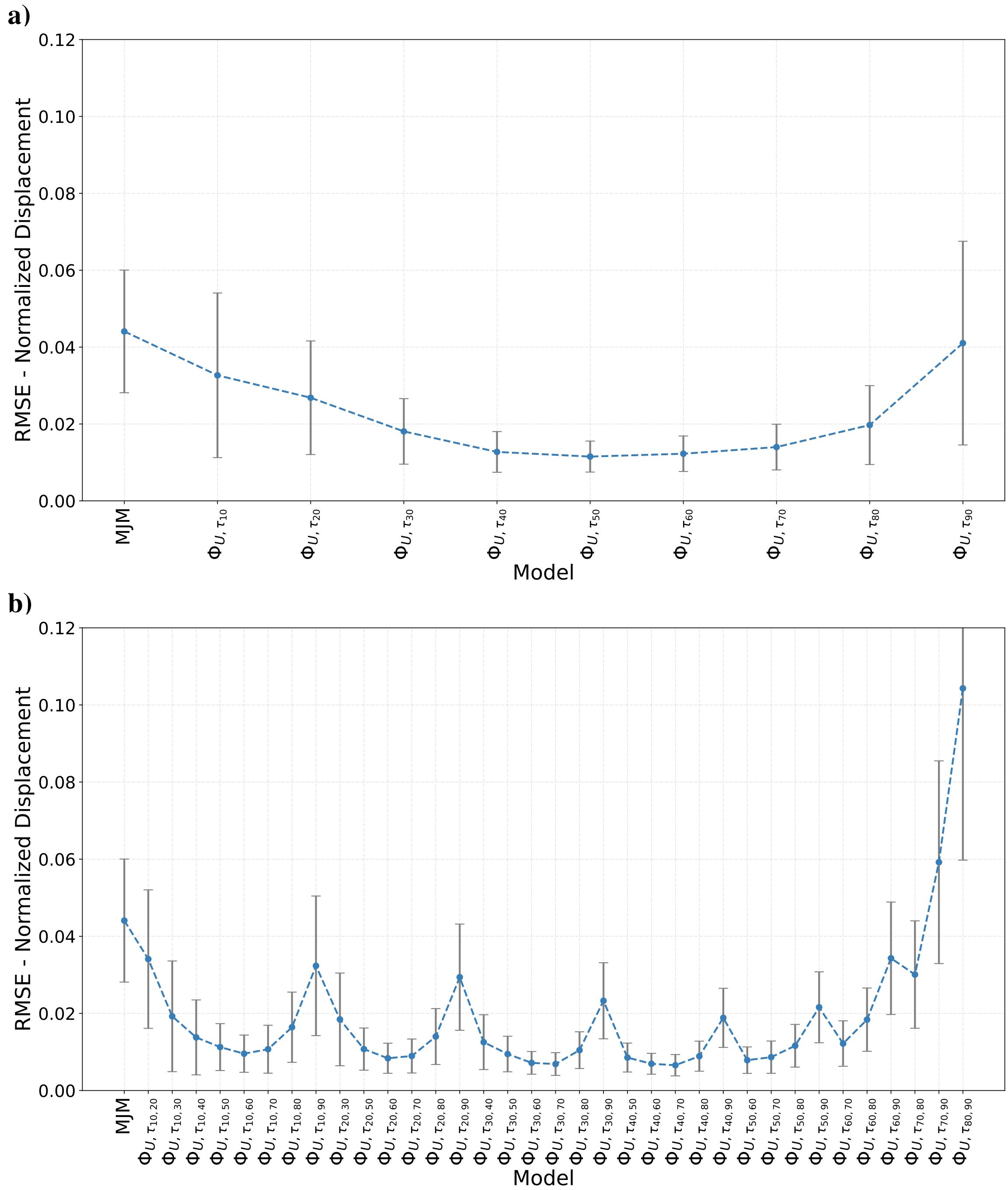


FIGURE 4.5: A comparison of RMSE for normalized trajectory for Φ_U models with different τ for a) $k = 1$ and b) $k = 2$ models. MJM is also shown for comparison.

with $k = 3$ constraints provided the most accurate approximations of user trajectories, the marginal improvement over the $k = 2$ models was minimal. Considering the diminishing returns, we propose to use $k = 2$. $\tau = 40\%$ and $\tau = 70\%$ for modelling user-specific trajectories and we call this model Φ_U^*.

4.5 Gesture-specific Empirically Modified Minimum Jerk Mode

While the Φ_U^* accounts for user-specific movement characteristics, it does not capture the variability in movement trajectories for the same user across different gestures or even the same gesture performed repeatedly, which can be influenced by factors such as fatigue, gesture length, complexity, etc. Φ_U^* predicts the trajectory a user would typically take on average, but the exact trajectory can only be determined as the gesture unfolds. Therefore, we aim to predict the trajectory of the gesture once a small percentage ($p\%$) of the gesture has been completed.

Here, we formulate the problem differently to Equation 4.13 that was for Φ_U^*. We explored different methods of formulating the optimization problem and empirically found a better approach that closely resembles the gesture trajectories observed in our data. Since we know the exact values of $x(\tau_k) = d_{\tau_k}$ for given τ, we impose these as constraints. Similar to the previous model, we relax the acceleration constraint and model the optimization problem as in Equation 4.14.

$$
\begin{aligned}
\underset{b_2,b_3,b_4,b_5}{\arg\min} \quad & \sum_{k=1}^{N} (r(\tau_k) - d_{\tau_k})^2 \\
\text{subject to} \quad & r(1) = 1, \\
& \dot{r}(1) = 0, \\
& r(\tau_p) = d(\tau_p), p = 1, 2, ..., M
\end{aligned}
\tag{4.14}
$$

In this approach, we enforce the displacement constraints $r(\tau_p) = d_{\tau_p}$ exactly, ensuring that the trajectory intersects the observed points at specified normalized times τ_p. The number of these observed points is denoted by M. By incorporating these constraints, we aim to capture the essential characteristics of user-performed gesture trajectories. We select d_{τ_k} We have the flexibility to select the value of d_{τ_l} in different ways. Since this represents a predicted value, it is determined as d_{τ_k} using the mean value obtained from empirical data. This value was calculated from the aggregated data across all users with the aim to make this model user independent.

Similar to the previous model, our optimization criterion is to minimize the deviations $r(\tau_k) - d_{\tau_k}$, with τ_k chosen based on empirical data. Here, we also remove the overshoot constraint, unlike in Equation 4.13, as some trajectories in the data may overshoot on certain occasions; this provides leniency and allows the model to better fit the empirical observations. By including both the displacement constraints and minimizing the deviations, we strive to capture the essential characteristics of the trajectory while permitting slight adjustments to fit the data more accurately. The final position and velocity constraints ensure the trajectory reaches the intended endpoint smoothly. In this model, the mean displacement values of the trajectories are used for d_{τ}, while d_{τ_k} values are computed by averaging aggregated data across all users. This approach aims to create a user-independent model. We denote this model as Φ_G.

This optimization problem is solved using Python's *cvxpy* library's default solver, employing quadratic programming techniques similar to the previous model. To achieve a solution with lower complexity, we aim to maintain a balance between performance and computational simplicity. Specifically, we wanted to keep $N = 1$ and $M = 1$, if feasible. Moreover, to enhance the responsiveness of the trajectory prediction, it is critical to minimize the value of τ_k. Since our objective is to anticipate the trajectory before reaching its midpoint, we restrict our exploration to the solution space where $\tau_k \leq 50\%$.

Figure 4.6a shows the average RMSE per user for the model Φ_G. The model's performance is limited due to several contributing factors. Although the trajectory was computed using the mean displacement, the x, y, and z axes exhibited distinct characteristics, indicating that applying the same coefficients across all axes may not effectively capture movement patterns. Additionally, the d_{τ_k} values were selected without considering the initial movement phase of the gesture, causing the latter part of the predicted trajectory to fail to reflect the characteristics of the initial phase.

To address these issues, we propose a refined model, denoted as Φ'_G. This model calculates d_{τ_k} separately for each axis, allowing individual coefficients to be derived for generating corresponding trajectories along the x, y, and z axes. Additionally, it employs an adjusted d'_{τ_k} value based on d_{τ_p} to improve alignment with the initial movement phase using the Equation 4.15. Note that $\mu_{d_{\tau_p}}$, $\sigma_{d_{\tau_p}}$ and $\sigma_{d_{\tau_k}}$ are calculated with aggregated data across all users.

$$\begin{aligned} \lambda &= (d_{\tau_p} - \mu_{d_{\tau_p}})/\sigma_{d_{\tau_p}}, \quad p = M \\ d'_{\tau_k} &= d_{\tau_k} + \lambda \sigma_{d_{\tau_k}}, \quad k = 1, 2, \ldots, N \end{aligned} \tag{4.15}$$

Finally, the model uses $M >= 1$ for enhanced prediction accuracy. We formulate the optimization problem for Φ'_G as in Equation 4.16. Note that we incorporate the d_{τ_p} constraints into the optimization formula for $M > 1$ to avoid over-constraining the problem

$$\begin{aligned} \underset{b_2,b_3,b_4,b_5}{\arg\min} \quad & \sum_{k=1}^{N} \left(r(\tau_k) - d'_{\tau_k}\right)^2 + \sum_{p=1}^{M-1} \left(r(\tau_p) - d_{\tau_p}\right)^2 \\ \text{subject to} \quad & r(1) = 1, \\ & \dot{r}(1) = 0, \\ & r(\tau_p) = d(\tau_p), \quad p = M \end{aligned} \tag{4.16}$$

Furthermore, we observed that on some occasions, users tend to overshoot the target location during the trajectory for a particular axis. In order to address this, we slightly modify the Equation, by adding an extra condition. Which is if $d(\tau_p)$ > 1 when $p = M$, we add an extra part to the optimization criteria which is $d(\tau_p)/dt = 0$ which ensures that it is the maximum overshoot. This tries to follow the user's behavior of trying to correct the trajectory in case of an overshoot.

Furthermore, we observed that users occasionally overshoot the target location along specific axes during the trajectory. To address this, we introduced a slight modification to the equation by adding an extra condition. If $d(\tau_p) > 1$ when $p = M$, we include an additional term in the optimization criterion:

$$\frac{d}{dt}d(\tau_p) = 0 \tag{4.17}$$

This condition ensures that the maximum overshoot is recognized, encouraging the model to adjust accordingly. By incorporating this term, the model aims to replicate user behavior, reflecting how users naturally correct their trajectories after an overshoot.

4.5.1 Model selection

Figure 4.6 shows the RMSE values for the Φ_G and Φ'_G models across different τ and p values. We observed that Φ'_G consistently achieved lower RMSE values than Φ_G for models with higher p values. For $p >= 30$, models $\Phi_{G,p_{30},\tau_{80}}$, $\Phi_{G,p_{40},\tau_{80}}$, and $\Phi_{G,p_{50},\tau_{80}}$ achieved RMSE values of 1.03 cm, 0.93 cm, and 1.10 cm, respectively. In contrast, the corresponding Φ'_G models outperformed these, with $\Phi'_{G,p_{30},\tau_{80}}$, $\Phi'_{G,p_{40},\tau_{80}}$, and $\Phi'_{G,p_{50},\tau_{80}}$ achieving RMSE values of 0.95 cm, 0.70 cm, and 0.56 cm, respectively. We denote the best-performing Φ'_G models for specific p values as $\Phi^*_{G_p}$. For example, the best-performing model for $p = 30$, which indicates that the trajectory is predicted after 30% of the gesture is completed, is denoted by $\Phi^*_{G_{30}}$.

These results indicate that $\Phi^*_{G_p}$ models can predict the trajectory more accurately when a higher percentage of the trajectory is completed. However, for early trajectory prediction, it is desirable for the p value to be smaller enabling earlier predictions.

4.6 Results

In this section, we evaluate the models Φ^*_U and Φ^*_G quantitatively. We show that our final user-specific empirically modified minimum jerk model (Φ^*_U) achieves an average improvement of 83.80% in approximating user-specific trajectories compared to the traditional MJM for all 20 users. Similarly, Φ^*_U reduces the RMSE by an average of 16.95% across all gestures.

We also demonstrate that our final gesture-specific empirically modified minimum jerk models (Φ^*_G) outperform MJM in predicting trajectories once the gesture has started. These models

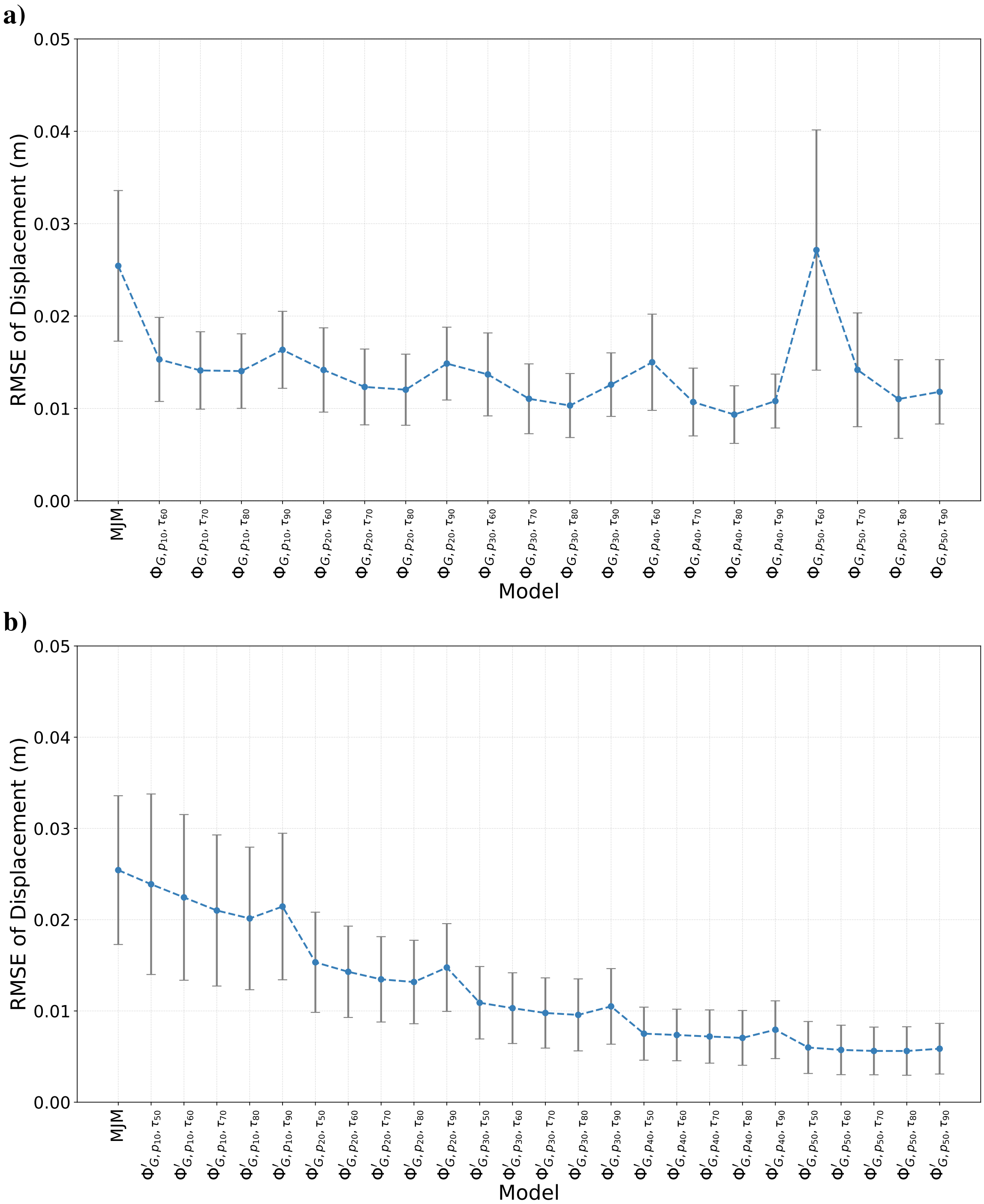


FIGURE 4.6: A comparison of RMSE for Minimum Jerk Model and a) Φ_G, b) Φ'_G models with different configurations.

reduce the average RMSE for all 20 users by 78.17%, 72.46%, and 62.70% for $\Phi^*_{G_{50}}$, $\Phi^*_{G_{40}}$, and $\Phi^*_{G_{30}}$, respectively.

Finally, to evaluate the generalizability and applicability of our models to new users, we perform quantitative testing with an external dataset. Results show that Φ_U^* improves the approximation of user-specific trajectories by 73.85% on average compared to the traditional MJM. It also reduces the RMSE by an average of 11.80% across all gestures. Additionally, the $\Phi_{G_{50}}^*$ model achieves a 77.29% average RMSE reduction across all users, demonstrating performance comparable to that observed with our initial dataset.

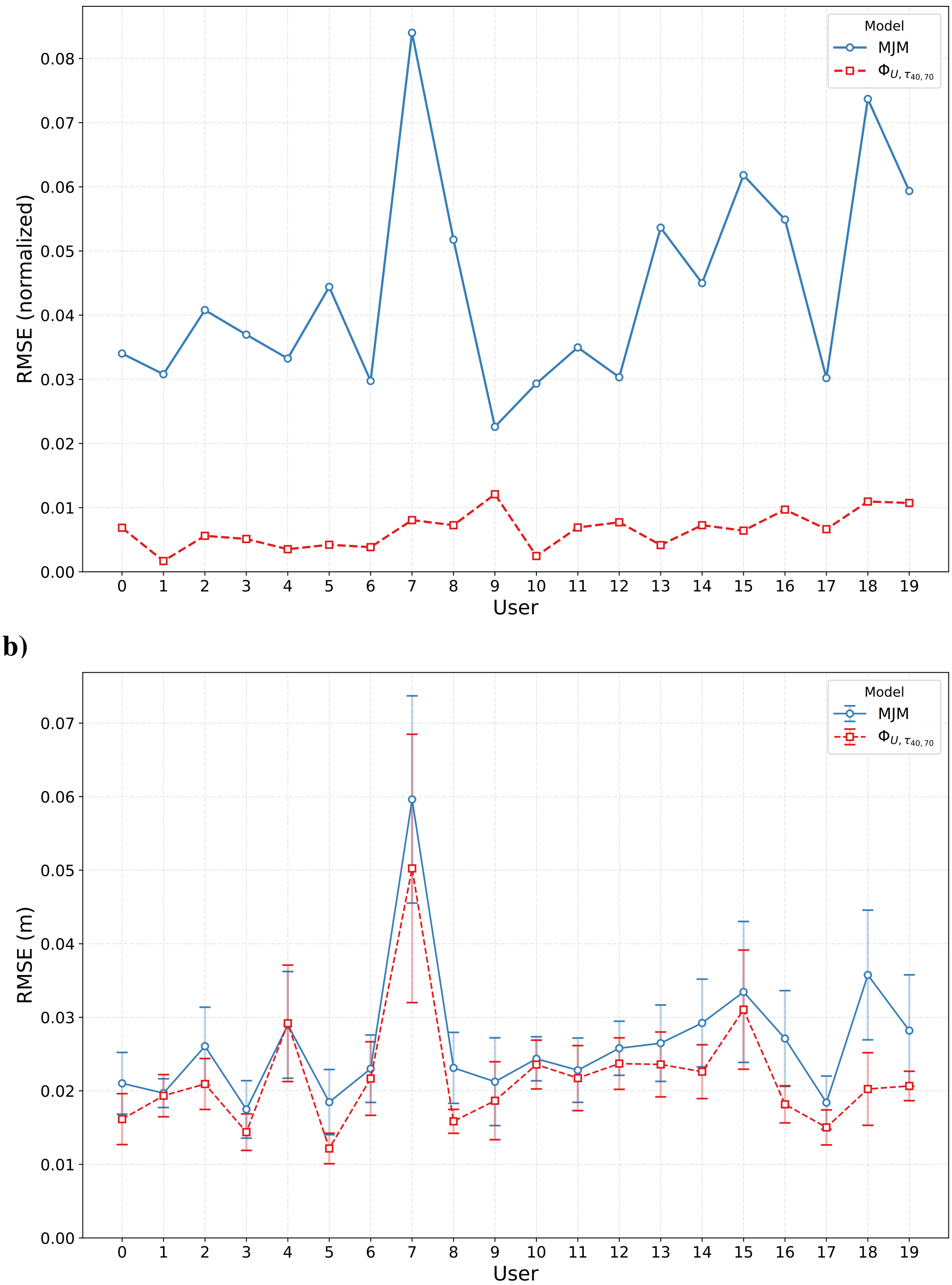
4.6.1 Φ_U^* Model Evaluation

Figure 4.7a illustrates how well MJM and $\Phi_{U,\tau_{40,70}}$ approximate the user's average trajectory. Across all users, it is evident that the $\Phi_{U,\tau_{40,70}}$ model consistently achieves better approximations compared to the standard MJM. The normalized error data shows substantial reductions in error when using the $\Phi_{U,\tau_{40,70}}$ model. For example, User 1 experienced one of the most significant reductions, with the RMSE decreasing from 3.40 cm to 0.16 cm, representing a reduction of approximately 94.53%. Conversely, the smallest improvement was observed for User 9, where the RMSE decreased from 2.26 cm to 1.20, a reduction of 46.51%. These results highlight the effectiveness of the $\Phi_{U,\tau_{40,70}}$ model over the traditional MJM for modelling users' approximate trajectories for reaching movements.

Figure 4.7b shows the RMSE values calculated for all gestures by each user with the models MJM and $\Phi_{U,\tau_{40,70}}$. The mean RMSE for each user generally decreases when employing the $\Phi_{U,\tau_{40,70}}$ model compared to the MJM. For example, User 18's mean RMSE decreased from 3.58 cm with the MJM to 2.02 cm with the $\Phi_{U,\tau_{40,70}}$ model, indicating a reduction of approximately 43.37%. This reduction represents the highest improvement observed among users. However, we see a slight increase in error for User 4. User 4 experienced an increase in mean RMSE, from 2.90 cm with the MJM to 2.92 m with the $\Phi_{U,\tau_{40,70}}$ model, reflecting a 0.75% increase in error. Furthermore, Users 6, 10, and 11 only show a slight improvement for the $\Phi_{U,\tau_{40,70}}$ model.

The substantial reductions in RMSE for several users suggest that incorporating additional constraints enhances the model's ability to closely match actual user trajectories. While

a)



b)

FIGURE 4.7: A comparison between MJM and Φ_U^* models. a) Error between each user's normalized trajectory and models. b) RMSE value calculated for all gestures by a user with the models.

this improvement is effective for approximating the user's average trajectories, it does not address the underlying issues of gesture consistency across different movements. The lower performance observed for Users 4, 6, 10, and 11 for model $\Phi_{U,\tau_{40,70}}$ indicates a high variance in how these users perform gestures, making it more challenging for the model to achieve precise predictions. Conversely, for users where the RMSE was significantly reduced, the lower variance in their gestures allowed the model to more accurately approximate their trajectories.

4.6.2 Φ_U^* Model User Coefficients

User	b_2	b_3	b_4	b_5
0	2.9126	-0.6564	-2.4251	1.1689
1	1.5330	2.9258	-5.4505	1.9918
2	3.3508	-2.0836	-0.8851	0.6180
3	2.9697	-1.9451	-0.0188	-0.0057
4	1.2531	3.7228	-6.2050	2.2291
5	3.9183	-4.7736	2.7925	-0.9371
6	1.9450	1.7377	-4.3105	1.6278
7	3.3890	0.4625	-6.0922	3.2406
8	4.3240	-5.9492	3.9264	-1.3012
9	2.1052	1.7784	-4.8725	1.9889
10	1.9178	1.8170	-4.3873	1.6526
11	2.5939	-0.9795	-0.8227	0.2083
12	2.8687	-1.0978	-1.4104	0.6396
13	1.0037	3.6484	-5.3078	1.6557
14	3.1470	-0.7026	-3.0358	1.5914
15	4.2294	-4.3106	0.9330	0.1482
16	4.2609	-4.9409	2.0991	-0.4191
17	2.4419	1.2884	-4.9026	2.1723
18	5.7940	-10.2429	8.1036	-2.6548
19	4.3063	-4.7369	1.5551	-0.1244

TABLE 4.1: Table showing b_2, b_3, b_4, and b_5 values for each user.

Table 4.1 presents the Φ_U^* coefficients for all 20 users and Table 4.2 provides a statistical overview of these coefficients. Implications and our observations of these coefficients are discussed further in Section 4.8.1.

Coefficient	Mean	Std	Min	Max
b_2	3.0132	1.2200	1.0037	5.7941
b_3	-1.2519	3.6670	-10.2429	3.7228
b_4	-1.5358	3.8869	-6.2050	8.1036
b_5	0.7745	1.4291	-2.6548	3.2406

TABLE 4.2: Mean, standard deviation, minimum, and maximum values for each coefficient.

4.6.3 Φ^*_G Model Evaluation

Figure 4.8a shows the average RMSE computed for all gestures by each user when applying the MJM and three variants of the Φ^*_G models. The mean RMSE across all users consistently decreases for the Φ^*_G models, with $\Phi^*_{G_{50}}$ being the most accurate. This indicates that the predicted trajectory becomes more accurate as a larger portion of the gesture is completed. The $\Phi^*_{G_{50}}$ model achieves a maximum RMSE reduction of 86.64% for User 15 and a minimum reduction of 69.17% for User 12, highlighting its effectiveness in improving trajectory prediction across different users.

Figure 4.8b shows the average RMSE computed for all users against the normalized time. For all Φ^*_G models, the error increases between the p value and the end of the gesture. This increase is smallest for the $\Phi_{G_{50}}$ model, while it is more pronounced for the $\Phi^*_{G_{30}}$ model. The results highlight that while the $\Phi^*_{G_{50}}$ model achieves relatively low prediction error, it has room to improve as errors gradually accumulate toward the latter part of the gesture.

Figure 4.9 shows a reaching gesture performed by a particular user (User 0) toward a target in the xz-plane, with no intended displacement in the vertical (y) direction. However, the user slightly overshoots in the y-axis. As depicted in Figure 4.9a, the MJM model fails to account for this deviation, while the Φ^*_G model better handles the overshoot. Figure 4.9b illustrates that the RMSE of the MJM model is significantly higher than that of the Φ_G models, with $\Phi_{G_{50}}$ providing the most accurate prediction among all evaluated models. Additionally, Figure 4.9c presents a 3D plot of the trajectory, clearly showing the overshoot in the vertical direction and Figure 4.1(c) shows the same gesture, but with non-uniform scales on axes to highlight the deviation in the vertical direction.

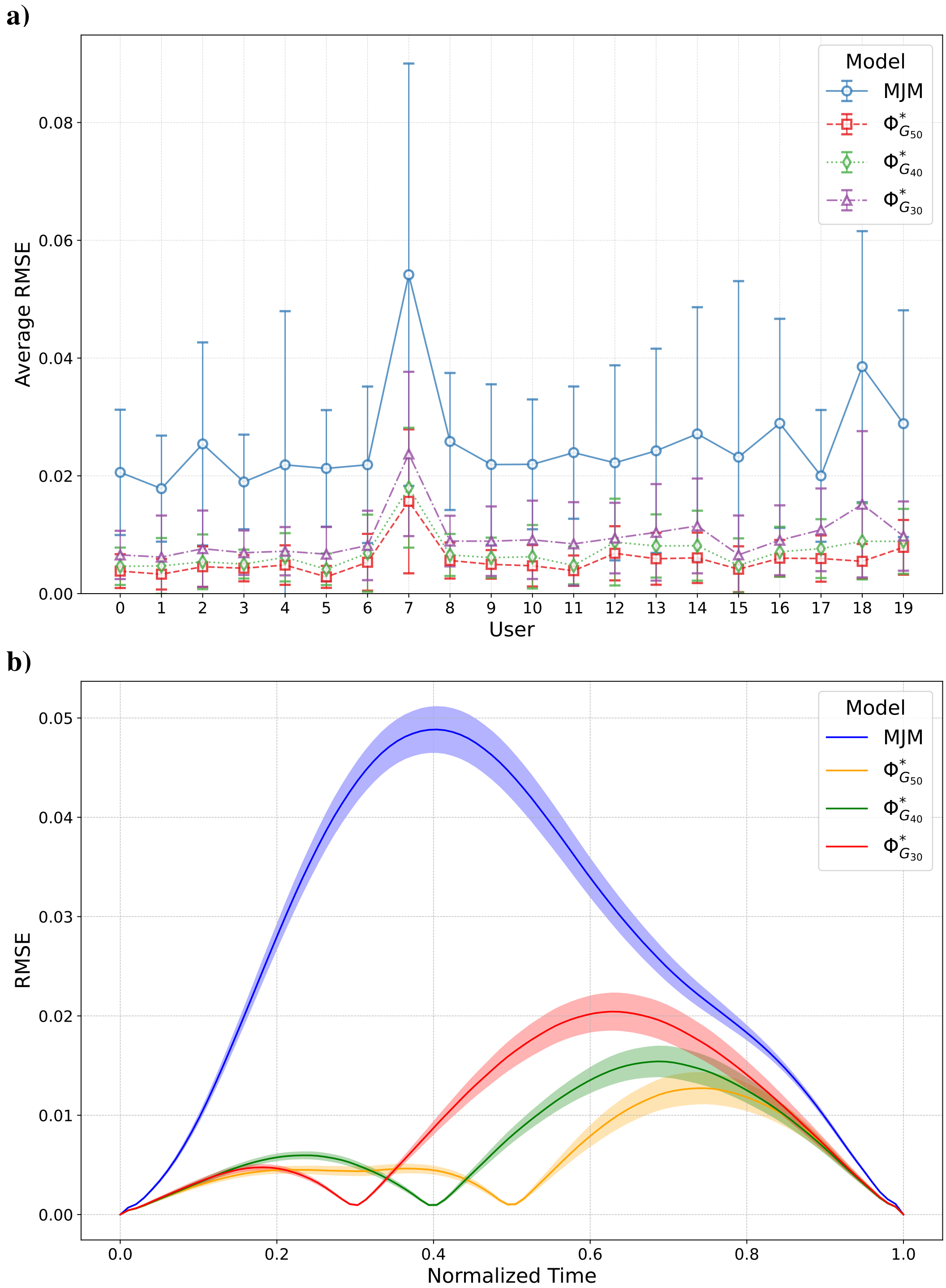


FIGURE 4.8: A comparison of RMSE for a) Average RMSE per user, b) Average RMSE for all users against normalized time for Minimum Jerk Model and $\Phi^*_{G_p}$ models.

a)

b)

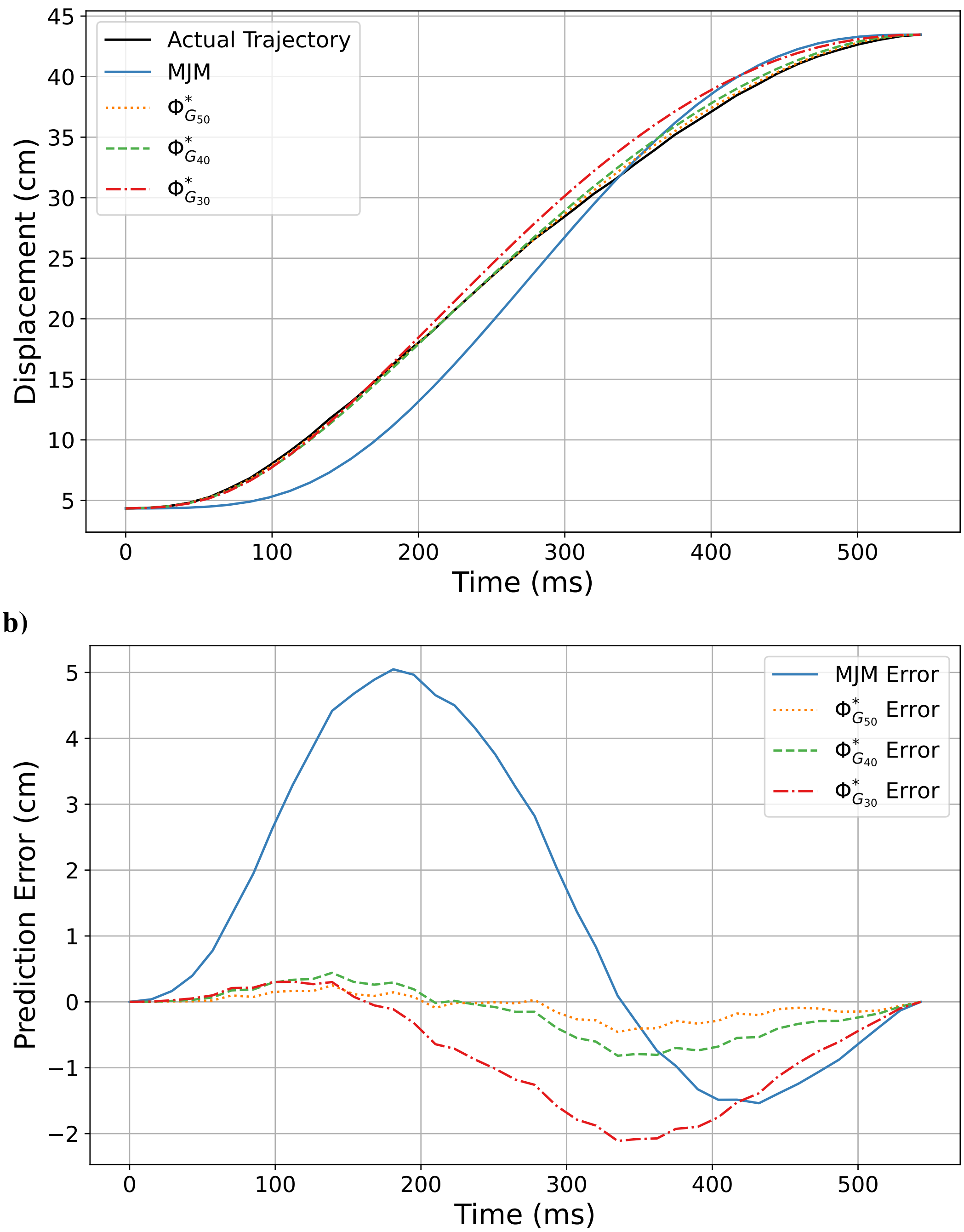


FIGURE 4.9: A comparison of Minimum Jerk Model and $\Phi^*_{G_p}$ models for one gesture of a user a) comparing displacement, b) comparing the RMSE of displacement, and c) showing 3D trajectory.

c)

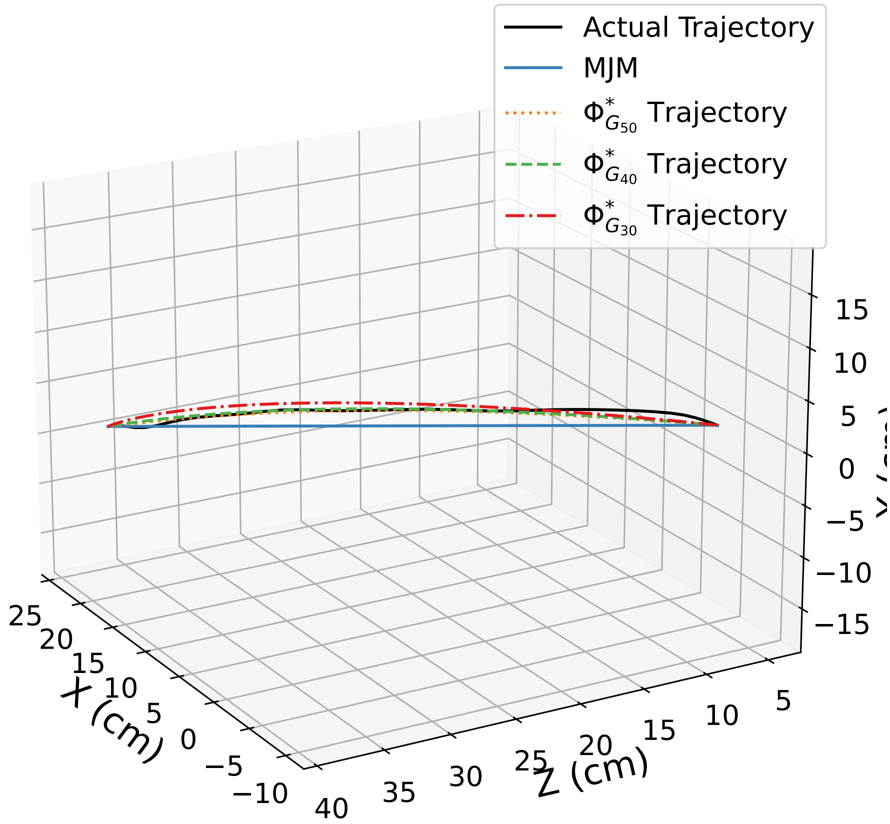


FIGURE 4.9: A comparison of Minimum Jerk Model and $\Phi^*_{G_p}$ models for one gesture of a user a) comparing displacement, b) comparing the RMSE of displacement, and c) showing 3D trajectory.

4.6.4 Verification of the Models with a New Dataset

To ensure the generalizability of our models, we tested them using a publicly available third-party dataset by Clarence et al. [27]. The dataset consists of reaching movements recorded from 11 users using a VR headset. For more details about the dataset, we refer readers to[27]. We preprocessed this dataset following the approach described in Section 3.2 and evaluated the performance of Φ^*_U and Φ^*_G models.

Figure 4.10a compares the RMSE values between the MJM and Φ^*_U, indicating that Φ^*_U better approximates the normalized trajectory, achieving a 73.85% average RMSE reduction across all users. Similarly, Figure 4.10b shows the average RMSE values computed for all gestures by each user, demonstrating that Φ^*_U outperforms the MJM, with an average RMSE reduction

a)

b)

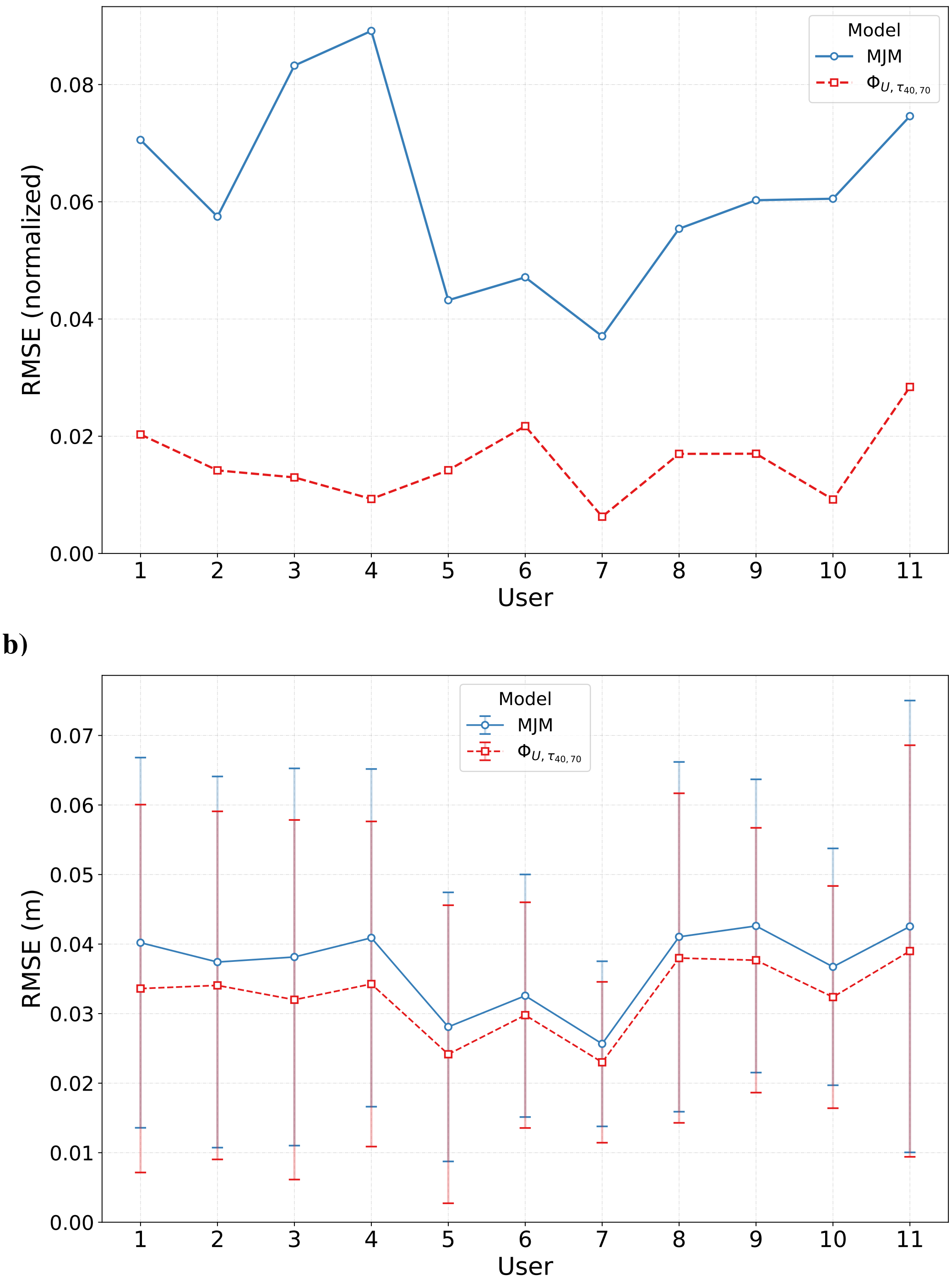
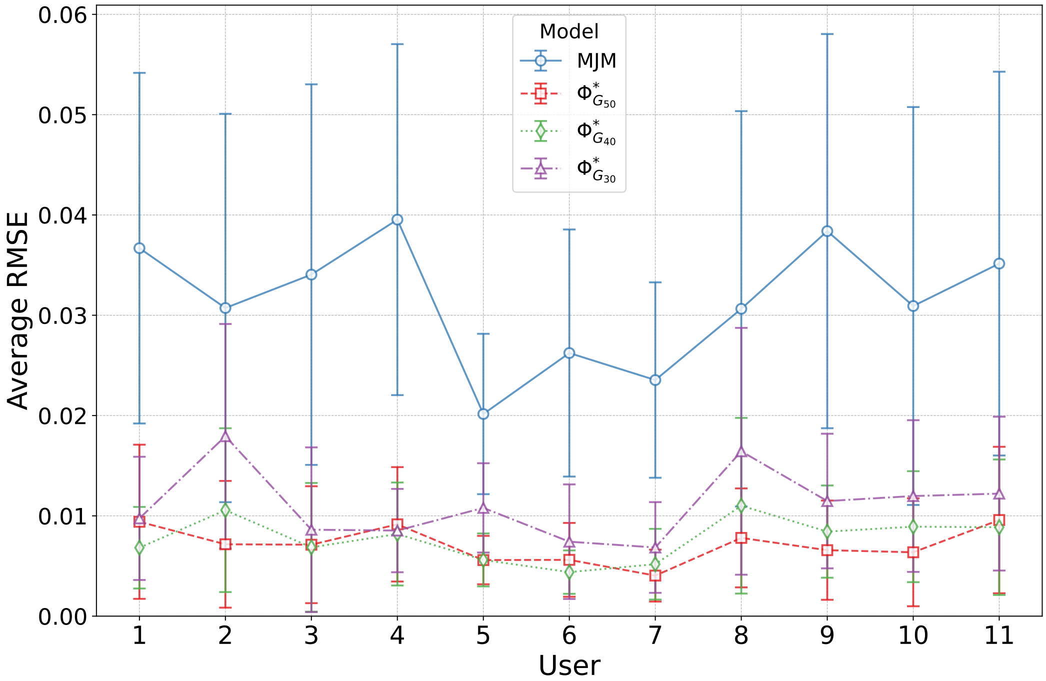


FIGURE 4.10: A comparison between MJM and Φ_U^* models. a) Error between each user's normalized trajectory and models. b) RMSE value calculated for all gestures by a user with the models for the new dataset.

a)



b)

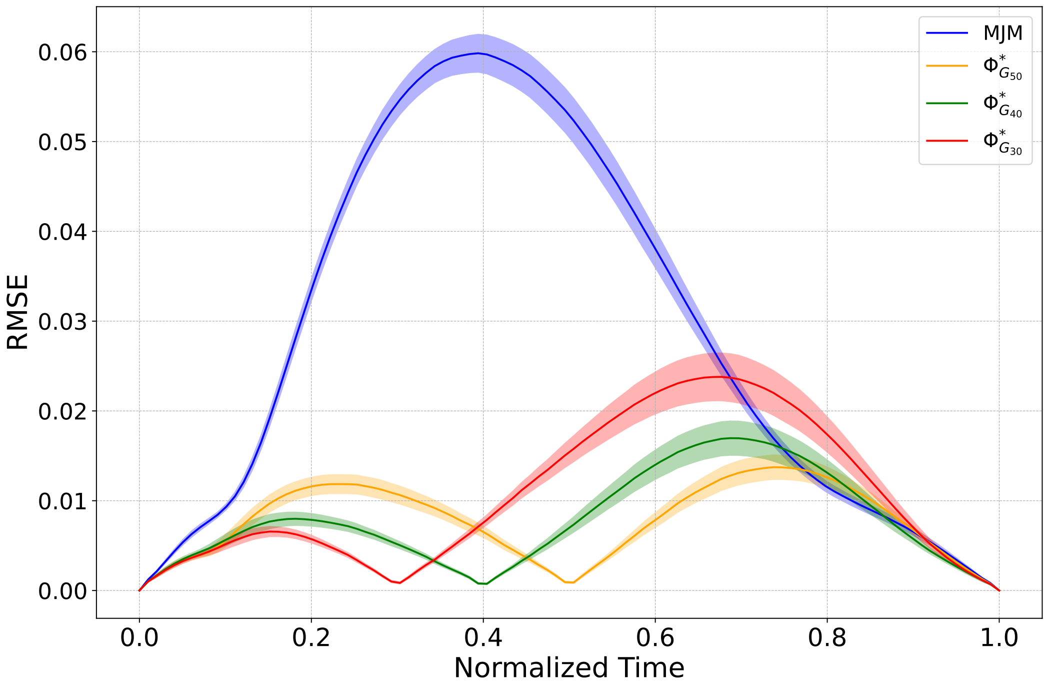


FIGURE 4.11: A comparison of RMSE for a) Average RMSE per user, b) Average RMSE for all users against normalized time for Minimum Jerk Model and $\Phi^*_{G_p}$ models for the new dataset.

of 11.80% across all users. Consistent with the previous dataset, we observe significant variability in how users perform gestures, which is not fully captured by the Φ_U^* model.

4.6.5 Φ_U^* Model User Coefficients for New Dataset

User	b_2	b_3	b_4	b_5
1	4.222993	-3.464049	-0.740880	0.981936
2	2.780025	1.824596	-6.989267	3.384646
3	5.547745	-8.409852	5.176468	-1.314362
4	5.123608	-6.301496	2.232168	-0.054280
5	3.803284	-5.165832	3.921813	-1.559264
6	4.035231	-4.215177	1.324661	-0.144715
7	3.181949	-1.769844	-1.006160	0.594055
8	2.961736	0.837590	-5.560390	2.761063
9	3.746464	-2.004458	-2.230475	1.488470
10	3.491412	-1.105892	-3.262453	1.876933
11	4.716337	-5.253809	1.358608	0.178864

TABLE 4.3: Table showing b_2, b_3, b_4, and b_5 values for each user for the new dataset.

Coefficient	**Mean**	**Std**	**Min**	**Max**
b_2	3.9646	0.8835	2.7800	5.5477
b_3	-3.1844	3.0879	-8.4099	1.8246
b_4	-0.5251	3.8036	-6.9893	5.1765
b_5	0.7449	1.5571	-1.5593	3.3846

TABLE 4.4: Mean, standard deviation, minimum, and maximum values for each coefficient for the new dataset.

Table 4.3 presents the Φ_U^* coefficients for each user in this new dataset. Among these users, User 2 exhibits the slowest initial acceleration with a coefficient of 2.78, while User 3 demonstrates the highest at 5.55. These values align with the user coefficient range reported in Table 4.1, indicating that the normalized trajectories remain comparable despite being collected under different experimental conditions.

Table 4.4 provides a statistical overview of these coefficients. A Welch's t-test was conducted using Table 4.2 and Table 4.4 to compare the two distributions, yielding p-values of 0.02, 0.13, 0.49, and 0.96 for the four comparisons of b_2, b_3, b_4, and b_5 coefficients respectively. Although the t-test value of b_2 coefficient falls below the conventional 0.05 threshold, the other three are notably higher. Therefore, the overall differences between these datasets are not pronounced, as any potential disparity in one measure is offset by the non-significance of the others.

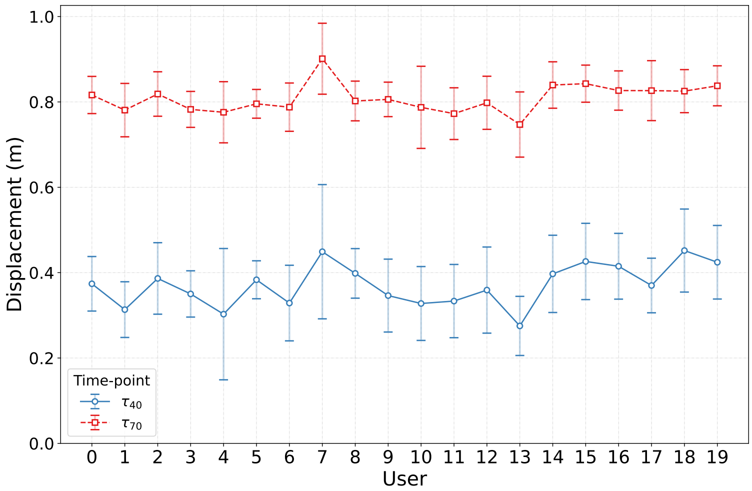
Similar to the evaluation of Φ_U^*, we assessed the performance of our Φ_G^* models using the new dataset. As shown in Figure 4.10a, which compares the average RMSE computed for all gestures by each user when applying the MJM and three variants of the Φ_G^* models, $\Phi_{G_{50}}^*$ emerges as the most accurate, achieving an average RMSE reduction of 77.29% across all users. The highest reduction is observed for User 9 with 82.87%, while the minimum reduction is 72.26% for User 11. Figure 4.8b shows the average RMSE computed for all users against the normalized time for this dataset. The performance trends of the Φ_G^* models closely resemble those observed in the previous dataset, as illustrated in Figure 4.8b. These results demonstrate that even with new users and in different environments, our models perform reliably well for reaching movements.

4.7 Trajectory Synthesis

In this section, we show how our models are utilized to synthesize user-specific hand trajectories for different reaching movements. Our approach not only generates trajectories for different movements but can also synthesize multiple plausible variations for the same movement, allowing repetitive actions to exhibit slight differences.

As shown in Figure 4.4d, the normalized trajectories for each user are different. These unique variations are captured by the corresponding d_{τ_k} values, and the distribution of values for $k = 40, 70$ is shown in Figure 4.12a. To generate a trajectory, we choose $\alpha \in [-1, 1] \subset \mathbb{R}$, where α is any real number between -1 and 1. The trajectory is then generated by utilizing the model Φ_U^* through solving Equation 4.13, where $d_{\tau_k} = \mu_{d_{\tau_k}} + \alpha\sigma_{d_{\tau_k}}$ for each user. Similar to the classical MJM, the trajectories synthesized by this technique produce a straight line. However, as shown in Figure 4.9, some observed trajectories are curved in the vertical (y) direction. To address this, we solve for the y direction separately using Equation 4.16, where $d_{\tau_p} = \mu_{d_{\tau_p}} + \beta\sigma_{d_{\tau_p}}$ and $\beta \in [0, 1] \subset \mathbb{R}$, to account for overshoots. Each different value of

a)



b)

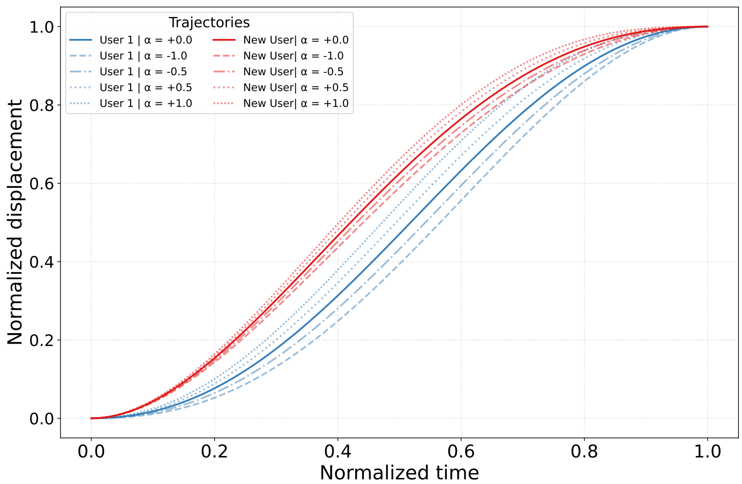


FIGURE 4.12: a) Comparison between each user's d_{τ_k} values for $k = 40, 70$. b) Synthesized trajectories for an existing user and a fabricated new user.

α and β produces a unique trajectory, allowing this technique to generate numerous distinct trajectories. This enables applications to introduce natural variation across repetitions, where the difference between trajectories is proportional to the difference between the α and β values used. Figure 4.12b shows synthesized trajectories for an existing user and a fabricated new user, where $\alpha = \beta$ is used to generate the trajectories.

In addition to generating trajectories for known users, it is also possible to fabricate new, unseen users. To enable this, we first calculate the distribution of the mean and standard deviation across all 20 users for $\mu_{d_{\tau_k}}$ and $\sigma_{d_{\tau_k}}$, which we denote as $\mu_{\bar{\mu}_{d_{\tau_k}}}$, $\sigma_{\bar{\mu}_{d_{\tau_k}}}$, $\mu_{\bar{\sigma}_{d_{\tau_k}}}$, and $\sigma_{\bar{\sigma}_{d_{\tau_k}}}$, respectively. Using these population-level statistics, new user profiles can be generated by selecting $\bar{\alpha}$ and $\bar{\beta}$ within the range $[-1, 1]$, and setting:

$$\mu_{d_{\tau_k}}^{new} = \mu_{\bar{\mu}_{d_{\tau_k}}} + \bar{\alpha}\sigma_{\bar{\mu}_{d_{\tau_k}}} \tag{4.18}$$

$$\sigma_{d_{\tau_k}}^{new} = \mu_{\bar{\sigma}_{d_{\tau_k}}} + \bar{\beta}\sigma_{\bar{\sigma}_{d_{\tau_k}}} \tag{4.19}$$

By varying $\bar{\alpha}$ and $\bar{\beta}$ within the defined range, a wide range of plausible trajectories for new, unseen user profiles can be synthesized.

4.8 Discussion

In this paper, we demonstrated that for reaching gestures in VR environments, classical MJM and MSM models exhibit significant errors when predicting users' point-to-point reaching movements. To address this limitation, we proposed a user-specific empirically modified minimum jerk model, which approximates users' likely trajectories based on empirical data. Additionally, we observed that some users display high variance in performing gestures, prompting the development of a gesture-specific empirically modified minimum jerk model, which accurately predicts the trajectory even when the gesture is only partially completed. Our results show that these models outperform the traditional MJM and achieve comparable performance on a new dataset with different users. Furthermore, the proposed models are mathematically derived, transparent, and computationally efficient, making them suitable alternatives to deep learning approaches for predictive systems.

In this section, we begin by discussing the results obtained from our study and the observations derived from their analysis. Finally, we explore the implications of the proposed system, its limitations, and potential directions for future research.

Coefficient	b_2	b_3	b_4	b_5
b_2	1.0000	-0.9700	0.8886	-0.7817
b_3	-0.9700	1.0000	-0.9735	0.9098
b_4	0.8886	-0.9735	1.0000	-0.9807
b_5	-0.7817	0.9098	-0.9807	1.0000

TABLE 4.5: Correlation matrix of the coefficients.

4.8.1 Effect of the coefficients in the Φ_U^* Model

In the Φ_U^* model, the normalized trajectory for each user is determined by four coefficients, b_2, b_3, b_4, and b_5, where each coefficient plays a role in shaping different phases of the movement. The coefficient b_2 affects the initial acceleration, where higher values lead to a steeper onset. The coefficient b_3 shapes the curvature during the early to mid-phase, and its variability reflects users' differing strategies in building up momentum. The coefficient b_4 contributes to deceleration, with negative values indicating a slowing down as the movement progresses. The coefficient b_5 impacts the smoothing of the trajectory towards the target, with positive values facilitating a gentle approach. The combined effect of these coefficients results in individualized movement profiles, reflecting the unique motor control characteristics of each user.

Table 4.1 shows the Φ_U^* coefficients for each user, while Table 4.2 provides a summary of their statistical measures. The coefficient b_2 ranges from 1.00 to 5.79, with a mean of 3.01 (SD = 1.22), indicating moderate variability among users in the initial acceleration phase of the trajectory. In contrast, b_3 exhibits higher variability, extending from -10.24 to 3.72 with a mean of -1.25 (SD = 3.67), suggesting significant differences in the curvature during the early to mid-phase of movement among users. Similarly, b_4 displays substantial variability with values between -6.21 and 8.10 and a mean of -1.53 (SD = 3.89), reflecting a general tendency towards deceleration in the mid to late phases of the trajectory. The coefficient b_5 ranges from -2.65 to 3.24 with a mean of 0.77 (SD = 1.43), indicating relatively lower variability and a consistent contribution to smoothing the trajectory towards the target.

Table 4.5 shows the correlations between these coefficients, revealing notable relationships that influence the trajectory's shape. Consecutive coefficients exhibit strong negative correlations,

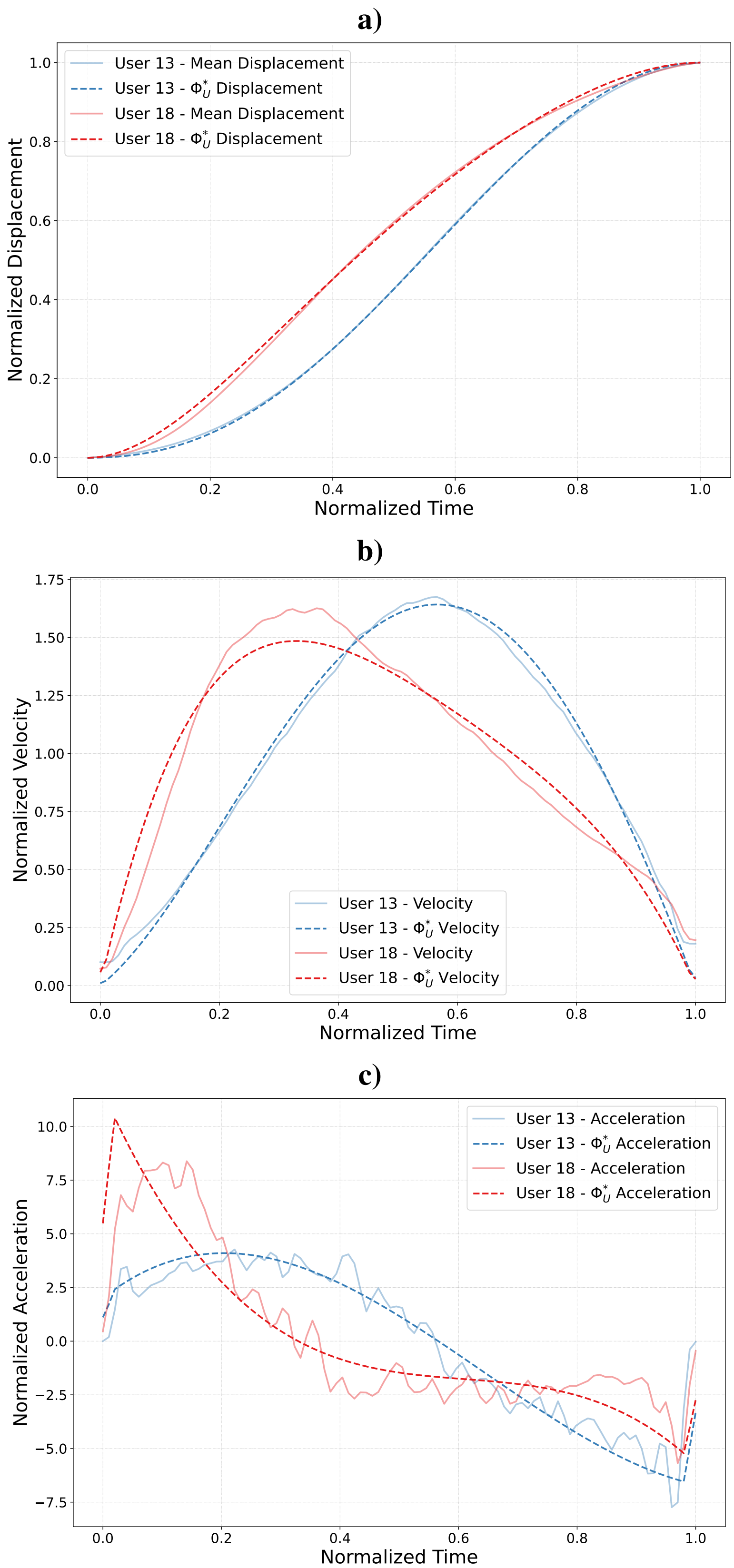


FIGURE 4.13: A comparison of a) displacement, b) velocity, and c) acceleration between the MJM and Φ_U^* models for User 13 and User 18, highlighting the differences in how these two users perform reaching gestures on average.

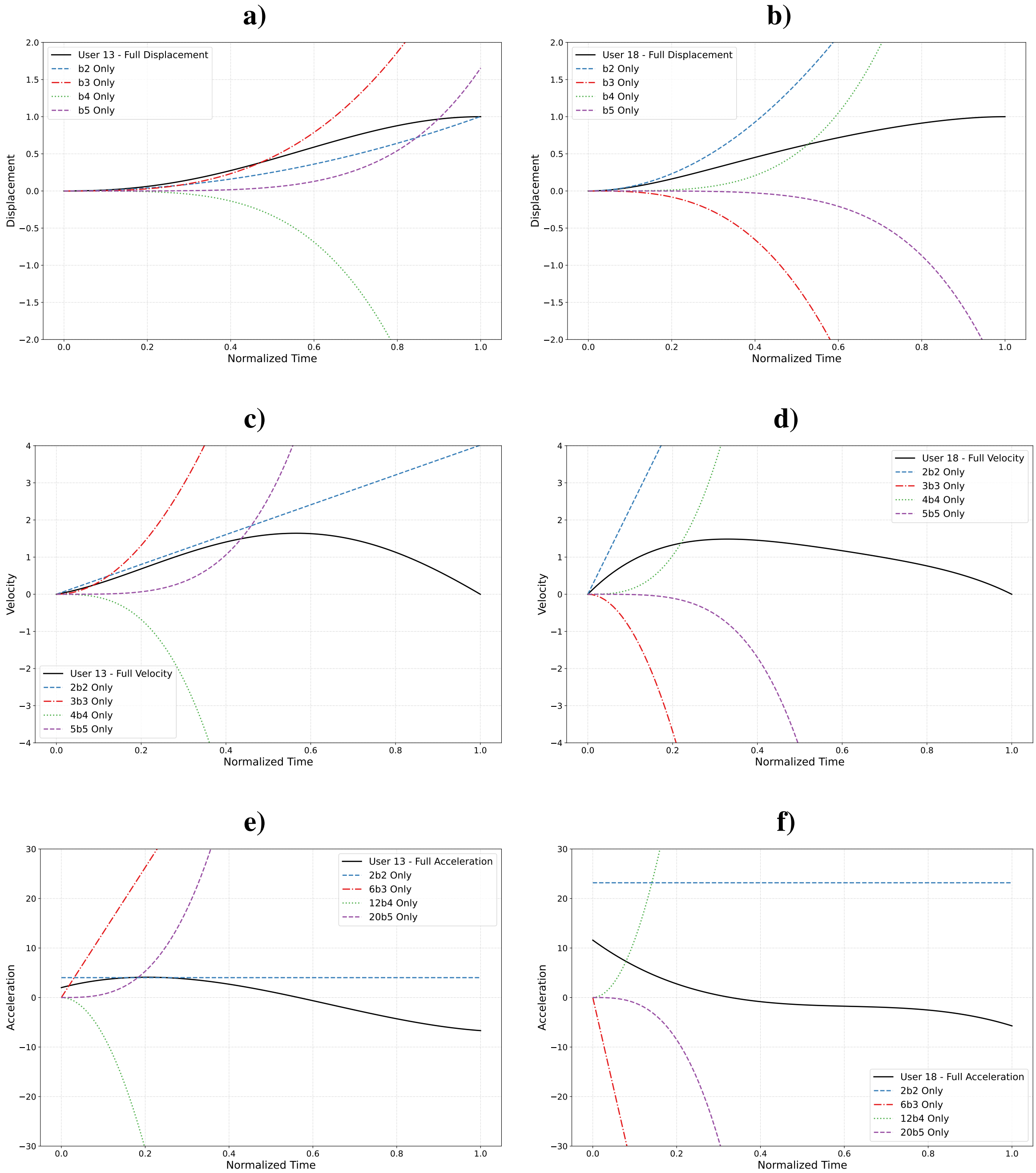


FIGURE 4.14: The movement profiles for Users 13 and 18, illustrating the effect of the b_1, b_2, b_3, and b_4 coefficients on displacement, velocity, and acceleration. For User 13, these effects are shown in plots a), c), and e), and for User 18, in plots b), d), and f), respectively.

such as b_2 with b_3 (r = -0.97), b_3 with b_4 (r = -0.97), and b_4 with b_5 (r = -0.98). These relationships indicate that increases in one coefficient are generally balanced by decreases in the subsequent coefficient. This pattern of alternating correlations suggests that compensatory mechanisms in motor control adjust the coefficients to maintain smoothness and efficiency in the trajectory.

Identifying users with extreme coefficient values provides insight into individualized movement strategies. User 18, with the highest b_2 value of 5.79, exhibits a trajectory characterized by rapid initial acceleration, suggesting a propensity for quick movement initiation. In contrast, User 13 has the lowest b_2 value of 1.00, potentially indicating a slower start or initial deceleration, depending on the interplay with other coefficients.

Figure 4.13 illustrates the movement profiles for Users 13 and 18. In the displacement profiles shown in Figure 4.13(a), User 18 exhibits a steeper initial displacement gain compared to User 13. This trend is also evident in the velocity and acceleration profiles, shown in Figures 4.13(b) and (c), respectively. User 18 demonstrates a higher acceleration phase, reaching peak velocity earlier than User 13. In contrast, User 13, while slower in the initial phase, achieves a higher peak velocity than User 18 and displays a shorter deceleration phase.

Figure 4.14 highlights how distinct coefficient values influence different phases of the trajectories. User 13 has b_1, b_2, b_3, b_4 coefficients as $1.00, 3.65, -5.31, 1.66$, while User 18's coefficients are $5.79, -10.24, 8.10, -2.66$, respectively. The higher value of b_2 for User 18 clearly contributes to their steeper initial displacement, as well as their initial velocity and acceleration profiles. However, as time progresses, the influence of the higher-order coefficients becomes more prominent due to their multiplication with higher powers of time. This example demonstrates how variations in coefficient values effectively capture users' individual movement patterns.

4.8.2 Limitations and Future Work

Predictive models are crucial in VR systems for anticipating future user interactions. For example, they can predict when a user is likely to interact with a virtual object, allowing the system to pre-render graphics and reduce system delays [138]. They also have potential applications such as haptic retargeting, as shown by Clarence et al. [27], who demonstrated that predictive models for reaching movements can be applied to haptic retargeting, where the user's real and virtual hands are decoupled, enabling a single physical object to provide haptic feedback for multiple virtual objects. Therefore, having simple yet accurate predictive models, such as those proposed in this paper, is important, as they can be easily deployed on resource-constrained devices such as VR headsets.

It is important to note that these models were evaluated only for aimed hand movements, primarily consisting of voluntary ballistic movements. We did not test them on complex movements involving sudden changes in direction or sequences of multiple point-to-point movements, and such cases were filtered out during data preprocessing.

Our models were trained and tested on reaching movements where users extend their arms primarily along the sagittal plane, which divides the body into left and right halves, representing forward motion. However, users can also perform other point-to-point movements, such as retracting their hands backward or moving their hands in vertical or horizontal directions. The methodology used to develop our predictive models has no inherent limitations preventing its application to these other types of movements. However, this requires further exploration to validate its effectiveness.

We developed and evaluated our models using datasets collected in a VR environment, with data captured through a VR headset. However, our models are not fundamentally limited to VR-based movements or VR headset data. We believe that our methodology can be applied to any 3D coordinate capture system, including motion capture setups and sensor-based technologies. It could also be extended to non-VR environments such as augmented reality (AR), mixed reality (XR), or real-world applications. In addition, predicting hand motion is essential for anticipating human actions, enabling safer and more In addition, predicting hand motion is essential for anticipating human actions, enabling safer and more efficient human-robot collaborative environments [54, 19, 190]. Furthermore, it has applications in rehabilitation [7] and in generating realistic human avatars for various interactive applications [180]. This adaptability opens up diverse possibilities for implementation across various platforms and use cases, though it requires further investigation in these different contexts.

4.9 Conclusion

In this paper, we introduced novel user-specific and gesture-specific empirically modified minimum jerk models to predict hand trajectories for reaching movements in VR environments. The user-specific model approximates a user's average trajectory based on empirical data, while the gesture-specific model predicts the trajectory when the gesture is partially completed, offering higher accuracy by accounting for gesture-specific variations. We developed and evaluated these models using a dataset containing over 2000 reaching gestures from 20 users. Additionally, we validated our approach using a third-party dataset with 11 users, demonstrating that the models generalize well across different datasets. Our predictive models are useful in VR applications for tasks such as pre-rendering graphics and enabling haptic retargeting. While this paper focuses exclusively on reaching gestures in VR, we believe extending this work to other point-to-point movements and application domains presents a promising direction for future research.

CHAPTER 5

Discussion and Conclusion

This final chapter reflects on the main findings of this thesis. It begins by discussing the user study and dataset, addressing its limitations and offering additional insights not covered in earlier chapters. The chapter revisits the research questions initially proposed and examines how each has been addressed in Chapters 2, 3, and 4. Furthermore, it explores the contributions, implications, potential applications, limitations, and directions for future research. Finally, the chapter concludes with a brief summary of the thesis, emphasizing its key contributions and insights.

5.1 User Study Dataset

In Chapter 3, we described the user study conducted to collect a comprehensive hand motion dataset in VR across four distinct applications. In this section, we provide a detailed exploration of the dataset, highlighting its limitations and offering our insights into how different attributes of the dataset relate to the predictive models.

5.1.1 Demographic and Physical Attributes

Table 5.1 shows the demographic and physical information for the 20 participants in the dataset. Among the participants only 7 (35%) were female and the other 13 (65%) were male with shows a considerable skew towards male participants in the gender distribution. This imbalance suggests that male and female participants are not equally represented, which may limit the ability to generalize findings across genders. Among all the participants, only 1 (5%) was left-handed, while the remaining 19 (95%) were right-handed. Although this left-handed

User	Gender	Dominant Hand	Age	Height (cm)	Weight (kg)	BMI	VR Usage (Num times)
0	Male	Left	39	188	84	23.77	2-5
1	Male	Right	21	181	75	22.89	5-10
2	Male	Right	21	193	99	26.58	1
3	Male	Right	21	182	70	21.13	1
4	Male	Right	21	190	78	21.61	2-5
5	Male	Right	25	172	63	21.30	5-10
6	Male	Right	21	186	70	20.23	5-10
7	Female	Right	33	160	45	17.58	0
8	Male	Right	18	178	70	22.09	2-5
9	Male	Right	26	181	80	24.42	5-10
10	Male	Right	22	173	78	26.06	2-5
11	Male	Right	19	181	88	26.86	2-5
12	Male	Right	18	170	60	20.76	2-5
13	Male	Right	18	185	67	19.58	5-10
14	Female	Right	18	167	53	19.00	2-5
15	Female	Right	18	163	51	19.20	2-5
16	Female	Right	22	157	78	31.64	0
17	Female	Right	23	160	60	23.44	5-10
18	Female	Right	22	155	43	17.90	0
19	Female	Right	22	162	58	22.10	0
Mean			22.4	174.2	68.5	22.41	
Std			5.13	11.58	14.21	3.37	

TABLE 5.1: Demographic details, BMI, and VR usage patterns for users, including the mean and standard deviation for age, height, weight, and BMI.

participant was not treated differently in any of our models, the dataset's heavy skew toward right-handed participants limits its ability to fully recognize or analyze effects of the dominant hand on movements.

The age range of participants spans from 18 to 39 years, with only 2 participants aged over 26 (33 and 39). This reflects a relatively narrow focus on young to early-middle-aged adults, with a mean age of 22.4 years (SD = 5.13). Such a limited age distribution may restrict the broader applicability of our findings to older populations, early teenagers, and children. With this dataset, we are unable to comment on whether the movement characteristics of these groups are different.

We calculated Body Mass Index (BMI) using participants' height (mean = 174.2 cm, SD = 11.58 cm) and weight (mean = 68.5 kg, SD = 14.21 kg) values. The resulting BMI values (mean = 22.41, SD = 3.37) were classified based on standards from the Australian Government Department of Health and Aged Care [1]. The majority of participants (65%) fell within the healthy weight range (BMI 18.5–25), reflecting a predominantly healthy-weight sample. Smaller proportions were observed in other categories, with 2 (10%) participants classified as underweight (BMI < 18.5), 3 (15%) as overweight (BMI 25–30), and only 1 (5%) as obese (BMI > 30). This distribution may also affect the application of our findings to populations with a broader BMI range. For example, according to 2022 statistics on the Australian population [166], the BMI distribution among adults was 1.6% underweight, 31.6% healthy weight, 34.0% overweight, and 31.7% obese, which is different from the dataset's distribution.

In addition, we assessed VR usage among participants, with 8 (40%) reporting 2–5 instances of use and 6 (30%) reporting 3–10 instances. Only 2 participants (10%) had used VR once, while 4 (20%) were completely new to VR. This indicates that 80% of participants in this dataset had prior experience with VR before this study.

5.1.2 Relationship with the User-Specific Models

In Chapter 4, we described the development of User-Specific Models (Φ_U^*) to represent each user's normalized trajectories for reaching movements. In this section, we examine the relationships between users' demographic and physical attributes and these models, specifically focusing on the user coefficients listed in Table 4.2.

Attribute	b_2	b_3	b_4	b_5
Age	0.03	0.08	-0.19	0.27
BMI	-0.09	0.04	0.01	-0.05
Gender (Numeric)	-0.57	0.42	-0.26	0.11
VR Usage (Numeric)	-0.62	0.56	-0.46	0.37

TABLE 5.2: Correlation coefficients for user attributes across b_2, b_3, b_4, and b_5.

We initially hypothesized apparent relationships between users' physical attributes, and the four coefficients (b_2, b_3, b_4, b_5). However, no direct relationships were identified. Table 5.2 presents the calculated Pearson correlation coefficients between age, BMI, gender, and VR

[1] https://www.health.gov.au/topics/overweight-and-obesity/bmi-and-waist

usage with these coefficients. For the calculations, gender was mapped as 0 for males and 1 for females, while VR usage was mapped to the mean of the corresponding usage range.

A higher b_2 coefficient corresponds to a faster initial acceleration phase. Our analysis indicates a moderate relationship between gender and b_2, with a correlation coefficient of -0.57. This suggests that male participants generally exhibited a faster initial acceleration phase. A similar relationship was observed with b_3, which influences the early to mid-phase of the movement. This finding highlights an intriguing relationship, suggesting that reaching movement characteristics vary by gender. However, due to the small sample size in the dataset, no definitive conclusions can be drawn, and further investigation is required.

We also observed a strong correlation between VR usage and the b_2 coefficient. Interestingly, this relationship is the inverse of what might have been expected. With a correlation coefficient of -0.62, it suggests that participants with less VR experience exhibited a faster initial acceleration phase. We are unable to hypothesize the reason for this observation, which requires further investigation.

Surprisingly, no strong correlations were observed between age or BMI and the coefficients. We initially hypothesized inverse relationships between age and BMI with b_2, but this was not the case. This outcome may be attributed to the lack of older and overweight participants in the dataset.

5.2 Research Questions

In this section, we revisit the research questions and briefly discuss how they have been addressed in this thesis.

RQ1 *How can the limitations and challenges of existing hand motion prediction and analysis approaches inform the development of effective solutions for VR applications?*

Chapter 2, Sections 3.2 and 4.2 focused on the key techniques used for predicting and analyzing human hand motion, specifically within the context of VR. We discussed various approaches, including statistical methods such as template matching and Hidden Markov Models (HMMs), as well as deep learning techniques like Long Short-Term Memory networks (LSTMs) that are commonly employed in hand prediction for VR environments.

Additionally, we explored hand motion prediction outside the VR context to gain a comprehensive understanding of the broader state of the art. This included methods such as kinematic modeling, regression-based approaches, other Deep Neural Networks (DNNs), and hybrid strategies like Physics-Informed Neural Networks (PINNs). Through these analyses, we gained a deeper understanding of the respective strengths and weaknesses of different prediction techniques across various domains.

However, we identified significant limitations in existing techniques when applied to hand motion prediction in VR. Statistical methods often struggle with generalizability in dynamic VR environments, making it difficult to maintain accuracy across diverse scenarios and new users. While deep learning methods are more accurate and generalizable due to their ability to learn complex relationships from large motion datasets, they are also highly computationally intensive. This computational demand poses challenges for deployment on resource-constrained devices such as standalone VR headsets, where processing power and energy efficiency are critical considerations.

RQ2 *How can effective prediction frameworks be developed to address challenges such as accuracy, computational efficiency, and interpretability?*

To address the challenges of efficiency, and accuracy in hand motion prediction techniques for VR, we developed and evaluated prediction methods that combine statistical approaches with kinematics and empirical data. Our goal was to create simple, computationally efficient, and interpretable prediction models that are easy to understand and deploy on real-world VR devices. Throughout this thesis, we demonstrated how empirical observations from multiple users and kinematic models can be integrated with traditional statistical methods to build more effective prediction models.

In Chapter 3, we explored how kinematics-driven, empirical data-based regression models can be developed for both structured and unstructured ballistic 3D hand movements in VR activities. We combined the well-known Minimum Jerk Model with regression models to create both personalized user-specific models and generalized gesture-specific models.

In Chapter 4, we introduced a method that combines well-known kinematic models with empirical observations from a user study to predict trajectories for reaching activities. We demonstrated how the trajectory prediction task can be formulated as an optimization problem, enabling a mathematical solution for accurate trajectory prediction.

The prediction models developed in this thesis are expressed as simple mathematical equations, aligning with our goal of creating computationally efficient and interpretable models. Additionally, we thoroughly validated these models, demonstrating their effectiveness across various scenarios.

RQ3 *How can generalized hand motion models be developed to perform comparably to personalized models?*

To improve the generalizability of hand prediction techniques in VR, we aimed to develop models that can be applied to new users, activities, and datasets without requiring extensive modifications.

In Chapter 3, we introduced a user- and activity-independent model capable of continuously predicting 3D hand trajectories in VR. We demonstrated that this model achieves performance comparable to personalized models without the need for additional training phases. Its effectiveness was further validated through a secondary user study involving new users and activities in VR.

In Chapter 4, we developed a gesture-specific empirically modified Minimum Jerk Model, designed to predict the trajectory of reaching movements based on the percentage of movement completed. The methodology we used to develop this approach is user-agnostic, and we successfully validated the model using a third-party dataset containing new users, confirming its generalizability.

5.3 Contributions

In this section of the thesis, we briefly outline its main contributions.

We conducted a user study with 20 participants, each performing four different VR activities for a one-hour period, as discussed in Chapter 3. Each participant completed a structured activity involving point-to-point movements and the other three activities consisted of playing three commercially available VR games. We collected the activity data using multiple sensor systems, including the OptiTrack motion capture system, an Oculus VR headset, and six IMU units fitted on both arms of the participants. This dataset is highly detailed and informative, unmatched by existing VR hand motion datasets, and holds significant potential for future research.

To the best of the authors' knowledge, we developed the first user- and activity-independent generalized 3D hand prediction model capable of continuously predicting a user's hand trajectory in Chapter 3. This model was created by combining empirical data and kinematic principles with a classical regression approach. It does not require additional training phases for new users or activities, making it adaptable and versatile. Furthermore, the model can be expressed as a simple mathematical equation, and it is both computationally efficient and easy to implement. We demonstrated that this model can predict hand positions with sub-centimeter accuracy for prediction intervals under 200 cm. It achieves low RMSE values of 0.80 cm, 0.85 cm, and 3.15 cm when predicting future hand positions at 100 ms, 200 ms, and 300 ms ahead, respectively.

In Chapter 4, we demonstrated the development of empirically modified Minimum Jerk Models by combining empirical observations with the traditional Minimum Jerk Model and formulating the problem as an optimization task. We introduced a user-specific model that approximates each user's average trajectory and a user-agnostic gesture-specific model that predicts the trajectory when the gesture is partially completed. Additionally, these models can also be expressed as a simple mathematical equation. Our findings show that the user-specific empirically modified model improves trajectory approximation by 83.8% and reduces RMSE by 16.9% across all users and gestures compared to the baseline. Gesture-specific models reduce average RMSE across all users and gestures by 62.7% and 78.2% when predicting trajectories after 30% and 50% of the gesture is completed, respectively.

5.4 Implications of Findings

Our predictive models, discussed in detail in Chapters 3 and 4, demonstrate the effectiveness of combining traditional statistical methods with empirical data and motion kinematics. In Chapter 3, we integrated classical regression techniques with kinematics by incorporating motion derivatives directly into the regression process trained with empirical data. In Chapter 4, we extended the traditional Minimum Jerk Model by incorporating empirical observations and reformulating the trajectory prediction problem as a mathematical optimization problem, enabling trajectory formulation by solving the optimization problem. Both predictive models indicate the potential to balance theoretical knowledge and real-world data, enabling the development of accurate and computationally efficient predictive models. Since these models can also be expressed as mathematical equations, they are straightforward to understand and interpret. In Chapter 3, we developed a user- and activity-independent 3D trajectory prediction model, applicable to new users and activities without requiring additional training phases. Similarly, in Chapter 4, our gesture-specific models demonstrate applicability for new users, highlighting their generalizability.

We hypothesize that the simplicity of our approach, which integrates mathematically driven yet empirically calibrated motion models, makes it adaptable to a wide range of domains. The well-established theoretical principles underlying the prediction model make it possible for developers and researchers to easily tailor the equations and constraints to suit the specific needs of various tasks and user populations. In contrast, deep learning models are highly complex and often require an additional training phase for modifications. This adaptability underscores the potential of our approach to contribute meaningfully to advancing motion prediction across diverse applications.

Reflecting on our results, we found ourselves asking an important question about why our models work and whether everyone follows a general movement pattern. Our findings support

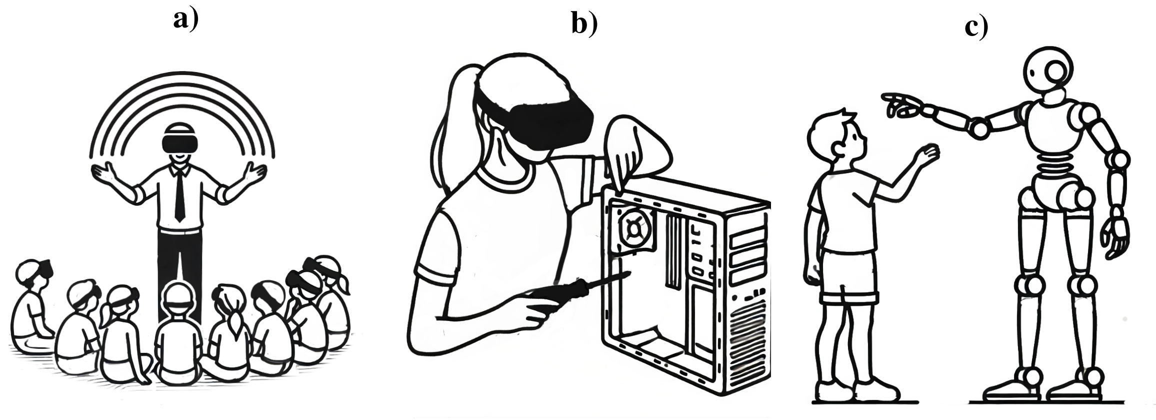


FIGURE 5.1: Potential applications: a) interactive storytelling, b) repair and maintenance tasks, c) human-robot collaborative environments.

this idea, suggesting the existence of a universal structure underlying human hand motion patterns. While individual users exhibit unique idiosyncrasies, this structure that governs the majority of hand movements can be effectively captured by our prediction models. We believe this explains why the generalized motion models achieve results comparable to personalized models trained for each user.

5.5 Potential Applications

There are a growing number of applications for VR, including immersive games like adventure or puzzle challenges, fitness programs such as guided yoga or virtual cardio sessions, interactive workouts like boxing or dance routines, and simulations of real-world activities such as driving, cooking, or medical training. However, hand motion prediction remains underutilized in the current commercial VR landscape, limited primarily to pre-rendering graphics to avoid system delays and haptic retargeting, where predictive models decouple a user's real and virtual hands, enabling a single physical object to provide haptic feedback for multiple virtual objects. We believe that models like ours can not only enhance these existing applications but have the potential to unlock a multitude of novel use cases.

We envision hand motion prediction to be helpful in enhancing accessibility for users with motor impairments by compensating for motion irregularities when doing tasks in VR. Even in the presence of motor impairments such as tremors, involuntary movements and coordination issues making it hard to perform smooth hand movements, and accurately predicting hand trajectory, the VR device can still provide fluid interactions, allowing users to perform the tasks with more confidence.

Similarly, we believe that our predictive models can be applied to interactive storytelling and teaching tasks in VR. By anticipating hand movements and understanding the context, the VR system can automatically generate visualizations to enhance engagement and make the storytelling or teaching experience more immersive. Figure 5.1(a) illustrates the authors' depiction of an interactive storytelling scenario, where the speaker's hand actions are predicted in advance and matched with the story's context to create VR visualizations, such as a rainbow in this example.

In addition, we see hand motion prediction as invaluable for real-world maintenance tasks in mixed reality. For instance, when a non-expert user performs a delicate repair, the system can predict the trajectory of their hand movement. If the predicted action risks damaging a component or deviating from the correct procedure, it can provide real-time warnings or corrective feedback. This can help reduce errors and enhance the user's confidence while performing the task. Figure 5.1(b) illustrates the authors' depiction of the above scenario, where a user is attempting to repair a computer with guidance provided by a VR application.

We also see our models being valuable beyond VR applications, particularly in human-robot collaborative environments. In Chapter 4, we discussed the intra-user and inter-user differences observed during reaching movements. If a robot were to perform reaching actions uniformly, its movements would appear more "robotic" than human-like. The user-specific models we developed could enable robots to replicate human-like movements, enhancing their ability to act more naturally and even imitate the movements of specific users. Figure 5.1(c) depicts a futuristic scenario where a robot performs human-like hand movements while giving directions.

5.6 Limitations and Future Work

This section discusses the limitations of the predictive models developed in this thesis and explores potential directions for future research.

5.6.1 Scope of Movement Data Used in Predictive Models

The data collected in our user study primarily consisted of aimed hand movements, even though it included both structured and unstructured movements in VR. For instance, in pointing tasks, the movements involved reaching a target, while in VR games, they often involved hitting or slashing a moving target. These aimed hand movements are largely voluntary and ballistic in nature, which are characterized by their quick, direct nature, where the hand moves toward a target with minimal trajectory adjustments, and the user does not change their intended target during the movement. For example, in pointing tasks, the movement involved reaching a target, or in other VR games, it involved hitting or slashing a moving target.

Nevertheless, we believe that our predictive frameworks are not restricted to specific types of movements and can be extended to others. These include involuntary hand motions influenced by external factors or lacking deliberate intent. They also include steering movements, where users continuously adjust the hand's target location in real-time, such as during line tracing. Additionally, movements performed while the user's feet are in motion, such as walking, jumping, or engaging in other dynamic activities, represent further opportunities for exploration.

This belief is supported by the design of our model in Chapter 3, which does not rely on underlying assumptions about the movement type. It uses only temporal positional data as input, with all other parameters derived from it. The model in Chapter 4 is more restrictive, and assumes point-to-point movements. However, we believe it can be extended beyond the tested reaching movements to any point-to-point scenario. Exploring and evaluating these models across various movement types would be an exciting direction for future research.

We developed and evaluated these models for movements within a 3D VR environment. However, our techniques are not theoretically limited to predicting hand movements in VR alone. Since our approach utilizes 3D location data, we believe it can be extended to other domains where 3D coordinate capture systems are employed. This includes non-VR environments such as augmented reality, mixed reality, and real-world applications. Our approach in developing these predictive models does not have any inherent constraints that would prevent its applications to these other environments. Therefore, it allows for a broader applicability of these models across various contexts. Exploring the models' effectiveness across these diverse scenarios presents an interesting direction for future research.

5.6.2 Extending Models Beyond Hand Motion

In this thesis, we focused on hand motion specifically and all our predictive models are based on temporal hand position data. However, we have several interesting considerations for potential extensions of our models. One possibility is adapting the proposed approach to motion trajectories of other body locations, specifically for the model presented in Chapter 3. For instance, incorporating the trajectories of the elbow and shoulder alongside the hand's trajectory could facilitate the development of a kinematic model for the entire arm. This extension could explore whether distinct models are needed for each joint or whether a universal model can effectively capture the trajectories of all joints. Additional extensions could involve considering the motion trajectories of both arms or legs, allowing for the analysis of coordinated limb movements and the potential development of models that account for full-body kinematics.

Another possible direction involves extending the models to angular data. Similar to temporal position data, these models could be applied to Euler angles or Quaternions to predict hand orientation. This approach is relevant to both models. For instance, derivatives of angular data could be incorporated into continuous prediction. Additionally, for reaching movements, assumptions about starting and ending angular positions could be used to predict changes in orientation along the movement trajectory.

5.6.3 Predicting Motion Dynamics: Force and Momentum

Another potential research direction involves predicting velocity and acceleration. These can either be obtained by calculating the time derivatives of the predicted positional data or by modifying the models to directly predict these dynamic quantities. Predicting velocity and acceleration offers the advantage of enabling the calculation of momentum and force if the hand's mass is known. For instance, in scenarios such as a user punching a target, momentum and force information could enhance the realism of rendered interactions. Additionally, applications could use this data to produce sound and haptic feedback proportional to the force, creating a more immersive user experience.

5.6.4 Multi-Modal Sensor Fusion

Our models are currently limited to using positional data. However, it would be interesting to explore how incorporating additional sensors could improve the accuracy and performance of the models. For instance, Electromyography (EMG) could be used to detect muscle activation, such as identifying the start of a movement. Another possibility is to incorporate head tracking and gaze data into the prediction process.

Other potential sensors include Inertial Measurement Units (IMUs), which could directly provide acceleration, angular velocity, and orientation data, and Galvanic Skin Response (GSR) sensors, which measure skin conductance and could offer insights into the user's emotional or physiological states that influence motion patterns. Exploring a combination of these sensors could open new possibilities for enhancing the predictive accuracy and adaptability of the models across diverse applications.

5.7 Conclusion

This thesis focused on developing computationally efficient, accurate, and generalizable hand motion prediction models for activities in VR environments. We began by collecting a comprehensive dataset of VR activities from 20 users, including both structured and unstructured movements, using multiple sensor systems. We proposed the first user- and activity-independent generalized 3D hand prediction model capable of continuously predicting a user's hand trajectory through a hybrid classical-regressive kinematics approach. Additionally, we introduced user- and gesture-specific empirically modified Minimum Jerk Models to predict reaching hand trajectories by formulating trajectory prediction as an optimization problem. Our models were thoroughly evaluated and validated with new users, activities, and datasets, demonstrating their generalizability. In conclusion, we believe this thesis makes a significant contribution to Human-Computer Interaction, particularly in the area of hand motion prediction. The techniques and models developed here lay a strong foundation for future innovations in this field, especially in advancing immersive VR devices and applications.

Bibliography

[1] Johnny Accot and Shumin Zhai. 'Beyond Fitts' Law: Models for Trajectory-Based HCI Tasks'. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. CHI '97. New York, NY, USA: Association for Computing Machinery, Mar. 1997, pp. 295–302. ISBN: 978-0-89791-802-2. DOI: 10.1145/258549.258760.

[2] Emre Aksan et al. 'A Spatio-temporal Transformer for 3D Human Motion Prediction'. In: *2021 International Conference on 3D Vision (3DV)*. Dec. 2021, pp. 565–574. DOI: 10.1109/3DV53792.2021.00066.

[3] A Deniz Aladagli et al. 'Predicting head trajectories in 360 virtual reality videos'. In: *2017 International Conference on 3D Immersion (IC3D)*. IEEE. 2017, pp. 1–6.

[4] Sadegh Aliakbarian et al. 'A stochastic conditioning scheme for diverse human motion prediction'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 5223–5232.

[5] Sadegh Aliakbarian et al. 'Contextually Plausible and Diverse 3D Human Motion Prediction'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 11333–11342.

[6] Amira Aloulou and Olfa Boubaker. 'Minimum Jerk-Based Control for a Three Dimensional Bipedal Robot'. In: *Intelligent Robotics and Applications*. Ed. by Sabina Jeschke, Honghai Liu and Daniel Schilberg. Berlin, Heidelberg: Springer, 2011, pp. 251–262. ISBN: 978-3-642-25489-5. DOI: 10.1007/978-3-642-25489-5_25.

[7] F. Amirabdollahian, R. Loureiro and W. Harwin. 'Minimum Jerk Trajectory Control for Rehabilitation and Haptic Applications'. In: *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*. Vol. 4. May

2002, 3380–3385 vol.4. DOI: 10.1109/ROBOT.2002.1014233. (Visited on 05/09/2024).

[8] Elena Arabadzhiyska et al. 'Saccade landing position prediction for gaze-contingent rendering'. In: *ACM Transactions on Graphics* 36.4 (2017). ISSN: 15577368. DOI: 10.1145/3072959.3073642.

[9] Stanislav Aranovskiy et al. 'Modeling pointing tasks in mouse-based human-computer interactions'. In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016, pp. 6595–6600.

[10] *Award Winning Motion Capture Systems*. https://www.vicon.com/. (Visited on 06/12/2024).

[11] Mahdi Azmandian et al. 'Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences'. In: *Proceedings of the 2016 chi conference on human factors in computing systems*. 2016, pp. 1968–1979.

[12] Ronald Azuma and Gary Bishop. 'A frequency-domain analysis of head-motion prediction'. In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 1995, pp. 401–408.

[13] Myroslav Bachynskyi and Jörg Müller. 'Dynamics of Aimed Mid-air Movements'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–12. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376194. (Visited on 26/08/2024).

[14] A. Terry Bahill and Jeffrey S. Kallman. 'Predicting Final Eye Position Halfway Through a Saccade'. In: vol. BME-30. 12. Dec. 1983, pp. 781–786. DOI: 10.1109/TBME.1983.325078.

[15] Emad Barsoum, John Kender and Zicheng Liu. 'HP-GAN: Probabilistic 3D Human Motion Prediction via GAN'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 1418–1427.

[16] Mattias Bratt, Christian Smith and Henrik I. Christensen. 'Minimum Jerk Based Prediction of User Actions for a Ball Catching Task'. In: *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2007, pp. 2710–2716. DOI: 10.1109/IROS.2007.4398989. (Visited on 05/09/2024).

[17] Daniel Bullock and Stephen Grossberg. 'Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation.' In: *Psychological review* 95.1 (1988), p. 49.

[18] Judith Butepage et al. 'Deep Representation Learning for Human Motion Prediction and Classification'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 6158–6166.

[19] Judith Bütepage, Hedvig Kjellström and Danica Kragic. 'Anticipating Many Futures: Online Human Motion Prediction and Generation for Human-Robot Interaction'. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. Brisbane, Australia: IEEE Press, May 2018, pp. 1–9. DOI: 10.1109/ICRA.2018.8460651. (Visited on 26/10/2024).

[20] Fabio Carrara et al. 'LSTM-based Real-Time Action Detection and Prediction in Human Motion Streams'. In: vol. 78. 19. Oct. 2019, pp. 27309–27331. DOI: 10.1007/s11042-019-07827-3.

[21] Elie Cattan et al. 'Reducing latency with a continuous prediction: Effects on users' performance in direct-touch target acquisitions'. In: *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. 2015, pp. 205–214.

[22] Yeonjoo Cha and Rohae Myung. 'Extended Fitts' law for 3D pointing tasks using 3D target arrangements'. In: *International Journal of Industrial Ergonomics* 43.4 (2013), pp. 350–355.

[23] Wesley P. Chan et al. 'An Experimental Validation and Comparison of Reaching Motion Models for Unconstrained Handovers: Towards Generating Humanlike Motions for Human-Robot Handovers'. In: *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*. July 2021, pp. 356–361. DOI: 10.1109/HUMANOIDS47582.2021.9555779. (Visited on 07/09/2024).

[24] Xiaoming Chen et al. 'ImmerTai: Immersive Motion Learning in VR Environments'. In: vol. 58. Jan. 2019, pp. 416–427. DOI: 10.1016/j.jvcir.2018.11.039.

[25] Yang Chen, Xingang Zhao and Jianda Han. 'Hierarchical projection regression for online estimation of elbow joint angle using EMG signals'. In: *Neural Computing and Applications* 23.3 (2013), pp. 1129–1138.

[26] Chris Christou. 'Virtual reality in education'. In: *Affective, interactive and cognitive methods for e-learning design: creating an optimal education experience*. IGI Global, 2010, pp. 228–243.

[27] Aldrich Clarence et al. 'Unscripted Retargeting: Reach Prediction for Haptic Retargeting in Virtual Reality'. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. Mar. 2021, pp. 150–159. DOI: 10.1109/VR50410.2021.00036. (Visited on 01/09/2024).

[28] Aldrich Clarence et al. 'Unscripted Retargeting: Reach Prediction for Haptic Retargeting in Virtual Reality'. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2021, pp. 150–159.

[29] Andrew Clark and Deshendran Moodley. 'A System for a Hand Gesture-Manipulated Virtual Reality Environment'. In: *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*. SAICSIT '16. New York, NY, USA: Association for Computing Machinery, Sept. 2016, pp. 1–10. ISBN: 978-1-4503-4805-8. DOI: 10.1145/2987491.2987511.

[30] Logan D. Clark, Aakash B. Bhagat and Sara L. Riggs. 'Extending Fitts' Law in Three-Dimensional Virtual Environments with Current Low-Cost Virtual Reality Technology'. In: vol. 139. July 2020, p. 102413. DOI: 10.1016/j.ijhcs.2020.102413.

[31] Xavier Corbillon, Francesca De Simone and Gwendal Simon. '360-Degree Video Head Movement Dataset'. In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. MMSys'17. New York, NY, USA: Association for Computing Machinery, June 2017, pp. 199–204. ISBN: 978-1-4503-5002-0. DOI: 10.1145/3083187.3083215.

[32] Enric Corona et al. 'Context-Aware Human Motion Prediction'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 6992–7001.

[33] E. R. F. W. Crossman and P. J. Goodeve. 'Feedback Control of Hand-Movement and Fitts' Law'. In: vol. 35. 2. Routledge, May 1983, pp. 251–278. DOI: 10.1080/14640748308402133.

[34] Yanli Cui. ‘An Efficient Approach to Sports Rehabilitation and Outcome Prediction Using RNN-LSTM’. In: *Mobile Networks and Applications* (June 2024). ISSN: 1572-8153. DOI: 10.1007/s11036-024-02355-3.

[35] Cheng-Long Deng et al. ‘Beyond Fitts’s Law: A Three-Phase Model Predicts Movement Time to Position an Object in an Immersive 3D Virtual Environment’. In: SAGE PublicationsSage CA: Los Angeles, CA, Mar. 2019. DOI: 10.1177/0018720819831517.

[36] Nachiket Deo and Mohan M Trivedi. ‘Convolutional social pooling for vehicle trajectory prediction’. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 1468–1476.

[37] Massimiliano Di Luca and Arash Mahnan. ‘Perceptual limits of visual-haptic simultaneity in virtual reality interactions’. In: *2019 IEEE World Haptics Conference (WHC)*. IEEE. 2019, pp. 67–72.

[38] Mohammed S Elbamby et al. ‘Toward low-latency and ultra-reliable virtual reality’. In: *IEEE Network* 32.2 (2018), pp. 78–84.

[39] Hesham Elsayed et al. ‘VRSketchPen: Unconstrained Haptic Assistance for Sketching in Virtual 3D Environments’. In: *26th ACM Symposium on Virtual Reality Software and Technology*. VRST ’20. Virtual Event, Canada: Association for Computing Machinery, 2020. ISBN: 9781450376198. DOI: 10.1145/3385956.3418953. URL: https://doi.org/10.1145/3385956.3418953.

[40] Behnam Esfahbod. *SFU Motion Capture Database*. 2011. URL: https://mocap.cs.sfu.ca/ (visited on 01/03/2021).

[41] Paul M. Fitts. ‘The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement’. In: vol. 47. 6. US: American Psychological Association, 1954, pp. 381–391. DOI: 10.1037/h0055392.

[42] T Flash and N Hogan. ‘The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model’. In: vol. 5. 7. July 1985, pp. 1688–1703. DOI: 10.1523/JNEUROSCI.05-07-01688.1985.

[43] Tamar Flash and Neville Hogan. ‘The coordination of arm movements: an experimentally confirmed mathematical model’. In: *Journal of neuroscience* 5.7 (1985), pp. 1688–1703.

[44] Nadine Fligge, Joseph McIntyre and Patrick van der Smagt. ‘Minimum Jerk for Human Catching Movements in 3D’. In: *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. June 2012, pp. 581–586. DOI: 10.1109/BioRob.2012.6290265.

[45] Laura Freina and Michela Ott. ‘A literature review on immersive virtual reality in education: state of the art and perspectives’. In: *The international scientific conference elearning and software for education*. Vol. 1. 133. 2015, pp. 10–1007.

[46] Silvia Gabrielli et al. ‘Co-Design of a Virtual Reality Multiplayer Adventure Game for Adolescents With Autism Spectrum Disorder: Mixed Methods Study’. In: *JMIR Serious Games* 11.1 (Dec. 2023), e51719. DOI: 10.2196/51719.

[47] Holly C. Gagnon et al. ‘The Effect of Feedback on Estimates of Reaching Ability in Virtual Reality’. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. Mar. 2021, pp. 798–806. DOI: 10.1109/VR50410.2021.00107.

[48] Nisal Menuka Gamage et al. ‘So Predictable! Continuous 3D Hand Trajectory Prediction in Virtual Reality’. In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. UIST ’21. New York, NY, USA: Association for Computing Machinery, Oct. 2021, pp. 332–343. ISBN: 978-1-4503-8635-7. DOI: 10.1145/3472749.3474753. (Visited on 13/08/2024).

[49] Stephan Joachim Garbin et al. ‘Dataset for Eye Tracking on a Virtual Reality Platform’. In: *ACM Symposium on Eye Tracking Research and Applications*. ETRA ’20 Full Papers. New York, NY, USA: Association for Computing Machinery, June 2020, pp. 1–10. ISBN: 978-1-4503-7133-9. DOI: 10.1145/3379155.3391317.

[50] Nirit Gavish et al. ‘Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks’. In: *Interactive Learning Environments* 23.6 (2015), pp. 778–798.

[51] Amirpouya Ghasemaghaei et al. ‘Towards Better Throwing: A Comparison of Performance and Preferences Across Point of Release Mechanics in Virtual Reality’. In: *Proceedings of the 50th Graphics Interface Conference*. GI ’24. New York, NY, USA: Association for Computing Machinery, Sept. 2024, pp. 1–11. ISBN: 9798400718281. DOI: 10.1145/3670947.3670963.

[52] Partha Ghosh et al. 'Learning Human Motion Models for Long-Term Predictions'. In: *2017 International Conference on 3D Vision (3DV)*. Oct. 2017, pp. 458–466. DOI: 10.1109/3DV.2017.00059.

[53] Marcello Giordano et al. 'An Investigation of Multimodal Kinematic Template Matching for Ray Pointing Prediction for Target Selection in VR'. In: Oct. 2024. DOI: 10.1145/3702319.

[54] Shiqiu Gong and Jing Zhao. 'MMJM: A High-Precision Motion Model for Human Arm Reaching Motion'. In: *2020 3rd International Conference on Mechatronics, Robotics and Automation (ICMRA)*. Oct. 2020, pp. 59–63. DOI: 10.1109/ICMRA51221.2020.9398359. (Visited on 03/09/2024).

[55] Anand Gopalakrishnan et al. 'A Neural Temporal Model for Human Motion Prediction'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 12116–12125.

[56] Henry Griffith, Subir Biswas and Oleg Komogortsev. 'Towards Reduced Latency in Saccade Landing Position Prediction Using Velocity Profile Methods'. In: *Proceedings of the Future Technologies Conference*. Springer. 2018, pp. 79–91.

[57] Lorena Gril et al. 'A Tensor-based Regression Approach for Human Motion Prediction'. In: vol. 39. 2. 2023, pp. 481–499. DOI: 10.1002/qre.3153.

[58] Tovi Grossman and Ravin Balakrishnan. 'Pointing at Trivariate Targets in 3D Environments'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '04. New York, NY, USA: Association for Computing Machinery, Apr. 2004, pp. 447–454. ISBN: 978-1-58113-702-6. DOI: 10.1145/985692.985749.

[59] Liang-Yan Gui et al. 'Adversarial geometry-aware human motion prediction'. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 786–803.

[60] Serhan Gül et al. 'Kalman Filter-based Head Motion Prediction for Cloud-based Mixed Reality'. In: *Proceedings of the 28th ACM International Conference on Multimedia*. 2020, pp. 3632–3641.

[61] Serhan Gül et al. 'Low-latency cloud-based volumetric video streaming using head motion prediction'. In: *Proceedings of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*. 2020, pp. 27–33.

[62] Gianpaolo Gulletta, Wolfram Erlhagen and Estela Bicho. 'Human-Like Arm Motion Generation: A Review'. In: vol. 9. 4. Multidisciplinary Digital Publishing Institute, Dec. 2020, p. 102. DOI: 10.3390/robotics9040102. (Visited on 07/09/2024).

[63] Haitao Guo and Yunsick Sung. 'Movement Estimation Using Soft Sensors Based on Bi-LSTM and Two-Layer LSTM for Human Motion Capture'. In: vol. 20. 6. Multidisciplinary Digital Publishing Institute, Jan. 2020, p. 1801. DOI: 10.3390/s20061801.

[64] Markus Hahn, Lars Krüger and Christian Wöhler. '3d action recognition and long-term prediction of human motion'. In: *International Conference on Computer Vision Systems*. Springer. 2008, pp. 23–32.

[65] Christopher M. Harris and Daniel M. Wolpert. 'Signal-Dependent Noise Determines Motor Planning'. In: vol. 394. 6695. Nature Publishing Group, Aug. 1998, pp. 780–784. DOI: 10.1038/29528. (Visited on 26/10/2024).

[66] Z. Hasan. 'Optimized Movement Trajectories and Joint Stiffness in Unperturbed, Inertially Loaded Movements'. In: vol. 53. 6. Apr. 1986, pp. 373–382. DOI: 10.1007/BF00318203.

[67] Rorik Henrikson et al. 'Head-Coupled Kinematic Template Matching: A Prediction Model for Ray Pointing in VR'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–14.

[68] Alejandro Hernandez, Jurgen Gall and Francesc Moreno-Noguer. 'Human motion prediction via spatio-temporal inpainting'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 7134–7143.

[69] Neville Hogan. 'An organizing principle for a class of voluntary movements'. In: *Journal of Neuroscience* 4.11 (1984), pp. 2745–2754.

[70] John M. Hollerbach. 'The Minimum Energy Movement for a Spring Muscle Model'. In: Sept. 1977.

[71] D. E. Holmes et al. 'Using Fitt's Law to Model Arm Motion Tracked in 3D by a Leap Motion Controller for Virtual Reality Upper Arm Stroke Rehabilitation'. In: *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. June 2016, pp. 335–336. DOI: 10.1109/CBMS.2016.41.

[72] Xueshi Hou et al. 'Head and Body Motion Prediction to Enable Mobile VR Experiences with Low Latency'. In: *2019 IEEE Global Communications Conference (GLOBECOM)*. Dec. 2019, pp. 1–7. DOI: 10.1109/GLOBECOM38437.2019.9014097.

[73] Xueshi Hou et al. 'Head and Body Motion Prediction to Enable Mobile VR Experiences with Low Latency'. In: *2019 IEEE Global Communications Conference (GLOBECOM)*. Dec. 2019, pp. 1–7. DOI: 10.1109/GLOBECOM38437.2019.9014097. (Visited on 14/05/2025).

[74] Xueshi Hou et al. 'Head and body motion prediction to enable mobile VR experiences with low latency'. In: *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE. 2019, pp. 1–7.

[75] Catalin Ionescu et al. 'Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments'. In: vol. 36. 7. July 2014, pp. 1325–1339. DOI: 10.1109/TPAMI.2013.248.

[76] Neziha Jaouedi, Noureddine Boujnah and Med Salim Bouhlel. 'A New Hybrid Deep Learning Model for Human Action Recognition'. In: vol. 32. Emerging Software Systems 4. May 2020, pp. 447–453. DOI: 10.1016/j.jksuci.2019.09.004.

[77] Dilip Chakravarthy Kavarthapu and Kaushik Mitra. 'Hand Gesture Sequence Recognition Using Inertial Motion Units (IMUs)'. In: *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*. Nov. 2017, pp. 953–957. DOI: 10.1109/ACPR.2017.159.

[78] ByeoungDo Kim et al. 'Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network'. In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2017, pp. 399–404.

[79] Joohwan Kim et al. 'NVGaze: An Anatomically-Informed Dataset for Low-Latency, Near-Eye Gaze Estimation'. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300780.

[80] A. Kiruluta, M. Eizenman and S. Pasupathy. 'Predictive Head Movement Tracking Using a Kalman Filter'. In: vol. 27. 2. Apr. 1997, pp. 326–331. DOI: 10.1109/3477.558841.

[81] Oleg V. Komogortsev and Javed I. Khan. 'Eye Movement Prediction by Kalman Filter with Integrated Linear Horizontal Oculomotor Plant Mechanical Model'. In: *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*. ETRA '08. New York, NY, USA: Association for Computing Machinery, Mar. 2008, pp. 229–236. ISBN: 978-1-59593-982-1. DOI: 10.1145/1344471.1344525.

[82] Pradeep Kumar et al. 'Coupled HMM-based Multi-Sensor Data Fusion for Sign Language Recognition'. In: vol. 86. Jan. 2017, pp. 1–8. DOI: 10.1016/j.patrec.2016.12.004.

[83] Jogendra Nath Kundu, Maharshi Gor and R. Venkatesh Babu. 'BiHMP-GAN: Bidirectional 3D Human Motion Prediction GAN'. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.01 (July 2019), pp. 8553–8560. ISSN: 2374-3468. DOI: 10.1609/aaai.v33i01.33018553.

[84] Francesco Lacquaniti, Carlo Terzuolo and Paolo Viviani. 'The Law Relating the Kinematic and Figural Aspects of Drawing Movements'. In: vol. 54. 1. Oct. 1983, pp. 115–130. DOI: 10.1016/0001-6918(83)90027-6.

[85] Chiara Talignani Landi et al. 'Prediction of Human Arm Target for Robot Reaching Movements'. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Macau, China: IEEE, Nov. 2019, pp. 5950–5957. ISBN: 978-1-72814-004-9. DOI: 10.1109/IROS40897.2019.8968559. (Visited on 13/08/2024).

[86] Edward Lank, Yi-Chun Nikko Cheng and Jaime Ruiz. 'Endpoint prediction using motion kinematics'. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 2007, pp. 637–646.

[87] Rynson WH Lau and Addison Chan. 'Motion prediction for online gaming'. In: *International Workshop on Motion in Games*. Springer. 2008, pp. 104–114.

[88] Steve LaValle. *The Latent Power of Prediction*. https://developer.oculus.com/blog/the-latent-power-of-prediction/. Accessed: 2021-04-01. 2013.

[89] Bianca Lento et al. '3D-ARM-Gaze: A Public Dataset of 3D Arm Reaching Movements with Gaze Information in Virtual Reality'. In: *Scientific Data* 11.1 (Aug. 2024), p. 951. ISSN: 2052-4463. DOI: 10.1038/s41597-024-03765-4.

[90] Chen Li et al. 'Convolutional sequence to sequence model for human dynamics'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 5226–5234.

[91] Maosen Li et al. 'Symbiotic Graph Neural Networks for 3D Skeleton-Based Human Action Recognition and Motion Prediction'. In: vol. 44. 6. June 2022, pp. 3316–3333. DOI: 10.1109/TPAMI.2021.3053765.

[92] Q. L. Li, Y. Song and Z. G. Hou. 'Estimation of Lower Limb Periodic Motions from sEMG Using Least Squares Support Vector Regression'. In: vol. 41. 3. June 2015, pp. 371–388. DOI: 10.1007/s11063-014-9391-4.

[93] Qin Li et al. 'Directed Acyclic Graph Neural Network for Human Motion Prediction'. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. May 2021, pp. 3197–3204. DOI: 10.1109/ICRA48506.2021.9561540.

[94] Qinghua Li et al. 'Data Driven Models for Human Motion Prediction in Human-Robot Collaboration'. In: vol. 8. 2020, pp. 227690–227702. DOI: 10.1109/ACCESS.2020.3045994.

[95] Haoyan Liu et al. 'An FPGA-Based Upper-Limb Rehabilitation Device for Gesture Recognition and Motion Evaluation Using Multi-Task Recurrent Neural Networks'. In: *IEEE Sensors Journal* 22.4 (Feb. 2022), pp. 3605–3615. ISSN: 1558-1748. DOI: 10.1109/JSEN.2022.3141659.

[96] Hongyi Liu and Lihui Wang. 'Human motion prediction for human-robot collaboration'. In: *Journal of Manufacturing Systems* 44 (2017), pp. 287–294.

[97] Lei Liu et al. 'Comparing aimed movements in the real world and in virtual reality'. In: *2009 IEEE Virtual Reality Conference*. IEEE. 2009, pp. 219–222.

[98] Xiaoli Liu et al. 'TrajectoryCNN: A New Spatio-Temporal Feature Learning Network for Human Motion Prediction'. In: vol. 31. 6. June 2021, pp. 2133–2146. DOI: 10.1109/TCSVT.2020.3021409.

[99] Zhilan Liu et al. 'Application of Immersive Virtual-Reality-Based Puzzle Games in Elderly Patients with Post-Stroke Cognitive Impairment: A Pilot Study'. In: *Brain Sciences* 13.1 (Jan. 2023), p. 79. ISSN: 2076-3425. DOI: 10.3390/brainsci13010079.

[100] Zitong Liu et al. 'Deep Learning-based Human Motion Prediction Considering Context Awareness for Human-Robot Collaboration in Manufacturing'. In: *Procedia CIRP*. 11th CIRP Conference on Industrial Product-Service Systems 83 (Jan. 2019), pp. 272–278. ISSN: 2212-8271. DOI: 10.1016/j.procir.2019.04.080.

[101] I. Scott MacKenzie. 'Fitts' Law as a Research and Design Tool in Human-Computer Interaction'. In: vol. 7. 1. Taylor & Francis, Mar. 1992, pp. 91–139. DOI: 10.1207/s15327051hci0701_3.

[102] I. Scott MacKenzie and William Buxton. 'Extending Fitts' Law to Two-Dimensional Tasks'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. New York, NY, USA: Association for Computing Machinery, June 1992, pp. 219–226. ISBN: 978-0-89791-513-7. DOI: 10.1145/142750.142794.

[103] Y. Maeda, T. Hara and T. Arai. 'Human-Robot Cooperative Manipulation with Motion Estimation'. In: *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180)*. Vol. 4. Oct. 2001, 2240–2245 vol.4. DOI: 10.1109/IROS.2001.976403. (Visited on 05/09/2024).

[104] Naureen Mahmood et al. 'AMASS: Archive of Motion Capture As Surface Shapes'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 5442–5451.

[105] Jim Mainprice and Dmitry Berenson. 'Human-Robot Collaborative Manipulation Planning Using Early Prediction of Human Motion'. In: *2013 IEEE/RSJ International*

Conference on Intelligent Robots and Systems. Nov. 2013, pp. 299–306. DOI: 10.1109/IROS.2013.6696368.

[106] Christian Mandery et al. 'The KIT Whole-Body Human Motion Database'. In: *2015 International Conference on Advanced Robotics (ICAR)*. July 2015, pp. 329–336. DOI: 10.1109/ICAR.2015.7251476.

[107] Wei Mao et al. 'Learning trajectory dependencies for human motion prediction'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 9489–9497.

[108] Jenny Margarito et al. 'User-Independent Recognition of Sports Activities From a Single Wrist-Worn Accelerometer: A Template-Matching-Based Approach'. In: vol. 63. 4. Apr. 2016, pp. 788–796. DOI: 10.1109/TBME.2015.2471094.

[109] Julieta Martinez, Michael J Black and Javier Romero. 'On human motion prediction using recurrent neural networks'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 2891–2900.

[110] Angel Martínez-González, Michael Villamizar and Jean-Marc Odobez. 'Pose Transformers (POTR): Human Motion Prediction With Non-Autoregressive Transformers'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 2276–2284.

[111] D. E. Meyer et al. 'Optimality in Human Motor Performance: Ideal Control of Rapid Aimed Movements'. In: vol. 95. 3. July 1988, pp. 340–370. DOI: 10.1037/0033-295x.95.3.340.

[112] Hyosang Moon et al. 'Human Arm Motion Planning against a Joint Constraint'. In: *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. June 2012, pp. 401–406. DOI: 10.1109/BioRob.2012.6290923. (Visited on 05/09/2024).

[113] Nathan Moore et al. 'ALS-SimVR: Advanced Life Support Virtual Reality Training Application'. In: *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*. VRST '19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1–2. ISBN: 978-1-4503-7001-1. DOI: 10.1145/3359996.3365051.

[114] Aythami Morales et al. 'Saccade Landing Point Prediction: A Novel Approach based on Recurrent Neural Networks'. In: *Proceedings of the 2018 International Conference on Machine Learning Technologies*. 2018, pp. 1–5.

[115] *Motion Capture Systems*. http://www.optitrack.com/index.html. (Visited on 06/12/2024).

[116] Atsuo Murata and Hirokazu Iwase. 'Extending Fitts' Law to a Three-Dimensional Pointing Task'. In: vol. 20. 6. Dec. 2001, pp. 791–805. DOI: 10.1016/S0167-9457(01)00058-6.

[117] Eri Nakano et al. 'Quantitative Examinations of Internal Representations for Arm Trajectory Planning: Minimum Commanded Torque Change Model'. In: vol. 81. 5. American Physiological Society, May 1999, pp. 2140–2155. DOI: 10.1152/jn.1999.81.5.2140. (Visited on 26/10/2024).

[118] Mathieu Nancel et al. 'Next-point prediction for direct touch using finite-time derivative estimation'. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 2018, pp. 793–807.

[119] Mathieu Nancel et al. 'Next-point prediction metrics for perceived spatial errors'. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 2016, pp. 271–285.

[120] Alexander Nguyen and Biyun Xie. 'Human Arm Motion Prediction in Reaching Movements'. In: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. Aug. 2021, pp. 1117–1123. DOI: 10.1109/RO-MAN50785.2021.9515461.

[121] Diederick C Niehorster, Li Li and Markus Lappe. 'The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research'. In: *i-Perception* 8.3 (2017), p. 2041669517708205.

[122] Ren Ohmura, Yuki Kusano and Yuta Suzuki. 'Pointing Gesture Prediction Using Minimum-Jerk Model in Human-Robot Interaction'. In: *Proceedings of the Second International Conference on Human-agent Interaction*. HAI '14. New York, NY, USA: Association for Computing Machinery, Oct. 2014, pp. 181–184. ISBN: 978-1-4503-3035-0. DOI: 10.1145/2658861.2658917. (Visited on 05/09/2024).

[123] Shigeharu Ono et al. ‘A Real-time Distance Learning System for Alpine Skiing Using Virtual Reality’. In: *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. Nov. 2023, pp. 1–8. DOI: 10.1109/TALE56641.2023.10398373.

[124] Shigeharu Ono et al. ‘Learning Support and Evaluation of Weight-Shifting Skills for Novice Skiers Using Virtual Reality’. In: *Adaptive Instructional Systems*. Ed. by Robert A. Sottilare and Jessica Schwarz. Cham: Springer International Publishing, 2022, pp. 226–237. ISBN: 978-3-031-05887-5. DOI: 10.1007/978-3-031-05887-5_16.

[125] Huang Panfeng, Xu Yangsheng and Liang Bin. ‘Global Minimum-Jerk Trajectory Planning of Space Manipulator’. In: vol. 4. 4. Institute of Control, Robotics and Systems, 2006, pp. 405–413. (Visited on 05/09/2024).

[126] Georgios Papadopoulos, Alexandros Doumanoglou and Dimitrios Zarpalas. ‘VRGestures: Controller and Hand Gesture Datasets for Virtual Reality’. In: *Advances in Computer Graphics*. Ed. by Bin Sheng et al. Cham: Springer Nature Switzerland, 2024, pp. 336–350. ISBN: 978-3-031-50075-6. DOI: 10.1007/978-3-031-50075-6_26.

[127] Sangheon Park et al. ‘3D Hand Tracking Using Kalman Filter in Depth Space’. In: vol. 2012. 1. Feb. 2012, p. 36. DOI: 10.1186/1687-6180-2012-36.

[128] Daniel Eger Passos and Bernhard Jung. ‘Measuring the accuracy of inside-out tracking in XR devices using a high-precision robotic arm’. In: *International Conference on Human-Computer Interaction*. Springer. 2020, pp. 19–26.

[129] Dario Pavllo, David Grangier and Michael Auli. ‘Quaternet: A quaternion-based recurrent model for human motion’. In: *arXiv preprint arXiv:1805.06485* (2018).

[130] Dario Pavllo et al. ‘Modeling human motion with quaternion-based neural networks’. In: *International Journal of Computer Vision* (2019), pp. 1–18.

[131] Roshan Lalintha Peiris et al. ‘Thermovr: Exploring integrated thermal haptic feedback with head mounted displays’. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 5452–5456.

[132] J. L. Ponton et al. 'Combining Motion Matching and Orientation Prediction to Animate Avatars for Consumer-Grade VR Devices'. In: *Computer Graphics Forum* 41.8 (2022), pp. 107–118. ISSN: 1467-8659. DOI: 10.1111/cgf.14628.

[133] Thiago M. Porcino et al. 'Minimizing Cyber Sickness in Head Mounted Display Systems: Design Guidelines and Applications'. In: *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*. Apr. 2017, pp. 1–6. DOI: 10.1109/SeGAH.2017.7939283.

[134] Kjetil Raaen, Ragnhild Eg and Ivar Kjellmo. 'Playing with delay: An interactive VR demonstration'. In: *Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems*. 2019, pp. 19–21.

[135] Ananth N. Ramaseri Chandra, Fatima El Jamiy and Hassan Reza. 'A Systematic Survey on Cybersickness in Virtual Environments'. In: vol. 11. 4. Multidisciplinary Digital Publishing Institute, Apr. 2022, p. 51. DOI: 10.3390/computers11040051.

[136] Nimesha Ranasinghe et al. 'Season traveller: Multisensory narration for enhancing the virtual reality experience'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13.

[137] Nimesha Ranasinghe et al. 'Vocktail: A virtual cocktail for pairing digital taste, smell, and color sensations'. In: *Proceedings of the 25th ACM international conference on Multimedia*. 2017, pp. 1139–1147.

[138] Joël Randrianandrasana et al. 'Multi-user predictive rendering on remote multi-GPU clusters'. In: *2018 IEEE Fourth VR International Workshop on Collaborative Virtual Environments (3DCVE)*. IEEE. 2018, pp. 1–4.

[139] *Rendering to the Oculus Rift | Meta Horizon OS Developers*. https://developers.meta.com/horizon/docum render/. (Visited on 05/12/2024).

[140] Magnus JE Richardson and Tamar Flash. 'Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis'. In: *Journal of neuroscience* 22.18 (2002), pp. 8201–8211.

[141] Miguel Fabián Romero Rondón et al. 'Track: a Multi-Modal Deep Architecture for Head Motion Prediction in 360° Videos'. In: *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2020, pp. 2586–2590.

[142] Dairazalia Sánchez, Monica Tentori and Jesús Favela. 'Activity Recognition for the Smart Hospital'. In: vol. 23. 2. Mar. 2008, pp. 50–57. DOI: 10.1109/MIS.2008.18.

[143] Y. Sato, M. Saito and H. Koike. 'Real-Time Input of 3D Pose and Gestures of a User's Hand and Its Applications for HCI'. In: *Proceedings IEEE Virtual Reality 2001*. Mar. 2001, pp. 79–86. DOI: 10.1109/VR.2001.913773.

[144] Maria T Schultheis and Albert A Rizzo. 'The application of virtual reality technology in rehabilitation.' In: *Rehabilitation psychology* 46.3 (2001), p. 296.

[145] Valentin Schwind et al. 'The Effects of Full-Body Avatar Movement Predictions in Virtual Reality Using Neural Networks'. In: *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*. VRST '20. New York, NY, USA: Association for Computing Machinery, Nov. 2020, pp. 1–11. ISBN: 978-1-4503-7619-8. DOI: 10.1145/3385956.3418941.

[146] Emanuele L. Secco, Antonio Visioli and Giovanni Magenes. 'Minimum Jerk Motion Planning for a Prosthetic Finger'. In: vol. 21. 7. 2004, pp. 361–368. DOI: 10.1002/rob.20018.

[147] Amir Shahroudy et al. 'NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 1010–1019.

[148] Qijia Shao et al. 'N-Euro Predictor: A Neural Network Approach for Smoothing and Predicting Motion Trajectory'. In: vol. 7. 3. Sept. 2023, 120:1–120:25. DOI: 10.1145/3610884. (Visited on 07/09/2024).

[149] Adwait Sharma et al. 'GraspUI: Seamlessly Integrating Object-Centric Gestures within the Seven Phases of Grasping'. In: *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. DIS '24. New York, NY, USA: Association for Computing Machinery, July 2024, pp. 1275–1289. ISBN: 9798400705830. DOI: 10.1145/3643834.3661551.

[150] Adwait Sharma et al. 'SoloFinger: Robust Microgestures While Grasping Everyday Objects'. In: *Proceedings of the 2021 CHI Conference on Human Factors in*

Computing Systems. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–15. ISBN: 978-1-4503-8096-6. DOI: 10.1145/3411764.3445197.

[151] Yinghan Shi et al. 'Grasping 3D Objects With Virtual Hand in VR Environment'. In: *Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI '22. New York, NY, USA: Association for Computing Machinery, Jan. 2023, pp. 1–8. ISBN: 9798400700316. DOI: 10.1145/3574131.3574428.

[152] Leonid Sigal, Alexandru O. Balan and Michael J. Black. 'HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion'. In: vol. 87. 1. Mar. 2010, pp. 4–27. DOI: 10.1007/s11263-009-0273-6.

[153] B. Stenger, P.R.S. Mendonca and R. Cipolla. 'Model-Based 3D Tracking of an Articulated Hand'. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 2. Dec. 2001, pp. II–II. DOI: 10.1109/CVPR.2001.990976.

[154] Mikhail Svinin, Igor Goncharenko and Shigeyuki Hosoe. 'On the Boundary Conditions in Modeling of Human-like Reaching Movements'. In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Sept. 2008, pp. 518–525. DOI: 10.1109/IROS.2008.4651151. (Visited on 01/10/2024).

[155] Karan Taneja et al. 'A Feature-Encoded Physics-Informed Parameter Identification Neural Network for Musculoskeletal Systems'. In: vol. 144. 121006. Sept. 2022. DOI: 10.1115/1.4055238.

[156] Yuka Tashiro et al. 'GAuze-MIcrosuture-FICATION: Gamification in Microsuture Training with Real-Time Feedback'. In: *Proceedings of the Augmented Humans International Conference 2023*. AHs '23. New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 15–26. ISBN: 978-1-4503-9984-5. DOI: 10.1145/3582700.3582704.

[157] Emanuel Todorov. 'Optimality Principles in Sensorimotor Control'. In: vol. 7. 9. Nature Publishing Group, Sept. 2004, pp. 907–915. DOI: 10.1038/nn1309.

[158] Emanuel Todorov. 'Stochastic Optimal Control and Estimation Methods Adapted to the Noise Characteristics of the Sensorimotor System'. In: vol. 17. 5. May 2005, pp. 1084–1108. DOI: 10.1162/0899766053491887.

[159] Emanuel Todorov and Michael I. Jordan. 'Optimal Feedback Control as a Theory of Motor Coordination'. In: vol. 5. 11. Nature Publishing Group, Nov. 2002, pp. 1226–1235. DOI: 10.1038/nn963.

[160] Emanuel Todorov and Michael I. Jordan. 'Smoothness Maximization Along a Predefined Path Accurately Predicts the Speed Profiles of Complex Arm Movements'. In: vol. 80. 2. Aug. 1998, pp. 696–714. DOI: 10.1152/jn.1998.80.2.696. (Visited on 07/09/2024).

[161] David Tolley et al. 'Windywall: exploring creative wind simulations'. In: *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 2019, pp. 635–644.

[162] Lina Tong et al. 'HMM-Based Human Fall Detection and Prediction Method Using Tri-Axial Accelerometer'. In: vol. 13. 5. May 2013, pp. 1849–1856. DOI: 10.1109/JSEN.2013.2245231.

[163] Y. Uno, M. Kawato and R. Suzuki. 'Formation and Control of Optimal Trajectory in Human Multijoint Arm Movement'. In: vol. 61. 2. June 1989, pp. 89–101. DOI: 10.1007/BF00204593.

[164] Etienne Van Wyk and Ruth De Villiers. 'Virtual reality training applications for the mining industry'. In: *Proceedings of the 6th international conference on computer graphics, virtual reality, visualisation and interaction in Africa*. 2009, pp. 53–63.

[165] Tran Huy Vu et al. 'Smartwatch-based Early Gesture Detection 8 Trajectory Tracking for Interactive Gesture-Driven Applications'. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018), pp. 1–27. ISSN: 2474-9567. DOI: 10.1145/3191771.

[166] *Waist Circumference and BMI, 2022 | Australian Bureau of Statistics*. https://www.abs.gov.au/statistics conditions-and-risks/waist-circumference-and-bmi/latest-release. Fri, 15/12/2023 - 11:30. (Visited on 30/12/2024).

[167] Jacob Walker et al. 'The pose knows: Video forecasting by generating pose futures'. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 3332–3341.

[168] Chen Wang et al. 'Experimental Validation of Minimum-Jerk Principle in Physical Human-Robot Interaction'. In: *Neural Information Processing*. Ed. by Long Cheng, Andrew Chi Sing Leung and Seiichi Ozawa. Cham: Springer International Publishing, 2018, pp. 499–509. ISBN: 978-3-030-04239-4. DOI: 10.1007/978-3-030-04239-4_45.

[169] Min Wang et al. 'A human motion prediction algorithm for non-binding lower extremity exoskeleton'. In: *2015 IEEE International Conference on Information and Automation*. IEEE. 2015, pp. 369–374.

[170] Zhen Wang, Hengsuo Xu and Hongliang Yuan. 'Research on Design and Experience of Immersive Virtual Reality Psychological Relaxation Game Based on Image'. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 740. 1. IOP Publishing. 2020, p. 012118.

[171] Elliott Wen et al. 'VR.Net: A Real-world Dataset for Virtual Reality Motion Sickness Research'. In: vol. 30. 5. May 2024, pp. 2330–2336. DOI: 10.1109/TVCG.2024.3372044.

[172] Allen W. Wiegner and M. Margaret Wierzbicka. 'Kinematic Models and Human Elbow Flexion Movements: Quantitative Analysis'. In: vol. 88. 3. Jan. 1992, pp. 665–673. DOI: 10.1007/BF00228196. (Visited on 01/10/2024).

[173] Erwin Wu et al. 'SPinPong - Virtual Reality Table Tennis Skill Acquisition Using Visual, Haptic and Temporal Cues'. In: *IEEE Transactions on Visualization and Computer Graphics* 27.5 (May 2021), pp. 2566–2576. ISSN: 1941-0506. DOI: 10.1109/TVCG.2021.3067761.

[174] Biyun Xie, Jing Zhao and Yu Liu. 'Human-like Motion Planning for Robotic Arm System'. In: *2011 15th International Conference on Advanced Robotics (ICAR)*. June 2011, pp. 88–93. DOI: 10.1109/ICAR.2011.6088543. (Visited on 07/09/2024).

[175] Ning Xie et al. 'Explainable deep learning: A field guide for the uninitiated'. In: *arXiv preprint arXiv:2004.14545* (2020).

[176] Michele Xiloyannis et al. ‘Gaussian Process Regression for Accurate Prediction of Prosthetic Limb Movements from the Natural Kinematics of Intact Limbs’. In: *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*. Apr. 2015, pp. 659–662. DOI: 10.1109/NER.2015.7146709.

[177] Jie Xu et al. ‘Probabilistic Human Motion Prediction via A Bayesian Neural Network’. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. May 2021, pp. 3190–3196. DOI: 10.1109/ICRA48506.2021.9561665.

[178] Ruize Xu, Shengli Zhou and Wen J. Li. ‘MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition’. In: vol. 12. 5. May 2012, pp. 1166–1173. DOI: 10.1109/JSEN.2011.2166953.

[179] Tingting Xue and Hui Liu. ‘Hidden Markov Model and Its Application in Human Activity Recognition and Fall Detection: A Review’. In: *Communications, Signal Processing, and Systems*. Ed. by Qilian Liang et al. Singapore: Springer, 2022, pp. 863–869. ISBN: 978-981-19039-0-8. DOI: 10.1007/978-981-19-0390-8_108.

[180] Katsu Yamane, James J. Kuffner and Jessica K. Hodgins. ‘Synthesizing Animations of Human Manipulation Tasks’. In: *ACM SIGGRAPH 2004 Papers*. SIGGRAPH ’04. New York, NY, USA: Association for Computing Machinery, Aug. 2004, pp. 532–539. ISBN: 978-1-4503-7823-9. DOI: 10.1145/1186562.1015756. (Visited on 26/10/2024).

[181] Xingyu Yang et al. ‘Collaborative Robot Dynamics with Physical Human–Robot Interaction and Parameter Identification with PINN’. In: vol. 189. Nov. 2023, p. 105439. DOI: 10.1016/j.mechmachtheory.2023.105439.

[182] Michael Yates, Arpad Kelemen and Cecilia Sik Lanyi. ‘Virtual reality gaming in the rehabilitation of the upper extremities post-stroke’. In: *Brain injury* 30.7 (2016), pp. 855–863.

[183] HyeonBeom Yi, Jiwoo Hong and Woohun Lee. ‘DexController : Hand-Held Controller Recognizing Grasp-Pose and Grasp-Force in Virtual Reality Defense Game’. In: *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*. VRST ’19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1–2. ISBN: 978-1-4503-7001-1. DOI: 10.1145/3359996.3365031.

[184] Soojeong Yoo and Judy Kay. 'VRun: Running-in-Place Virtual Reality Exergame'. In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. OzCHI '16. New York, NY, USA: Association for Computing Machinery, Nov. 2016, pp. 562–566. ISBN: 978-1-4503-4618-4. DOI: 10.1145/3010915.3010987.

[185] Soojeong Yoo, Sunkyung Kim and Youngho Lee. 'Learning by Doing: Evaluation of an Educational VR Application for the Care of Schizophrenic Patients'. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–6. ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3382851.

[186] Soojeong Yoo and Callum Parker. 'Towards a Metaverse in Health Informatics: 3D Visualisation of Physical Activity from VR Gaming'. In: *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Mar. 2023, pp. 136–137. DOI: 10.1109/VRW58643.2023.00034.

[187] Shuo Yu, Hsinchun Chen and Randall A. Brown. 'Hidden Markov Model-Based Fall Detection With Motion Sensor Orientation Calibration: A Case for Real-Life Home Monitoring'. In: vol. 22. 6. Nov. 2018, pp. 1847–1853. DOI: 10.1109/JBHI.2017.2782079.

[188] Jie Zhang et al. 'Physics-Informed Deep Learning for Musculoskeletal Modeling: Predicting Muscle Forces and Joint Kinematics From Surface EMG'. In: vol. 31. 2023, pp. 484–493. DOI: 10.1109/TNSRE.2022.3226860.

[189] Jihai Zhang et al. 'Chinese Sign Language Recognition with Adaptive HMM'. In: *2016 IEEE International Conference on Multimedia and Expo (ICME)*. July 2016, pp. 1–6. DOI: 10.1109/ICME.2016.7552950.

[190] Jing Zhao et al. 'Human Arm Motion Prediction in Human-Robot Interaction Based on a Modified Minimum Jerk Model'. In: vol. 35. 3-4. Taylor & Francis, Feb. 2021, pp. 205–218. DOI: 10.1080/01691864.2020.1840432. (Visited on 05/09/2024).

[191] Shulin Zhao et al. 'Déjà View: Spatio-Temporal Compute Reuse for 'Energy-Efficient 360° VR Video Streaming'. In: *2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)*. IEEE. 2020, pp. 241–253.

[192] Hongyu Zhou et al. 'CoplayingVR: Understanding User Experience in Shared Control in Virtual Reality'. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8.3 (Sept. 2024), 148:1–148:25. DOI: 10.1145/3678508.

[193] Tim Zindulka, Myroslav Bachynskyi and Jörg Müller. 'Performance and Experience of Throwing in Virtual Reality'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–8. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376639.

Appendix A

In this appendix, we include the original documents used during the user study discussed in Chapter 3. These documents include;

- Participant Information Statement
- Participant Consent form
- Background Information Sheet
- Task Questionnaire
- Post Study Questionnaire
- User Study Script

A.1 Participant Information Statement

THE UNIVERSITY OF SYDNEY

School of Computer Science
Faculty of Engineering

ABN 15 211 513 464

Dr. Anusha Withana
Lecturer, School of Computer Science

Room 309
J12 Computer Science Building
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 8627 7056
Facsimile: +61 2 9351 3838
Email: anusha.withana@sydney.edu.au
Web: http://www.sydney.edu.au/

Arm Movement Modelling and Prediction
PARTICIPANT INFORMATION STATEMENT

(1) What is this study about?

You are invited to take part in a research study that investigates the use of a movement sensing and tracking systems for arm movement modelling and prediction. Physical movement sensing technologies are used to track human activities and used in variety of applications such as wellbeing (e.g. your fitness app), healthcare (e.g. rehabilitation), age care, etc. In this study, we will explore how new a sensing and tracking systems can be used for modelling arm movement. The aim is to gain further understanding into how well arm movements can be modelled and predicted, and how consistently it can work with individual differences between people.

You have been invited to participate in this study to take part in a user study and provide information on your experience. This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the research. Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.
Participation in this research study is voluntary.

By giving your consent to take part in this study you are telling us that you:

- ✓ Understand what you have read.
- ✓ Agree to take part in the research study as outlined below.
- ✓ Agree to the use of your personal information as described.

You will be given a copy of this Participant Information Statement to keep.

(2) Who is running the study?

The study is being carried out by the following researchers:

- Dr. Anusha Withana (Lecturer), The School of Computer Science
- Nisal Menuka (PhD Student), The School of Computer Science

The study that is being conducted will form the basis for research papers in the area of wearable computing and human computer interaction.

(3) What will the study involve for me?

You will be asked to wear number of sensors in your body to collect data about your arms movements and complete four activities while wearing a Virtual Reality (VR) headset. The tasks will include four VR games and activities which include several physical tasks including moving your arms and legs. This will allow us to record your movements during the activity and build an arm movement model with the recorded data. Some other technologies such as smartphone may be used to assist the study. A brief interview will take place after you have used the sensing applications to obtain feedback on your experience. Study will be audio and video recorded. These recordings will be used to document information only and will not be shared or published in any form. Some demographic and anthropometric information will also be asked including age, weight, height and arm length.
Personal information will not be shared or published in any form.

(4) How much of my time will the study take?

The study will take up to 60 minutes to complete. Trialling the technology will take up to 10 minutes. And post interview discussion will take up to 10 minutes. One of the researchers will explain the purpose of the study and the questions and will coordinate the discussion.

(5) Who can take part in the study?

Anyone that is 18 years or older, healthy and injury-free are welcome to participate.

(6) Do I have to be in the study? Can I withdraw from the study once I've started?

Being in this study is completely voluntary and you do not have to take part. Your decision whether to participate will not affect your current or future relationship with the researchers or anyone else at the University of Sydney.

If you decide to take part in the study and then change your mind later, you are free to withdraw at any time. You can do this by informing the researcher carrying out the study that you no longer wish to participate.

You are free to stop the experiment or discussion at any time. Unless you say that you want us to keep them, any recordings will be erased and the information you have provided will not be included in the study results. You may also refuse to answer any questions that you do not wish to answer during the interview.
You are free to opt out from audio/video recordings of the study.

(7) Are there any risks or costs associated with being in the study?

:

Our system is built with skin friendly adhesives. It is documented that potential exists for people to experience allergies to these materials. Not everyone experiences this, but if you experience any discomfort or develop any symptoms, you can quit the study at any time. Furthermore, if you are aware of any allergies to above materials, you should not engage in this study

It is well documented that potential exists for people to experience motion sickness while wearing a head mounted display to view a virtual information. Not everyone experiences this, but if it does occur you may have a break at any time and quit the study.

Otherwise, aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study. However, if you feel any discomfort or fatigue during the study, you are free to stop the study or request for a break. You can do this by informing the researcher carrying out the study that you no longer wish to participate.

(8) Are there any benefits associated with being in the study?

We cannot guarantee that you will receive any direct benefits from being in the study.

(9) What will happen to information about me that is collected during the study?

Some personal information will be asked of you, for example age group, gender, and some anthropometric information (weigh, height, etc.).

Video and audio recordings will be used for interpretation and analysis only. A transcription service may be employed to transcribe interviews; however, your personal information will be kept confidential.

All the information will be stored without reference to identifiable information such as your name.

All aspects of the study, including the results, will be strictly confidential and only the researchers will have access to information on participants. The results of the study will be published in student theses, journal publications and conference papers. Material collected in the course of the research will be kept stored securely in an online storage managed by The University of Sydney.

By providing your consent, you are agreeing to us collecting personal information about you for the purposes of this research study. Your information will only be used for the purposes outlined in this Participant Information Statement, unless you consent otherwise.

Your information will be stored securely, and your identity/information will be kept strictly confidential, except as required by law. Study findings may be published, but you will not be individually identifiable in these publications.

We will keep the information we collect for this study, and we may use it in future projects. By providing your consent you are allowing us to use your information in future projects. We don't know at this stage what these other projects will involve. We will seek ethical approval before using the information in these future projects.

(10) Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

(11) What if I would like further information about the study?

When you have read this information, Dr. Anusha Withana will be available to discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact Nisal Menuka Gamage (email: nisal.gamage@sydney.edu.au, mobile: [Redaction]).

(12) Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by indicating on the participant consent form. This feedback will be in the form of one-page summary. You will receive this feedback after the study is finished.

(13) What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney *2019/553* As part of this process, we have agreed to carry out the study according to the *National Statement on Ethical Conduct in Human Research (2007).* This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

This information sheet is for you to keep

A.2 Participant Consent Form

THE UNIVERSITY OF SYDNEY

School of Computer Science
Faculty of Engineering

ABN 15 211 513 464

Dr. Anusha Withana
Lecturer, School of Computer Science

Room 309
J12 Computer Science Building
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 8627 7056
Facsimile: +61 2 9351 3838
Email: anusha.withana@sydney.edu.au
Web: http://www.sydney.edu.au/

Arm Movement Modelling and Prediction
PARTICIPANT CONSENT FORM

I, ... [PRINT NAME], agree to take part in this research study.

In giving my consent I state that:

- I understand the purpose of the study, what I will be asked to do, and any risks/benefits involved.
- I have read the Participant Information Statement and have been able to discuss my involvement in the study with the researchers if I wished to do so.
- The researchers have answered any questions that I had about the study and I am happy with the answers.
- I understand that being in this study is completely voluntary and I do not have to take part. My decision whether to be in the study will not affect my relationship with the researchers or anyone else at the University of Sydney now or in the future.
- I understand that I can withdraw from the study at any time.
- I understand that I may stop the interview at any time if I do not wish to continue, and that unless I indicate otherwise any recordings will then be erased and the information provided will not be included in the study. I also understand that I may refuse to answer any questions I don't wish to answer.
- I understand that personal information about me that is collected over the course of this project will be stored securely and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.
- I understand that the results of this study may be published, and that publications will not contain my name or any identifiable information about me.
- As part of the study, audio, video and photograph recordings will be collected. These recording will not include your face or any identifying material. If you wish to, you can opt-out from these recordings.

I consent to:

- **Audio-recording** YES ☐ NO ☐
- **Video-recording** YES ☐ NO ☐
- **Photographs** YES ☐ NO ☐
- **Being contacted about future studies** YES ☐ NO ☐

I would like to review my interview transcripts YES ☐ NO ☐

I would like to receive feedback about the overall results of this study YES ☐ NO ☐

If you answered **YES**, please indicate your preferred form of feedback and address:

☐ Postal: ____

____

☐ Email: ____

....

Signature

....

PRINT name

....

Date

A.3 Background Information Sheet

THE UNIVERSITY OF SYDNEY

BACKGROUND INFORMATION SHEET

Project title: Arm Movement Modelling and Prediction

1. DOCUMENT SUMMARY

As part of the study, some background data will be recorded for comparison and analysis purposes. This will include demographics data such as age (year only), gender, etc. Also, depending on the particular study, some physiological information (e.g. height, length and measurements of limbs, etc.) will be collected. All this information will be collected before the main quantitative study.

2. BACKGROUND DATA COLLECTED

Part I: Background Data

1. Age?
   (Number of years only)

2. Gender identity?
   (Male/Female/Other)

3. Have you used a VR headset before this experiment?
   (never / 1 time/ 2-5 times / 5-10 times / more than 10 times/ Prefer not to answer)

Part II: Anthropometric Data

- Height?
  (In cm)
- Weight?
  (In Kg)
- Length of arm?
  (In cm)
- Dominant hand? (left/right)

A.4 Task Questionnaire

THE UNIVERSITY OF SYDNEY

TASK QUESTIONNAIRE

Project title: Arm Movement Modelling and Prediction

Task 1

On a scale of 1-7, please indicate your experience regarding completed activity.

		1 strongly disagree	2	3	4 neutral	5	6	7 strongly agree
1	The activity was physically demanding							
2	The activity was mentally demanding							
3	The activity contained many arm movements in straight lines							
4	The activity contained many arm movements in curved lines							
5	The activity contained fast movements							
6	The activity contained slow movements							
7	I am satisfied with how I completed the activity							
8	I enjoyed completing the activity							
9	It was comfortable to wear the VR headset during the activity							
10	It was comfortable to wear the sensor setup during the activity							

THE UNIVERSITY OF SYDNEY

Task 2

On a scale of 1-7, please indicate your experience regarding completed activity.

		1 strongly disagree	2	3	4 neutral	5	6	7 strongly agree
1	The activity was physically demanding							
2	The activity was mentally demanding							
3	The activity contained many arm movements in straight lines							
4	The activity contained many arm movements in curved lines							
5	The activity contained fast movements							
6	The activity contained slow movements							
7	I am satisfied with how I completed the activity							
8	I enjoyed completing the activity							
9	It was comfortable to wear the VR headset during the activity							
10	It was comfortable to wear the sensor setup during the activity							

THE UNIVERSITY OF SYDNEY

Task 3

On a scale of 1-7, please indicate your experience regarding completed activity.

		1 strongly disagree	2	3	4 neutral	5	6	7 strongly agree
1	The activity was physically demanding							
2	The activity was mentally demanding							
3	The activity contained many arm movements in straight lines							
4	The activity contained many arm movements in curved lines							
5	The activity contained fast movements							
6	The activity contained slow movements							
7	I am satisfied with how I completed the activity							
8	I enjoyed completing the activity							
9	It was comfortable to wear the VR headset during the activity							
10	It was comfortable to wear the sensor setup during the activity							

THE UNIVERSITY OF SYDNEY

Task 4

On a scale of 1-7, please indicate your experience regarding completed activity.

		1 strongly disagree	2	3	4 neutral	5	6	7 strongly agree
1	The activity was physically demanding							
2	The activity was mentally demanding							
3	The activity contained many arm movements in straight lines							
4	The activity contained many arm movements in curved lines							
5	The activity contained fast movements							
6	The activity contained slow movements							
7	I am satisfied with how I completed the activity							
8	I enjoyed completing the activity							
9	It was comfortable to wear the VR headset during the activity							
10	It was comfortable to wear the sensor setup during the activity							

A.5 Post Study Questionnaire

THE UNIVERSITY OF SYDNEY

POST STUDY QUESTIONNAIRE

Project title: Arm Movement Modelling and Prediction

1. Which activity;
   a. Was the most physically demanding?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

   b. Was the most mentally demanding?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

   c. Contained most movements in straight lines?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

   d. Contained most movements in curvy lines?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

   e. Contained fastest movements?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

   f. Contained slowest movements?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

   g. Were you most satisfied with completing?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

   h. Was the most enjoyable?
      1. Boxing 2. Beatsaber 3. Table Tennis 4. Dancing

2. What are the two activities that you felt had most similar movements?

3. What are the two activities that you felt had most different movements?

4. Do you have any comments or suggestions?

A.6 User Study Script

Project title: Arm Movement Modelling and Prediction

User study - Script

Thank you for willing to participate in this research study. Being in this study is completely voluntary and you do not have to take part. Your decision whether to participate will not affect your current or future relationship with the researchers or anyone else at the University of Sydney.

If you decide to take part in the study and then change your mind later, you are free to withdraw at any time. You can do this by informing the researcher carrying out the study that you no longer wish to participate.

You are free to stop the experiment or discussion at any time. Unless you say that you want us to keep them, any recordings will be erased and the information you have provided will not be included in the study results. You may also refuse to answer any questions that you do not wish to answer during the interview.

You are free to opt out from audio/video recordings of the study.

Introduction

In this study, our goal is to investigate how arm movement during VR activities can be modelled and predicted.

You will be asked to complete several VR activities while wearing a VR headset. Your movement data will be recorded through IMU sensors and Optitrack motion tracking software along with what you see in the VR headset will be recorded. Additionally, whole session will be video recorded.

Background Data Collection

Please complete the following questionnaire. (give background questionnaire)

If you have any questions, please ask. If you do not know your arm length, the researcher will help you to measure it.

Experiment setup

You will be asked to wear number of sensors in your body to collect data about your arms movements and complete four activities while wearing a Virtual Reality (VR) headset. The researcher who is conducting the study will make sure the sensors are attached properly and will assist in wearing the sensors. If you feel uncomfortable, please not hesitate to raise the issue with the researcher.

Checklist

1. 6 IMU sensors
    a. Right Arm and Left arm
        i. Shoulder
        ii. Elbow
        iii. Wrist

Make sure the orientations and locations are consistent. Make sure the bands are tight but comfortable.

2. 4 trackers
    a. Back – between shoulders
    b. Waist – front
    c. Left foot
    d. Right foot
3. Processing software – check incoming data
4. Optitrack – Check whether all markers are properly displayed

Calibration
First, we will record some movement data without the VR headset.
Please look at the screen in front of you and repeat the movement after each recording is finished.

VR Setup
Next, we will record data of some tasks with VR headset.

Make sure the VR headset is connected to the pc and the screen can be seen.
Ready the VR Screen Recording.
Help the participant to wear the headset. Make sure it is tight and comfortable.

VR Setup
For each VR task do the following.

Please follow the instructions given by the researcher to navigate to the intended application.
Give instructions to do the task.

After each task is finished, give the task questionnaire.

Post Study
Give post study questionnaire.

Thank you for participating in this user study.