BEHAVIOR SHIFT TO ALTERED PHYSICS LAW OF STANDING: A PREDICTION FROM THE REINFORCEMENT LEARNING CONTROLLER OF POSTURAL CONTROL

by

Jiyu Wang

B.KIN., The University of British Columbia, 2019

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Kinesiology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Behavior shift to altered physics law of standing: a prediction from the Reinforcement Learning Controller of postural control

submitted by Jiyu Wang in partial fulfillment of the requirements for

the degree of Master of Science

in Kinesiology

Examining Committee:

Dr. Jean-Sébastien Blouin, School of Kinesiology, UBC
Supervisor

Dr. Romeo Chua, School of Kinesiology, UBC
Supervisory Committee Member

Dr. Calvin Kuo, School of Biomedical Engineering, UBC
Supervisory Committee Member

Abstract

A central question to our understanding of postural control is the overall goal of standing balance. Current opinions of the topic diverge: researchers have argued that minimization of movement variability or overall exerted torque could be potential goals of balance control.

The purposes of the thesis were to (1) model standing balance control using the Markov Decision Process framework and identify best parameter combinations that represent the physiological characteristics of standing and (2) probe the goal of standing using computational simulations and a custom-designed robotic balancing platform with altered standing balance dynamics.

Human standing balance in the anterior-posterior direction was modeled using the Markov Decision Process framework, and the Q-learning algorithm was applied to solve the control problem. Performance of the model was evaluated by comparing the range, root mean square, mean power frequency and 99% power bandwidth of the simulated center of mass data with empirical evidence. In the experimental study, participants (n = 3) were asked to balance on the robotic balancing platform during perturbations in which torque bias terms were added to the load-stiffness relationship of standing. The exerted torque and body angle were recorded and analyzed.

The simulated quiet standing behavior from the Markov Decision Process model resembled the frequency characteristics of standing with larger variability in the time series analysis. In the experimental study, two participants balanced at a more backward (forward) angle when positive (negative) torque bias terms were added, which matched the predictions from my hypothesis. However, the size of the angle shifts differed from the hypothesis and they

did not maintain the same torque level as my hypothesis which predicts participants would maintain their torque.

In conclusion, the Markov Decision Process model generated behavior close to human balance control given specific parameters. While the direction of body angle shifts observed in the human data and Markov Decision Process model simulated data matched the prediction from my hypothesis of torque minimization, the experimental results did not fully support the statement that people always seek to maintain their torque levels during standing.

Lay Summary

Standing with two legs is a complex task, although often taken for granted. We need to sense our motion and coordinate our muscles in order to maintain an upright posture without falling. What are we trying to do when we coordinate our muscles? This question probes the underlying goal of standing and has raised many debates and plausible assumptions. Commonly assumed goals include minimizing deviation from the upright posture and minimizing joint torques (or effort). Knowing the goal of standing is crucial for understanding behavioral adaption in novel situations. Therefore, the purpose of this thesis is to provide insights onto the goal of standing using a computational framework and a novel experimental paradigm.

Preface

The data from the current thesis were collected in the Sensorimotor Physiology Laboratory at the University of British Columbia, Vancouver campus. The protocol of this thesis was approved by the University of British Columbia Human Research Ethics Committee and the ethics approval number is H18-03702.

I was the lead investigator of the projects in this thesis where I was responsible for all major areas of concept formation, data collection and analysis and thesis composition, with guidance from the supervisory committee composed of Dr. Jean-Sébastien Blouin, Dr. Calvin Kuo and Dr. Romeo Chua. Dr. Jean-Sébastien Blouin and Dr. Calvin Kuo were involved throughout the projects in concept formation and thesis document edits. Candy Xiyao Liu, Alex Liu, Cassa Courtney, Randy Gao, and Jarrett Lee contributed to computational data collection.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AP: Anterior-Posterior
CNS: Central Nervous System
COM: Center of Mass
MDP: Markov Decision Process
ML: Medial-Lateral
MPF: Median power frequency
RL: Reinforcement learning: RL
RMS: Root Mean Square

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Jean-Sébastien Blouin for all the support he has given me over the last few years. My first exposure to research came in the summer of 2018 when I started as an undergraduate research assistant supervised by Dr. Blouin. He introduced me to how research can be rigorous and creative like an art at the same time. He has supported my learning in many ways, by letting me develop independent problem-solving abilities as well as providing guidance to all aspects of research. I would also like to thank Dr. Calvin Kuo, who inspires my interest in computational modeling to tackle physiology issues. He has gone above and beyond his role as a committee member and provided invaluable feedback and support throughout every phase of my thesis. Lastly, I would like to thank Dr. Romeo Chua who has always provided detailed feedback on my projects. Dr. Chua always asked the right questions and brought new perspectives into the approaches in my thesis, broadening the scope and understanding of my projects.

I would also like to thank my fellow lab members in the Sensorimotor Physiology Lab and Human Motion Biomechanics Lab. The members of these labs were always willing to provide help and support when I came across problems in my projects. I am fortunate to have the opportunity to work with all of them.

I would like to acknowledge that this thesis was enabled in part by remote cloud support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.com).

Lastly, I would like to thank my family and friends for their ongoing support throughout this degree.

Dedication

I dedicate this thesis to my friends and family.

1 Introduction

Humans evolved by adapting to the changing environmental constraints in order to survive. We went through a long period to learn to stand and move around with two legs. As physiologists and engineers, we are interested in understanding how and why behavioral changes take place in response to the environmental or structural changes. Advancements in robotics and computer simulations allow us to reveal the mechanisms of balance adaptations by combining predictions from classic computational frameworks with manipulations of balance tasks to induce short-term changes in balancing behavior.

Standing balance, although often taken for granted, is a complex neural process that has been studied extensively by researchers in the field of motor control. When standing upright, the central nervous system (CNS) receives and processes multiple noisy sensory inputs and then generates appropriate balance-correcting motor commands based on those inputs. Understanding the physiological mechanisms underlying how humans maintain their upright posture is crucial for researchers to gain more insights on how sensory inputs are integrated and processed. Computational models are common approaches used by physiologists and engineers to represent standing balance as understanding a physiological system or process is essential in "designing significant experiments and correctly interpreting their results" (Robinson, 2011). The complexity in computational models is often difficult or even impossible to achieve in experimental diagrams. Therefore, a properly validated model that accurately predicts motor outcomes can be used to explain properties of neural control dynamics from empirical data, establishing a link from empirical findings to more abstract concepts (Popovych et al., 2019). As such, computational models are powerful tools for understanding general principles governing our brains' neural complexity.

Several researchers have applied a Markov Decision Process (MDP) framework and reinforcement learning (RL) models to represent human motor control processes (Jamali et al., 2018; Michimoto et al., 2016; Selinger et al., 2019). MDP is a mathematical framework typically used to model stochastic decision-making processes and RL represents a collection of computational methods that generate solutions for MDP frameworks. Given its potential for modeling motor control tasks (covered in details in Chapter 2), my thesis work was dedicated partly to modeling human standing balance using the MDP framework.

The MDP framework enforces two concepts closely related to physiology: reward functions and exploration. Reward functions evaluate the decisions made and they are the key element in MDP models as they shape the learned behavior. Reward functions in MDP frameworks and cost functions in traditional feedback controllers work in similar ways as reward functions encourage certain behavior whereas cost functions penalize undesired behavior patterns. Both of them reflect the goal of motor control tasks. In the context of postural control, there is no consensus regarding the goal of standing. Commonly assumed goals of standing include minimizing movement variability and joint torques (Kiemel et al., 2011; Peterka, 2002). Therefore, the experimental approach in the thesis aims to provide some insights onto this.

Another classic topic of interest in RL theories is the trade-off between exploration and exploitation, which represents a trade-off between exploring the environment and exploiting the current knowledge of the said environment at the same time. Constant exploration of the system conflicts the goal of minimizing movement variability, as exploratory behavior can sometimes raise variability. However, recent research findings suggested movement variability can be beneficial in motor control and learning tasks and should not be simply treated as unwanted noise (Carpenter et al., 2010; Wu et al., 2014). Traditional feedback-based controllers mostly aim

to reduce or minimize movement variability, without addressing its role in learning (Asai et al., 2009; Peterka, 2002; Van Der Kooij & Peterka, 2011). RL models, on the other hand, can be used to represent the dual roles of variability by maximizing a reward function that minimizes movement variability while enforcing the exploratory behavior throughout learning.

In summary, I seek to provide a computational framework to represent how humans learn to balance physiologically in this thesis. The experimental approach in this thesis will provide insights on the goal of standing. The comparison of the experimental and computational results will evaluate whether MDP framework can represent how humans learn and adapt to novel situations.

In Chapter 2, I provide an introduction to basic principles and mechanisms of standing balance control and then present and evaluate current computational approaches. In Chapter 3, I describe the methodology of the computational and empirical approaches in this thesis. In Chapter 4, I present the results obtained from both the computational and experimental study. In Chapter 5, I discuss the indications of the obtained results and how they relate to the current theories regarding the control of standing balance. Limitations and future directions will also be discussed here.

2 Literature Review

2.1 Biomechanics and physiology standing balance

2.1.1 Biomechanics of standing

To balance successfully without falling, the body's center of mass (CoM) needs to be maintained within the base of support. During standing, a small deviation from the upright posture will result in a rotatory torque that moves the body away from the center because of the gravity acting on the body. The inherent stiffness from the lower limbs is not sufficient to counteract the gravitational influence (Woodhull et al., 1985), and therefore, a corrective torque contributed mainly by the lower limb muscles will be needed to maintain balance. When the human body leans forward, activation of plantar-flexor muscles (i.e., soleus and Gastrocnemius) will be needed to generate a plantar-flexing torque that counteract the gravitational torque acted on the body.

The biomechanics of standing in the anterior-posterior (AP) direction is often represented using a single-link inverted pendulum that pivots and sways around the ankle joint in the AP direction (R. C. Fitzpatrick et al., 1992; Peterka, 2002). A second-order differential equation can describe the dynamics of a single-link inverted pendulum that simulates standing balance:

$$I\ddot{\theta} + b\dot{\theta} - mgLsin\theta = T \quad (1)$$

Here, θ is the angular position of body's CoM relative to the ankle joint, m is the body mass, L is the length from the CoM to the ankle joint, T is the torque at the ankle joint, I is the mass moment of inertia about the ankle joint and b is the viscoelasticity parameter that represents damping effect from the soft tissues in the lower limbs. When humans stand with minimal body sway, the angular velocity $\dot{\theta}$ and acceleration $\ddot{\theta}$ will be close to zero, and the body will remain a quasi-static posture. In a quasi-static posture, the ankle torque is approximately proportional to

the body angle (Equation 2), and this relationship reflects the load-stiffness of standing (see Figure 2 for quasi-static load stiffness curve).

$$T = -mgL\theta \quad (2)$$

Balance in the medial-lateral (ML) direction is more complicated than in the AP direction, as the control of balance in the ML direction involves more degrees of freedom (i.e., ankle joints, hip joints and upper body). Because of the complexity in ML balance control, both the computational and experimental approaches in this thesis only focused on behavior in the AP direction.

2.1.2 Physiological characteristics of standing

Humans don't stand perfectly still and we are always swaying. To quantify the swaying behavior, researchers observed and recorded the movement of CoM during standing (Hasan et al., 1996; Soames & Atha, 1982). Table 1 presents time and frequency characteristics of CoM movements recorded in a study conducted by Hasan et al., (1996). In quiet standing, postural sway demonstrated low-frequency oscillatory behavior with median power frequency (MPF) lower than 1 Hz and the majority of the frequency content (i.e., 99%) below 5 Hz (Hasan et al., 1996; McClenaghan et al., 1996; Soames & Atha, 1982).

	Mean + standard deviation	95% confidence interval
Range (mm)	10.33 ± 3.90	7.21 - 13.45
RMS (mm)	2.69 ± 3.19	0.14 – 5.24
MPF (Hz)	0.65 ± 0.12	0.5540 – 0.7460
99% power bandwidth (Hz)	3.15 ± 0.85	2.47 – 3.83

Table 1: Mean, standard deviation and 95% confidence intervals of the range, RMS, median power frequency and 99% power bandwidth of the CoM movement calculated from the findings of Hasan et al., 1996.

Detecting changes in body motion (i.e., both low-frequency sways and high-frequency oscillations) is essential for maintaining a steady standing posture (Fitzpatrick & McCloskey, 1994). Many sensory systems contribute to sensing body orientation and velocity and muscle activation levels during standing, such as vestibular, visual, auditory and proprioceptive systems. The inputs from those sensory systems contains delayed, incomplete and noisy information that are then further processed by the CNS to provide estimates of where the body is.

2.1.3 Control mechanisms of standing

The exact mechanism of how people maintain balance is poorly understood. How people control their balance precisely has always drawn interest from researchers because standing poses a fundamental question of how humans sense their own movements and then coordinate motor commands based on those estimates. The nature of active control mechanisms responsible for stabilizing the body upright is still under debate. Commonly proposed control mechanisms include continuous feedback control and intermittent feedback control. Intermittent control refers to serial production of ballistic muscle activations and continuous control refers to gradient activations of muscles (Loram & Lakie, 2002). It has been shown that intermittent control mechanism was effective for overcoming negative influence imposed by the delay in the system as well as sensory noises (Loram et al., 2011; Yoshikawa et al., 2016). Some physiology evidence suggested that intermittent control existed in the human motor system (Gawthrop et al., 2011; Huryn et al., 2014; Loram et al., 2011), which led to the proposal of intermittent controllers to model motor control tasks. In 2011 Loram and colleagues performed an experiment that asked the participants to balance a virtual joystick with visual feedback using either continuous contact (continuous control) or intermittent taps (intermittent control). Their results showed that the participants were able to complete the task with both controls at ease.

However, the performance (i.e., minimizing the position and velocity) of task completed with intermittent control was superior to continuous control. Yoshikawa and colleagues (2016) designed and compared performance of continuous and intermittent feedback controllers to explain the experimental data of human stick balancing. Their results also demonstrated that the intermittent controller was able to characterize human stick-balancing behavior, and the intermittent controller's fitted results were more accurate than the continuous controller. Asai and colleagues (2009) compared the performance of continuous and intermittent controllers in human postural control in standing. Simulation analysis showed that the region in the feedback space where stability can be achieved was larger for the intermittent controller than continuous controller, indicating superior robustness of the intermittent controller.

2.1.4 Goal of standing

Given the physiological observations of human standing balance, researchers have aimed to identify the underlying goal that drives the balance behavior. One of the commonly assumed goals in standing balance is to minimize deviation from a fixed set-point or minimizing movement variability (Mauer & Perterka, 2005). With this assumption, postural sway is considered unwanted noise caused by internal and external perturbations such as breathing, heart beats and air flow. Other assumptions include minimization of muscle activation (Kiemel et al., 2011; Welch & Ting, 2008). Results from a study conducted by Kiemel and colleagues (2011) compared the experimentally acquired feedback with optimal feedback using different cost functions (i.e., minimizing deviation from the upright posture, center of pressure variability, and muscle activation) and they reported that the feedback acquired from the experiment was the closest to the optimal feedback that minimizes muscle activation.

2.2 Computational models of standing balance

Given the experimental biomechanical and physiological observations, researchers have attempted to use computational approaches to explain the observed balancing behavior and understand the neural mechanisms underlying its control (Mauer & Perterka, 2005; Michimoto et al., 2016; Asai et al., 2009; Forbes et al., 2018). In the following section, I will present two categories of computational approaches that have been used to model human standing balance.

2.2.1 Traditional feedback-based controllers

One of the most widely used models within the context of standing balance is the optimal feedback-based model. Models that are based on optimality theories aim to minimize cost functions, which reflect the goal of a task, and have provided explanation to physiological phenomena in motor control tasks, including postural sway during quiet standing and balance responses to external perturbations (Hidenori & Jiang, 2006; Peterka, 2002; Van Der Kooij et al., 2001). These models typically have two components: an abstract representation of the human biomechanics (the plant) and a controller that produces optimal actions based on delayed feedback. The dynamics of the plant can be simplified using linear differential equations. Common feedback-based controllers include Proportional-Derivative-Integral (PID) controllers (Asai et al., 2009; Hidenori & Jiang, 2006) and Linear Quadratic Regulator (LQR) controllers (Kuo, 1995; Lockhart & Ting, 2007). PID control refers to a close-loop mechanism that utilizes feedback based on error values in the form of proportional (P), derivative (D) and integral (I) terms. LQR control refers to a control mechanism in which the system dynamics is linear, and the cost functions are defined in quadratic terms. In the context of human standing, PID controllers generate corrective torques that are proportional to (P), proportional to the time derivative (D) of and proportional to the time domain integral (I) of the error between the upright

posture (i.e., 0 degrees body angle) and the actual body angle (Hidenori & Jiang, 2006; Maurer & Peterka, 2005). LQR controllers minimize quadratic cost functions in which one term penalizes deviations from the upright posture similar to PID models and another term penalizes muscle activation (Li et al., 2012; Lockhart & Ting, 2007; Van Der Kooij & Peterka, 2011; Welch & Ting, 2008). Feedback-based controllers have been shown to account for the low-frequency body sway observed experimentally in quiet standing (Mauer & Perterka, 2005; Welch & Ting, 2007). However, those low-frequency oscillations in their simulations were mainly driven by noises added to the corrective torques generated by the controllers instead of specific control actions. In the PID controller designed by Bottaro and colleagues (2005), the amplitude of the noise added to the system was almost twice as large as the controllers' actions (i.e., 7.33 N vs. 4.78 N), exceeding the physiological range (Dijkstra, 2000; Kiemel et al., 2002; Bottaro et al., 2005).

2.2.2 Reinforcement learning controllers

Another type of computational framework commonly used to model motor control and motor learning tasks is the reinforcement learning controller. In motor control tasks, the central nervous system carries out motor commands based on sensory inputs from interacting with the environment. This process is similar to sequential decision-making tasks, where the best decisions need to be executed after carefully evaluating the situations. Decision-making processes are commonly characterized by a Markov Decision Process (MDP) framework, a mathematical framework that models the action selecting process whose consequences can be partially stochastic (Sutton & Barto, 1998). There are several terminologies commonly used in the framework. The agent, who interacts with the environment, makes decisions at each time step. The environment then provides feedback on the consequences of the decisions and the

agent evaluates them. There are four key elements in a typical MDP framework. The state space S contains state variables that provide sufficient description of the states that the agent is in. The state variable contains all the information that the agent needs to make a decision. The action space A contains possible actions (i.e., decisions) to make, which affects the next state $s'$the agent transitions into. The reward function R is the immediate reward provided by the environment when the agent moves from state s to state s' given action a. The transition probability $P(s'|s,a)$ is the probability of transitioning to state s'given the current state s and action a. In an MDP framework, actions are evaluated based on the expected future rewards when the agent makes decisions. The overall goal of the MDP framework is to generate and learn an optimal policy that maps the states of the system to the actions that maximize the reward (Sutton & Barto, 1998). In the context of standing balance, the agent is the CNS that performs the decisions (i.e., generating motor commands to counteract gravitational torques) in order to maintain an upright posture. The environment refers to the physical human body that moves based on the action inputs. Every time an action is executed, the consequences of the action, sensed by the sensory systems, are fed back to the agent. The agent then updates its knowledge about the actions and their consequences (i.e., expected future rewards).

Reinforcement learning provides a collection of algorithms that solve for the optimal policy in a defined MDP framework. Those algorithms can be classified into two categories: model-based and model-free algorithms. Model-based algorithm requires the agent to form a representation (i.e., model) of the environment. A typical model is the transition probability matrix, which contains probabilistic distributions of state transitions given the selected actions. One example of model-based algorithms is dynamics programming, which updates the knowledge of the model based on remembered knowledge from the last time step and the new

information learnt. Model-free algorithms do not require the agent to form any representation of the environment; instead, it interacts with the environment through trial and error and updates its knowledge of how good each action (i.e., action values) based on the expected future rewards. The classic Temporal Difference Learning algorithm is a model-free algorithm (Sutton & Barto, 1998). In Temporal Difference Learning, the agent updates its action values using the reward prediction error, which is the difference between predicted rewards and the actual rewards received. The RL algorithm used in my thesis is a special case of Temporal Difference learning: Q-learning. Q-learning is a model-free framework that learns the action value functions, which evaluates how good an action is in a particular state. The algorithm updates the learnt action value function through trial-and-error interactions with the environment until convergence, which resembles the trial-and-error learning when humans encounter new tasks.

2.3 Comparison of the advantage and disadvantages of the models presented above

2.3.1 Dual role of variability

One of the major differences between the feedback-based controllers and RL controllers lie in how they deal with variabilities. Traditional feedback control theories build on the assumption that motor variability needs to be eliminated (Todorov & Jordan, 2002). The rationale behind this argument is that improving performance of a motor task often requires reducing the variability of movement consequences, which is the aim of motor skill acquisition. For example, when basketball players perform free throws, they intend to make the execution of each shot as consistent as possible and reduce variabilities in their shooting technique. However, even among high performance athletes it is impossible to eliminate variability completely because the biological systems contain inherent noises (i.e., noises during muscle contraction and

noises from sensory receptors and sensory integration process) (Faisal et al. 2008; Renart & Machens 2014; Stein et al. 2005).

Recently, researchers proposed an alternative view on variability: it is more than just an undesired feature of motor systems. The alternative hypothesis regarding the new role of movement variability is that it serves as a meaningful exploration of the environment and can facilitate motor learning in new tasks (Wu et al., 2014; Dhawale et al., 2017; Sternad, 2018). In human motor systems, there is a large number of degrees of freedom in the action space compared with the task space. To perform a movement, the number of possible joint configurations, muscle activation patterns and even recruitment of motor neurons exceeds the ones required by the desired motion. Therefore, there are often more than one solution or strategy to perform the desired movement. Each strategy, while achieving the same outcome, likely poses different costs to the motor system depending on the goal of the movement (i.e., minimizing energy expenditure or movement variability).

When the brain goes through the different strategies in order to select the optimal strategy that minimizes the cost function, variability irrelevant to system noise occurs. To better characterize variability and clarify the context at which variabilities occur, researcher defined task-relevant variability as variability that is relevant to achieving the desired movements and task-irrelevant variability as variability that is irrelevant to achieving the desired movements (Bernstein, 1967). Task-irrelevant variability reflects how much the system explores and jumps between different motor strategies that lead to the same goal and task-relevant variability reflects how skillful motor commands are carried out.

Recent empirical evidence in both animals and humans research has provided support to the existence of task-irrelevant variability. In animal studies, adult songbirds perturbed during

singing purposefully shifted the pitches of their songs to optimize performance (Kao et al., 2005). Several pieces of evidence in human behavioral research have shown that task-relevant variability reduced over practice and task-irrelevant variability maintained their level with practice (Kang et al. 2004, Latash & Anson 2006, Scholz & Schöner 1999, van Beers et al. 2013). These results indicated that movement variability was still present in the system even after movement patterns were mastered. Then what is the purpose of the task-irrelevant variability?

Wu and colleagues (2014) examined the relationship between variability of movement in a point-to-point reaching task in earlier learning phase and the overall speed of learning. They reported that a higher baseline motor variability was correlated with a faster learning rate of the motor task, which suggested a positive role of variability in motor learning tasks. If such statement is true, then it is reasonable to hypothesize that exploration will increase at the beginning of learning and will decrease as people gradually master a skill with practice. Such hypothesis was supported by behavioral evidence in the motor learning literature. Uehara and colleagues (2019) asked participants to perform a goal-directed pointing task toward a pre-determined goal (unknown to participants). Binary feedback (i.e., successful or not) was given to participants when they completed each trial. Exploration was defined as the magnitude of change in pointing directions as a function of whether the previous trial succeeded or not. They found that exploration was higher after failed trials than successful trials. This indicates failure to obtain positive rewards facilitate movement changes, which might be a way to search (or explore) for strategies with higher reward. Pekny et al. (2015) also found that task-irrelevant movement variability increased as the probability of reward decreased. The largest trial-to-trial movement variability took place when completing a trial without receiving any reward, aiming to

increase the reward. This empirical evidence suggests that task-irrelevant variability changes with respect to the progress of learning, indicating correlations between exploration levels and learning progresses.

Unfortunately, it is very difficult for most traditional feedback-based controllers to reconcile the apparent conflict in the dual roles of variability, as they treat variability as detriment to performance and aim to minimize it. On the other hand, RL frameworks rely on the assumption that movement variability (i.e., exploration) is necessary for learning and adaptation to happen when humans encounter new tasks and environments. In RL models, the agent exploits its current knowledge of the environment to make decisions and also continually explores the environment in order to update its knowledge with the goal to maximize rewards in the long run (Sutton & Barto, 1998). This is known as the exploration-exploitation trade-off (Sutton and Barto, 1998). Continuous exploration is a prerequisite to find the optimal policy in an MDP framework as it increases the exposure to a variety of possible strategies to complete the task.

The features of RL framework address the dual roles of movement variability discussed previously because the framework can set up reward functions to minimize variability and still enable continuous exploration at the same time. Thus, they can be used to represent movements in both the learning stage and the mastering stage of motor tasks under static as well as dynamic environments.

2.3.2 Neuromechanical evidence of RL

MDP and RL frameworks have been used to model both continuous (i.e., walking and standing) and discrete (i.e., sit-to-stand movement) motor tasks (Michimoto et al., 2016; Jamali et al., 2017; Selinger et al., 2019; Song et al., 2020). Michimoto et al. (2016) modeled human standing balance using an MDP framework and their reward function penalized changes in the

distance from the upright posture and amplitudes of change in consecutive torques. With adequate parameters, the model exhibited low-frequency oscillatory behavior similar to human quiet standing. In addition, they found that with different weights placed on penalizing the smoothness of torque, the MDP model exhibited characteristics of both continuous and intermittent feedback control mechanisms of standing balance. This shows the model reproduced two plausible mechanisms of control. However, the connection between the model and physiological mechanisms was not clearly established in their paper. There is a lack of research evidence in the standing balance literature that shows humans aim to minimize swaying velocity and smoothness of torques during standing. Despite that the model exhibited behavior similar to the intermittent controller, it might simply be due to the fact that the output torque was regulated to be similar to the preceded output from the last time step.

In the general context of motor control, MDP frameworks have been applied to other motor tasks. Jamali et al. (2017) modeled the sit-to-stand task using an RL model. In their model, they assumed that people minimize joint torques when performing the task. Thus, they penalized the amplitude of torques at each time step and compared their model simulation (from a Q-learning algorithm) with actual kinematics and kinetics (i.e., joint torques and trajectories) of participants executing sit-to-stand movements. High resemblance of those kinematic variables (i.e., joint positions) between simulated and experimental data was reported from the study, which indicates that reinforcement learning can provide an effective controller to model motor control tasks like sit-to-stand movements. Selinger and colleagues (2019) modeled locomotion with an RL model to investigate whether people would alter their step frequencies (i.e., changing strategies) to optimize energy consumption when the dynamics of walking changed. They manipulated the relationship between energetic consumption and step frequency by shifting the

energetic minima to a different frequency and observed how people responded to the changes. Their results showed that participants adapted their gait patterns and their behavior converged gradually to the new minima of energy consumption, which was consistent with the predictions from their RL model. This indicates that a simple RL model with an energy minimization goal can predict how people adapt to changed landscapes of locomotion.

MDP frameworks have been applied to interpret a large amount of behavioral and neurobiological data. In particular, the activity of dopamine neurons has been studied extensively using the RL framework. In most of RL models, the avenue for the agent to update its knowledge of the environment and the consequences of its actions (i.e., internal representation of self and world) is through the reward prediction error. The reward prediction error is the core of multiple RL algorithms such as Temporal Difference learning and Q-learning (Niv, 2009). Dopamine has been proposed as a messenger for reward error signals (Glimcher & Bayer, 2005; Schultz et al., 1997), as the outputs produced by dopaminergic neurons were consistent with the scalar reward prediction error signal (Schultz et al., 1997). The notion that dopamine activity signals reward prediction was first supported by the Pavlovian conditioning tasks, an instance of predictive learning (Yerkes & Morgulis, 1909) in which people learned the causal relationship between two events. For example, touching a sizzling pot will most likely cause burning sensation on the skin. In one Pavlovian conditioning experiment, monkeys were presented with food as rewards preceded by auditory or visual cues (e.g., tone or light indicators). When the monkeys were first exposed to the task, the dopamine neurons were activated in response to the reward stimuli. After a number of trials, dopaminergic responses to the same reward disappeared (Schultz, 1986). In trials when the rewards weren't delivered following the same visual or auditory cues, dopamine neurons activity was depressed below their basal firing thresholds at the time that the reward

should have occurred. The experimental results showed that dopaminergic activity was not only correlated with the amount of rewards, it also reflected the difference between the actual reward received and the expected reward. This study provides empirical evidence that dopamine neuron activity is correlated with reward, specifically the reward prediction error.

The connection between Temporal Difference error and dopaminergic neuron activity motivates the development of Q-learning to model standing balance in this thesis, as the Q-learning algorithm facilitates learning using the Temporal Difference error updates. The connection between computation theories and experimental results provides a quantitative framework for psychological and biological studies.

2.4 Purposes and hypothesis

The first objective of my thesis is to model human standing balance with the MDP framework and identify the best parameter combinations that represent characteristics of quiet standing behavior observed in previous empirical evidence in the anterior-posterior direction. The behavior of the MDP model will be evaluated based on the range and root mean square (RMS) of the body angles and frequency characteristics (i.e., mean power frequency) of the body angle. I hypothesize that an MDP model with properly chosen parameters will replicate the main characteristics of standing balance.

The second objective is to generate predictions from the MDP framework with parameters mimicking features of human standing balance to determine how participants respond to changing landscape of the standing balance dynamics. Specifically, I will add a torque bias term to the physical laws of standing (in computer simulation and experimentally using a robotic platform), creating a mismatch between the expected body orientation and torque required to maintain this orientation. Given that my MDP model rewards minimization of the generated

torque, I hypothesize that it will balance the body (i.e., inverted pendulum) at different angles for the applied torque offsets in order to maintain the generated torque constant. I further hypothesize that participants will also maintain their torque level but balance at different positions when a torque bias term is added to the dynamics of standing balance.

Overall, the proposed computational and experimental work for my thesis will provide some insights regarding whether people aim to maintain their preferred torque at the same level or to maintain their body orientation during standing.

3 Methodology

3.1 Computational modeling

3.1.1 Model description

I chose to model standing balance using an MDP framework because it not only generates behavior directed by quantified task goals similar to traditional feedback-based controllers, but also represents the exploratory behaviors when the agent faces a new task. The objective of the framework is to learn the optimal policy for quiet standing through trial and error in order to simulate the process in which humans learn to balance under new conditions.

3.1.1.1 Model setup and training

First, I provide an overview of the different components of the model. The model agent only contains the decision-making components. In the general context of motor control, the agent would refer to the neurons in the brain responsible for making those decisions. For standing balance specifically, the model agent is responsible for determining a one-dimensional AP ankle torque. The ankle torque then acts on the environment of the model which includes standing balance in the AP direction, sensory and motor dynamics. In the standing balance context, ankle torques generate changes in the simulated inverted pendulum angle and angular velocity. These represent the model environment states that are sensed and utilized by the agent to make decisions on the next action to take. When each action is executed, the environment evaluates the decision and the new state that the agent transitions into and delivers a reward to the agent. The reward function in this MDP framework was defined to penalize falling and magnitude of the ankle torque.

Several parameters were implemented in the design of the environment to better replicate physiological properties of human standing: motor noise, two sources of sensory noise which

included sensory noise on the angle estimate (i.e., angle noise) and sensory noise on the angular velocity estimate (i.e., angular velocity noise), and metabolic cost. Here is a brief description of each parameters.

(1) Motor noise refers to the noise added to the torque output at each time step. Physiologically, it represents the noise from the process of generating the torque at the ankle joints by activating the muscles (Faisal et al., 2008). In the execution, it was represented as a 1/f noise, whose spectral amplitude is inversely proportional with frequency. To obtain such noise profile, I first obtained white noises of desired amplitudes (i.e., noises with uniform frequency spectrums), and then passed them through a filter that approximated the 1/f power spectrum. This term is similar to the motor noise component in PID controllers (Bottaro et al., 2005; Masani et al., 2008).

(2) Angle noise refers to the noise added to the estimated body angle at each time step. Physiologically, it represents the uncertainty when estimating the body angle relative to space. The process of obtaining the angle noise was the same as motor noise.

(3) Angular velocity noise refers to the uncertainty in sensing the body angular velocity at each time step. The process of obtaining the angular velocity noise was the same as motor noise.

(4) Metabolic cost refers to negative reward given to the controller proportional to the amplitude of the torque.

As the control of standing balance naturally contains sensory and motor delays, I also implemented the delay in the environment. The delay in the sensorimotor loop during natural standing balance is around 100 to 160ms (Kuo 2005; Forbes et al., 2018). In the environment, I implemented the sensory delay, motor delay and reward delay. Intuitively, it means that the

presence of state, effect of actions, delivery of rewards were delayed to the agent. I implemented the delay using a buffer array (i.e., queue) that stored the state transitions, action selected and rewards at each time step, and returned the delayed variables to the agent.

During the training phase of the model, the agent explored the environment through trial-and-error interactions. At each time step within a training episode, the agent selected an action instructed by the policy. The policy that the agent followed was the $\epsilon - greedy$ Q-learning (the details can be found in Appendix A). $\epsilon - greedy$ means that the agent selects a random action with the probability of ϵ and selects actions that maximizes the overall reward with the probability of $1 - \epsilon$. The purpose of using the $\epsilon - greedy$ algorithm is to ensure that the controller continuously explores the state space and action space. Q-learning refers to a collection of algorithms that updates the action values through reward prediction errors. In other words, every time an action was executed and a reward was given, the agent calculated the difference between the predicted reward and the actual reward received following an action selection and used the difference to update the action value of such action, which refers to the expected future rewards following this action. After selecting and implementing the action, the agent transitioned to the next state, determined by the environment dynamics. The agent kept going through this action-selection-state-transition iteration until the episode finished. The episode reached the end when either of the following conditions was met:

(1) The pendulum fell, which was defined as when either of the following was met
    a. when the angle of the inverted pendulum exceeded 7.5 degrees (0.13 radians) or -4.5 degrees (-0.08 radians)
    b. The magnitude of the angular velocity exceeded 0.4 radians/s.
(2) The length of the episode reached 2 minutes.

When an episode ended, the environment was reset. Then the next episode began and the agent started at a random state. Figure 1 represents the schematic of the MDP model. The details of the Q-learning algorithm and environment setup can be found in Appendix A and B respectively.

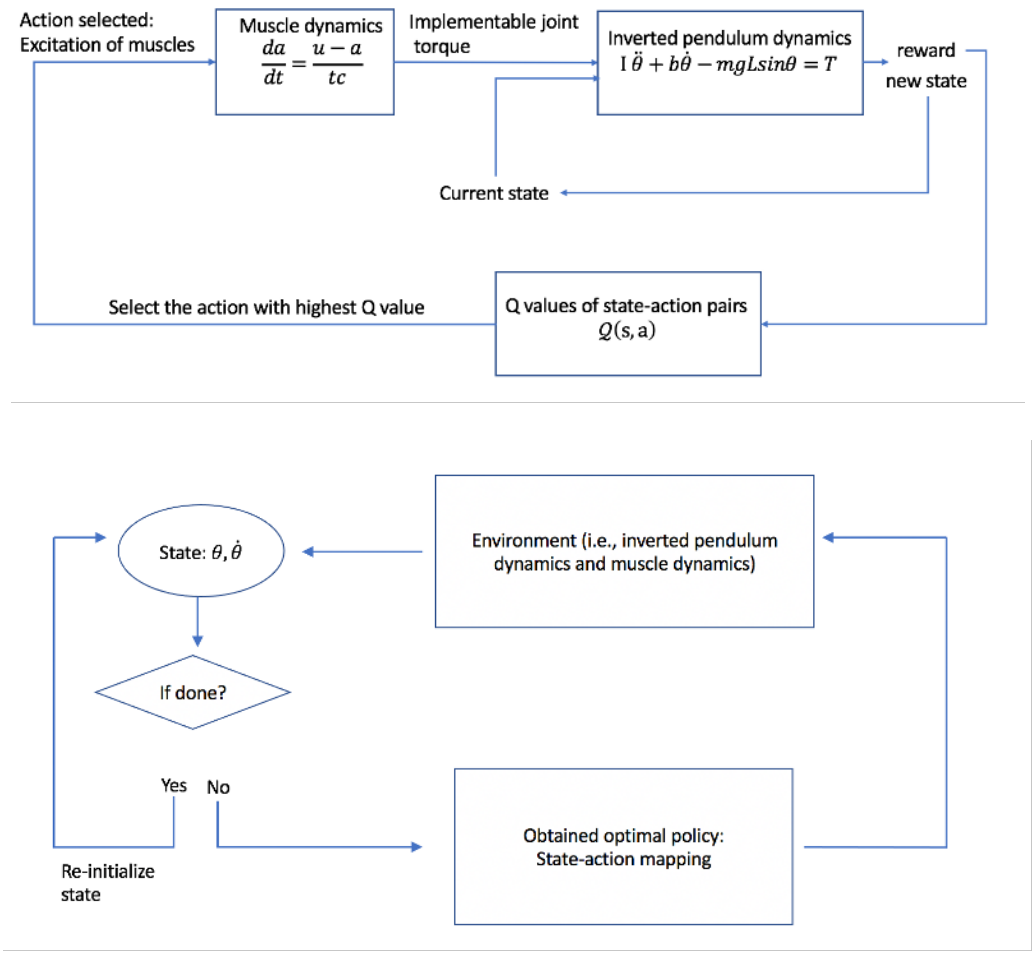


Figure 1. Schematic diagram of the MDP model workflow. Top: Training. The agent selected actions with the highest Q-value in the current state. The action was then fed into the muscle dynamics to yield the joint torque, which was then fed into the dynamics of standing balance. The environment then outputted the new state and the reward that evaluated the new state and action selected. The agent then used the state transition information and reward received to update the Q values. Bottom: Testing. Starting from a random state, the policy outputted an action, which was then fed into the inverted pendulum environment

to yield the next state. If the pendulum fell or the time step reached maximum, simulation would terminate. Otherwise, the iteration continued.

3.1.1.2 Testing

At the end of the training phase, the optimal policy was saved and then used to generate a trial of simulated balancing data by mapping the actions to the states. The simulation started with a randomly assigned state (i.e., angle and angular velocity) and was then propagated using the policy with the same environment (i.e., same parameters). The simulation terminated either when the pendulum fell or when the duration of the simulation lasted for two minutes. The reason why I selected this duration is that longer duration of simulation allowed for low-frequency oscillations to be revealed in the balance control (Visser et al., 2008). Carpenter and colleagues (2000) reported at least 60s of recording duration to obtain reliable and stable measures of postural sway.

3.1.2 Parameter space configuration

Four parameters and their effect on the simulation outcomes were investigated: motor noise, angle noise, angular velocity noise and metabolic cost. To explore the effect of each parameter on the simulated behavior, grid searches were performed to obtain simulation results of parameter combinations with different values. The value of motor noises amplitudes explored in this thesis were [0, 0.5, 1, 2, 3, 4, 6, 8, 12, 16, 20] Nm. The values of angle noise amplitudes included in the grid search were [0, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.04] radians. The values of angle noise amplitude explored in the grid search were [0, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.04] radian/s. The values of the metabolic cost investigated in this thesis were [0, 1, 2, 3, 4].

The purpose for creating a parameter space is to compare the time and frequency characteristics of the CoM movement (i.e., range, RMS, mean power frequency and 99%

frequency bandwidth) of the simulated behavior with the characteristics reported in the literature (See Chapter 2 Table 1 for exact values). The extraction of those characteristics will be presented in section 3.3.1. The script of the model was written in Python (3.7) and I ran the model remotely using Jupyter Notebook on Compute Canada servers.

3.1.3 Simulation of the experimental conditions using the MDP model

The purpose of this part of my thesis was to replicate changes in the physical laws of standing balance (i.e., addition of a torque offset) using the MDP model. The added torque bias terms to the biomechanics of standing balance in the MDP model were identical to the experimental study (Note: details of the experimental conditions are listed in Section 3.2.3). The MDP model was first trained with no knowledge of the environment using the Q-learning algorithm (i.e., the setup of the environment and algorithm was identical to Section 4.1.1). After the agent finished learning the normal dynamics of standing, the torque bias terms were added into the environmental dynamics. Then the agent continued learning with the learned knowledge (i.e., Q-values) until the algorithm converged again. The torque and inverted pendulum angle before and after the added torque bias terms were recorded. The bias term values matched the ones I used in the experimental conditions. Three trials of learning were conducted for each experimental condition, which also matched the empirical approach. The simulated behavior from the MDP model (i.e., simulated body angle) was then compared with the predicted behavior from the hypothesis to see whether the inverted pendulum maintained the torque level and balanced at different positions.

3.2 Experimental work

3.2.1 Sample and recruitment

Three participants (2 males, age: 26.67 $\pm$5.03 years (mean$\pm$SD)) with no previously known history of neuromuscular diseases or postural disorders were recruited for this study. The recruitment strategy was convenience sampling. Paper-based flyers and electronic messages were sent around the UBC community during the recruitment process. Potential participants were not excluded from the study based on ethnicity, gender or socio-economic status. Descriptions of the project and the consent forms were sent to the participants via email prior to the testing day and written consent forms were obtained from all participants on the day of testing. Procedures of the experiment were verbally explained to the participants before the experiment began. The experiment was approved by the University of British Columbia Human Research Ethics Committee and the ethics approval number is H18-03702.

3.2.2 Apparatus

The experiment took place in the Sensorimotor Physiology Laboratory on a customized robotic device that operates based on the mechanics of an inverted pendulum. A servo-motor (OMNUC G5 R88M-K1K530H-BS2; OMRON, Japan) rotated the backboard of the robotic platform around the ankle joint in the sagittal plane to simulate the whole-body sway in the anterior-posterior (AP) direction. The resolution of the motor encoder is 2^{13} pulses per revolution, resulting in 0.00004823 – 0.00004876 degree per count, dependent on the backboard angle. The control of the motors was programmed through LabVIEW (National Instrument, USA). Participants were instructed to stand on two force plates (BP250500; AMTI, USA), which were secured on top of the ankle-tilt platform. They were strapped onto the robot through belts located at the waists and shoulders. The reference system for the robotic balancing platform was

defined as the following: positive x axis pointed to the left of the participants when they stood on the platform (i.e., facing away from the backboard), positive y axis pointed forward when they stood on the platform and positive z axis aligned with gravity, pointing downward. Using this reference system, anterior body lean was positive (as the rotation vector pointed to the positive x axis) and posterior body lean was negative. When participants pushed down on their toes, the torque/moment they generated on the ground (i.e., plantar-flexor torque) was positive along the x axis (dorsi-flexor torque was negative), which was opposite to the torque applied to the body. Throughout the experiment, the ankle-tilt platform and force plates were Earth-fixed and the backboard moved forward (i.e., positive whole-body angle) and backward (i.e., negative whole-body angle) to simulate the whole-body movement in the AP direction.

A PXI computer with data acquisition board (PXI-6289; National Instrument, USA) recorded the forces and moments applied by the ground to the participants at 500 Hz and fed these values into a real-time state space model that represented the inverted pendulum dynamics. Due to technical issues (i.e., one of the force plates was not functioning properly), the dynamics of the robotic platform was driven only by the force plate under the left foot, assuming that the weight on the two force plates was evenly distributed and the torques applied underneath the feet in the AP direction were identical. The AP motor then moved the backboard to the calculated whole-body angle calculated based on the inverted pendulum equations. The motors were controlled by a FPGA board (PXIe-7846R; National Instrument, USA) which ran at 2 kHz. To simulate the physical limit of standing balance, virtual limits were implemented to the robot (+6 degrees and -3 degrees). When the limits were reached, the robot simulated a spring and damper to impede participants from further exceeding the limit and helped them restore their balance.

The built-in inverted pendulum of the robotic platform can be represented using following equation:

$$I\ddot{\theta} + b\dot{\theta} - mgLsin\theta = T \quad (3)$$

Here, θ is the angular position of body's center of mass relative to the ankle joint. m is the body mass. L is the length from the center of mass to the ankle joint. T is the torque at the ankle joint. I is the mass moment of inertia about the ankle joint and b is the viscoelasticity parameter and was manually programmed to be zero in the current experimental setup because of the inherent viscoelastic property of the ankle joint. I manually inputted participants mass, height and distance from the center of mass to ankle joint into a custom-written LabVIEW program to match the inverted pendulum model with each participant's anthropometric characteristics.

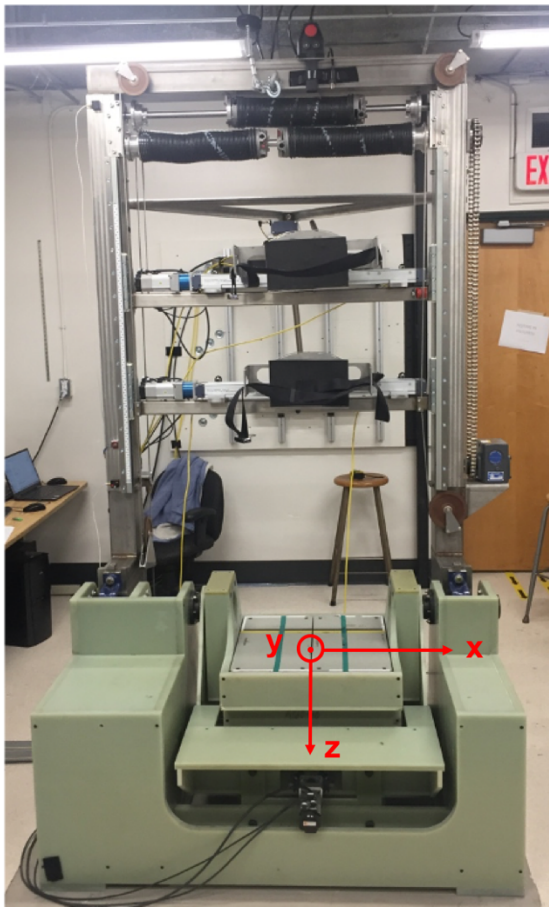


Figure 2. Apparatus used in the experiment was the robotic balancing platform.

3.2.3 Experimental perturbation

The goal of the experiment was to manipulate the governing standing balance dynamics (Equation 3) simulated in the robotic platform. The quasi-static torque-angle relationship (i.e., when humans stand with minimal movement and angular velocity is close to zero, which make the terms $I\ddot{\theta}$ and $b\dot{\theta}$ in equation (3) close to zero) can be expressed as $T = -mgL\theta$ when the angle θ is small (as $sin\theta \approx \theta$ in radians when θ is small). Under normal dynamics, the position associated with minimal torque (i.e., torque = 0) is 0 degree (defined as the body's CoM aligned with the ankle joints). To manipulate the position associated with minimum torque, a bias term T_{pert} was added to the equation such that the quasi-static dynamics became $T + T_{pert} = -mgL\theta$. Thus, the new position associated with minimum exerted torque (zero torque) became $\theta_{pert} = \frac{-T_{pert}}{mgL}$, as $T(\theta_{pert}) = 0$. For example, when a positive torque perturbation (T_{pert}) was added, the new angle minima would be slightly backward compared with natural body angle if people aim to minimize the torque. The control of perturbations was implemented to the LabVIEW program that controls the dynamics of the robotic platform.

During the experiment, the torque bias term was implemented as a linear ramp that reached the target torque bias term in 6 seconds. Meanwhile, the AP angle of the robot (i.e., body angle of participants) was brought to the target angle (differed for distinct conditions) within 6 seconds using a sigmoid function with extra motion of the backboard added to obscure the true motion of the backboard and disturb participants' perception of their body angle in space. I called the extra motion of the backboard the "angle perturbation". The angle perturbation consisted of the summation of four cosine waves with fixed frequencies (i.e., 0.3, 0.7, 1.1, 1.7 Hz). Note that participants were not in control of the balance simulation when the torque bias was added. The angle perturbation continued until the exerted AP torque and body angle were

both within a pre-defined limit from the target angle θ_0 and target torque $-mgL\theta_0 - T_{pert}$. To be more specific, the perturbation stopped when both the two following criteria were met:

$$|\theta - \theta_0| < 0.3\ deg\ (4)$$

$$\left|T - \left(-mgL\theta_0 - T_{pert}\right)\right| < 3\ Nm\ (5)$$

The stop criteria ensured that participants balanced at a body angle close to the target angle value while the exerted torque value was close the torque required to hold the body angle at the target angle under the standing balance dynamics in the experimental condition. This allowed for a smooth transition when participants gained back control of the robot with the torque perturbation added.

In the following section, I use the term "perturbation" to define the process in which the torque bias term and backboard angle are brought to their target values. Participants were not informed about the purpose of the experiment or the nature of the perturbations. They were only instructed that during the trial they would experience perturbations and their task was to maintain their balance naturally throughout the trials.

3.2.4 Experimental Conditions and Procedures

When the subjects first came in, I measured their height, weight, location of center of mass, leg length, pelvis depth (from right anterior superior iliac spine to right posterior superior iliac spine ipsilaterally), and pelvis width (from left anterior superior iliac spine to right anterior superior iliac spine). Hip joint width was estimated using the equation from Harrinton et al. (2007).

I then obtained their zero-degree posture, which is the body alignment at which the body is perpendicular to the ground and the exerted torque is zero. It was important for each experimental trial to start with the zero-degree posture, otherwise the actual standing balance

dynamics would not be the same for each trial. I asked participants to stand on the force plates with their ankle joints aligned with the AP motor's axis of rotation and their feet width the same as the estimated hip joint width. Then the height of the hip and shoulder clamps were adjusted so that they were around the greater trochanter and mid-deltoid region respectively. The widths of clamps were adjusted to make sure they were as close to the body as possible. Participants were strapped onto the backboard with foam pads between their body and the clamps. Depths of the shoulder and hip clamps were also adjusted so that the horizontal force was as small as possible (ideally smaller than 20N) and AP torque was also close to zero. Then their posture would be the zero-degree upright posture, as the AP torque is directly proportional to the body angle based on the quasi-static relationship of standing. The depths of the clamps were also recorded, and the feet positions were traced onto the white paper on the force plate.

I first conducted three quiet standing trials, each lasting 120 s. Participants were asked to stand on the force plates with feet within the traces on the white paper and their upper body was strapped onto the backboard. Before each trial started, I checked on the AP torque to make sure it was still close to zero. If not, minor adjustments were made to the feet positions to ensure that AP torque was close to zero. During each trial, body angle and AP torque (i.e., torque applied to the body) were recorded. Then, average angles of each quiet standing trial were calculated. Averaged body angle of quiet standing θ_0 was calculated by averaging the mean body angles from each quiet standing trial.

I then performed a series of experimental perturbation trials. Each trial contained two perturbations. The first perturbation occurred at the beginning of the trial with no added torque perturbation and brought participants back to the averaged body angle θ_0 during quiet standing trials $\theta_0$followed by 90s of quiet standing. Then, the second perturbation took place with an

added torque bias and corresponding displacement of the backboard angle to the target angle with amplitudes dependent on the experimental condition, followed by another 90s of quiet standing. The purpose of having two perturbations in the same trial was to compare the body angles after each perturbation. Given that there might be variability in preferred posture in each trial, the first perturbation served as a control perturbation where no torque bias was added. By comparing the differences in the applied torques and resulting whole-body angle after two perturbations, I eliminated the potential influence of stance positions and other sources of between-trial variability.

I used four different torque bias terms in this experiment: $\pm mgl * 0.6/57.3\ Nm$ and $\pm mgl * 1.2/57.3$ Nm. To frame them in terms of percentage of mgl in order to normalize them with respect to individual participants, the four conditions were $\pm 1.05\% mgl$ and $\pm 2.1\% mgl$. If participants maintained the same torque under normal and altered dynamics, they would experience a shift in their body angle of 0.6 and 1.2 degrees respectively. From the previous pilot testing sessions (data not presented in this thesis), I found that at $1.05\% mgl$, the torque bias term was imperceptible, whereas it was perceptible sometimes when the added torque was $2.1\% mgl$.

With the same torque bias term, there were two different target angles participants were brought to during the perturbation. The first target angle was the mean angle identified in the quiet standing session (i.e., θ_0). Experimental trials that brought participants to θ_0 were referred to as the preferred angle trials (i.e., pink dots in Figure 3a) for future references. The second angle was the angle at which participants maintained the same torque as the quiet standing condition, which was $\theta_0 - \frac{T_{pert}}{mgL}$. Experimental trials that brought participants to $\theta_0 - \frac{T_{pert}}{mgL}$ were referred to as the preferred torque trials for future references (i.e., blue dots in Figure 3a). In

total, there were eight conditions. Figure 3 shows the landscape of each condition, and Table 2 lists the added torque bias terms and the target angles at which the participants were brought to after the 2nd perturbation for all of the experimental conditions.

Based on my second hypothesis, balance behavior of the MDP controller and the participants in the experiment would shift backward with positive added torque if the goal of standing was indeed to maintain their torque at a constant level. Correspondingly, the hypothesis predicts a forward-shifted behavior with negative added torque. Their new preferred body angles after the added torque bias term would be $\theta_{pert} = \theta_0 + (\frac{-T_{pert}}{mgL})$. In other words, for both the preferred angle trials and preferred torque trials, their new preferred body angles would align with the target angles that participants were brought to during the preferred torque trials (blue dots in Figure 3) if the goal of standing is to maintain the torque level.

Condition	Torque bias term	Target angle
1	$1.05\%mgl$	θ_0
2	$1.05\%mgl$	$\theta_0 - 0.6$
3	$2.1\%mgl$	θ_0
4	$2.1\%mgl$	$\theta_0 - 1.2$
5	$-1.05\%mgl$	θ_0
6	$-1.05\%mgl$	$\theta_0 + 0.6$
7	$-2.1\%mgl$	θ_0
8	$-2.1\%mgl$	$\theta_0 + 1.2$

Table 2. Experimental conditions: listed by the added torque bias terms and the target angles

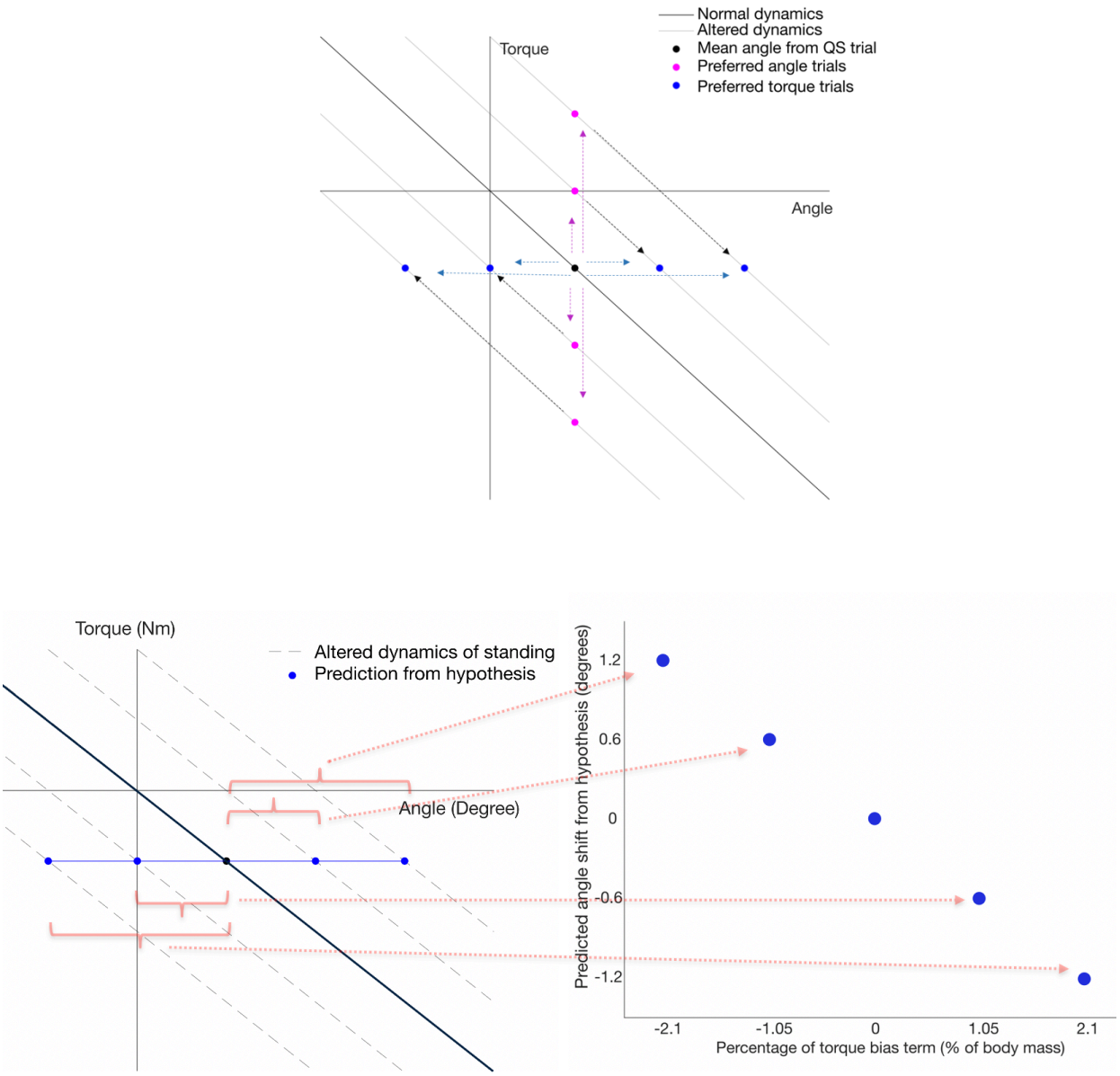


Figure 3. Top: Schematic representation of the experimental conditions using the load-stiffness landscape. The black line represents normal load-stiffness relationship of standing. Grey lines represent altered load-stiffness relationship with added torque bias terms. The black dot represents the mean body angle obtained from quiet standing trials. Blue dots represent the preferred torque trials and magenta dots represent the preferred angle trials. Bottom: My second hypothesis predicts that participants will maintain their torque with added torque bias term (blue dot and line in the left figure), and they will shift their body

angles (right). The dotted red arrow shows the transitions of the body angles the left figure to the body angle differences in the right figure.

Participants performed three trials for each condition and each trial lasted 196s – 278s depending on the duration of the perturbation. I also conducted 4 control trials in which both perturbations brought participants to their preferred angles with no added torque. The purpose of including the control trials was to establish the baseline variability of participants' preferred posture when no torque bias term was added. The results from the control trials can also inform me about whether the order of the perturbation influenced the behavior (i.e., whether participants' responses to the first perturbation would be different from the second perturbation when all the other conditions were the same)

The order of control and perturbation trials was randomized. During the trial, the participants put on ear plugs and had their eyes closed all the time. Lights in the testing room were also turned off throughout the trials.

3.3 Data analysis

3.3.1 MDP model simulation analysis

Reinforcement learning policy obtained from the simulation data were saved and were used to generate simulated data of balance control. Inverted pendulum angle and ankle torque were saved from the simulation data. I calculated CoM movement based on the inverted pendulum angle using Equation (6).

$$CoM = L * \theta \quad (6)$$

Here L is the distance from CoM to the ankle joint, and θ is the inverted pendulum angle.

Range and RMS measures were computed for the CoM data. Frequency analysis methods were applied to both CoM and torque data (i.e., Fast Fourier Transform). I performed discrete Fourier transform with frequency resolution 0.0084 Hz on single window with length 120

seconds. Spectral analysis is a technique that computes the distribution of energy at each frequency in the given bandwidth, providing insight on periodicity of the system (McClenagan et al., 1994). It has been used previously to describe characteristics of various physiology systems, including human postural control (Bensel et al., 1968; Hasan et al., 1996; Soames & Atha, 1982). In order to select a set of parameters that captured the characteristics of standing balance, I used the following indicators: range and RMS for the time series CoM data; MPF and 99% of frequency bandwidth of the CoM data in the frequency domain. MPF was calculated using the equation:

$$MPF = \frac{\sum_j f_j p_j}{\sum_j p_j} \quad (7)$$

Here, f_j are the frequency values from power spectrum and p_j is the power spectrum at frequency f_j. When calculating the MPF, I disregarded the frequency component above 5 Hz. This is to conform to the data processing method in Hasan et al., (1996) so that I could compare their reported values to ours. 99% of frequency bandwidth refers to the frequency below which 99% of the total energy is accounted for.

Then I compared the computed variables using the simulation results with the data previously reported in the literature (Table 1 from Chapter 2) and identified the parameter combinations with time and frequency characteristics that falls within 95% confidence interval presented in Table 1.

To further explore the relationship between MDP parameters and frequency characteristics of simulated standing balance, multiple regression analysis was applied to assess linear relationship between independent variables (i.e., motor noise, sensory noises and metabolic cost) and dependent variables (i.e., MPF and 99% frequency bandwidth). I computed

the coefficients for each independent variable and the r square value of the regression analysis to quantify the strength of the linear correlation.

3.3.2 Analysis of the MDP-simulated behavior shifts under the same experimental condition

Using the MDP model that best matched the physiological characteristics of balance (see Results), I calculated the mean and standard deviation of the simulated standard behavior (i.e., body angle) before and after the added torque bias terms. This procedure enabled quantitative predictions (i.e., body angle shifts due to added torque bias terms) from the MDP model under the altered dynamics and clear comparisons from the predicted behavior by the hypothesis. I expected the MDP model would predict an angle change in the pendulum angle that would match the offset torque added to the simulation according to the second hypothesis. That is, the pendulum angle shift should be $-\frac{T_{pert}}{mgL}$ (See Section 4.2.4 for justification). Any deviation from this prediction would be tested against the experimental data.

3.3.3 Experimental data analysis

For the experimental results, the mean and standard deviation of the body angle recorded from the robotic balancing platform were calculated for all quiet standing trials. For the experimental conditions (including the control conditions), I split and extracted the data (i.e., body angle and torque) after the perturbations ended during quiet standing for both the 1st and 2nd perturbations. I then fitted a Least Square exponential model to the quiet standing period after the perturbation.

$$y = Ae^{Bt} + C \quad (8)$$

The reason for selecting an exponential decay model was that the rate of change in participants' behavior decreased gradually and stabilized at a certain level after some time. This

resembled the behavior of an exponential decay model. The exponential regression analysis provided a standardized method to quantify when the behavior stabilizes and at which level it stabilized. Here, the A term represents the starting point of the fit, B term represents the time constant of the exponential decay, or in other words, the rate of the behavior shift. Finally, C term is the asymptote value, which is the level at which the behavior stabilizes.

I defined convergence of standing balance behavior after each perturbation by whether the fitted exponential decay function had decayed to a level that was in the range of within one standard deviation of the normal quiet standing angle (calculated from the quiet standing trials) of the predicted asymptote level at the end of the standing period (i.e., t = 90s). If the fitted exponential function decayed to the said level when the trial ended, then it was considered converged, or in other words, the behavior had reached a steady level. If the trial converged, then I defined the preferred body angle after the perturbations using the asymptote value (i.e., C term) of the exponential fit. Figure 4 illustrates the definition of the convergence criteria in this study.

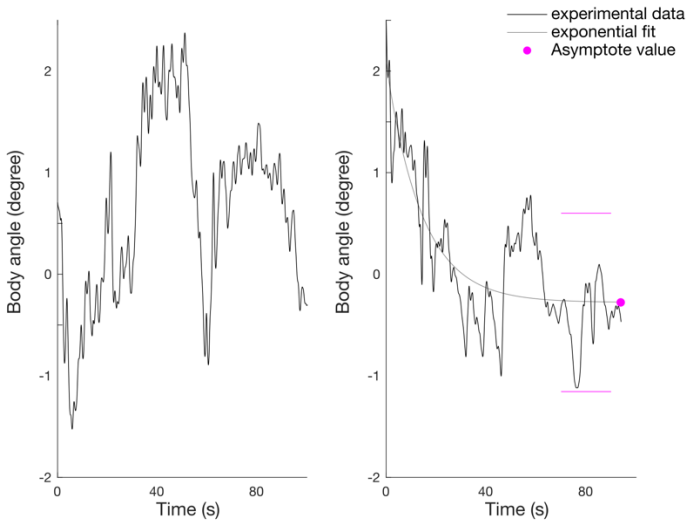


Figure 4. A schematic description of criterion of convergence. The graph on the left shows the body angle over time in a quiet standing trial. The right figure shows the raw angle trace after the 1st perturbation from one experimental trial (black) and the fitted exponential curve (gray) to the angle trace. The magenta dot represents the asymptote value from the exponential regression and the magenta lines represent 1 standard deviation of body angle during quiet standing trials away from the asymptote values. In this example, the fitted exponential curve is within the bands of the magenta lines, and therefore, the trial is considered converged.

To quantify the effect of the added torque bias term on the preferred body angle during standing, I conducted two analysis. First, I determined whether the torque bias introduced in the second perturbation produced observable changes in the preferred angle. Here, I first quantified the baseline variability of the preferred body angle after the first perturbation with no added torque in each trial. Specifically, I calculated the mean, standard deviation, and 95% confidence interval of the preferred angle after the 1st perturbation from all of the experimental trials. Knowing the baseline variability provided me with a range of reasonable values for preferred body angle in conditions without torque bias. Then I compared the preferred angle after the 2nd perturbation with torque bias of the experimental trials to examine whether it was within the 95% confidence interval of the preferred angle baseline variability. This provided information on whether introducing a torque bias produced changes in preferred angle that are not explained by natural variability. Next, I calculated the difference of the preferred angle after the 1st and 2nd perturbation of the same trial for both the control trials and experimental trials. I then compared the differences of the preferred angle in experimental trials against control trials. By comparing the difference between the behavior after two perturbations, the between-trial variability (i.e., from minute changes in the stance position) can be minimized.

Second, in order to investigate whether the participants maintained their torque at a constant level predicted by the hypothesis, I conducted linear regression analysis on the preferred angle after the 2nd perturbations in all experimental trials and control trials. The slope of the regression line informed me whether the participants maintained their exerted torque (i.e., horizontal slope) or the body angle (i.e., vertical slope) or a combination of both when the dynamics of standing changed. All data analyses were performed using Matlab R2020a (Mathwork, USA).

4 Results

4.1 Results from Computational Approach

4.1.1 Simulation results from parameter searches

Of all 3520 simulations, 1497 simulations lasted for 2 minutes without falls. The four parameters from the standing balance simulation environment (i.e., motor noise, angle noise, angular velocity noise and metabolic cost) determined the characteristics of the simulated balancing behavior, as they defined the environmental dynamics which then determined the optimal policies. Figure 5 shows two example simulations from the MDP controller with different sets of parameters. The first simulation was generated with all sources of noise set to zero (motor noise: 0 Nm; angle noise: 0 rad; angular velocity noise: 0 rad/s; metabolic cost: 0); identified as (0 Nm, 0 rad, 0 rad/s, 1). Note that this parameter set order and format will be used for the remainder of the thesis when referring to specific parameter combinations. Both time and frequency domain analysis showed high frequency oscillations with small low-frequency components, which is far from how humans behave. In contrast, the second simulation generated from the parameter set (8 Nm, 0.02 rad, 0.02 rad/s, 4) showed low-frequency oscillations. Figure 6 shows regions from the parameter space in which the MDP controller generated successful simulations. The general trend was that with high parameter values it was less likely for the controller to balance successfully. The density or probability of successes showed a decreasing trend when the parameters increased. Similar trend can be observed from Figure 7, where the number of successes reduced with increased value for the four parameters. Indeed, there were successful simulations with high parameters values (e.g., 20Nm, 0.008rad, 0.04rad/s, 4) and failed simulations with small parameter values (e.g., 0.5Nm, 0rad, 0.002rad/s, 1).

Out of 1497 successful simulations, 541 demonstrated 99% power bandwidth between 2.47 and 3.93 Hz, 104 demonstrated MPF between 0.55 and 0.75 Hz, and 29 passed both the 99% power bandwidth and MPF criteria. None of the simulations, however, yielded RMS and range values within the physiological limits (i.e., 7.21 – 13.45 mm and 0.14 – 5.24 mm respectively) of human standing balance. Therefore, the results of the simulations only partially supported my first hypothesis because the MDP model could not replicate all the characteristics of standing balance. Among the 29 simulations matching the frequency characteristics of balance, the mean of the range of CoM during the 2-minute period was 74.2 ± 20.0 mm and the RMS of the CoM is 22.3 ± 10.3 mm whereas the smallest range and RMS values from those simulations were 34.4mm and 10.1 mm. I selected the model that exhibited the smallest CoM range to minimize difference from previous experimental results. The identified parameter combination was 8 Nm for motor noise, 0.004 rad for angle noise, 0.006 rad/s for angular velocity noise and 1 for metabolic cost; identified as (8 Nm, 0.004 rad, 0.006 rad/s, 1).

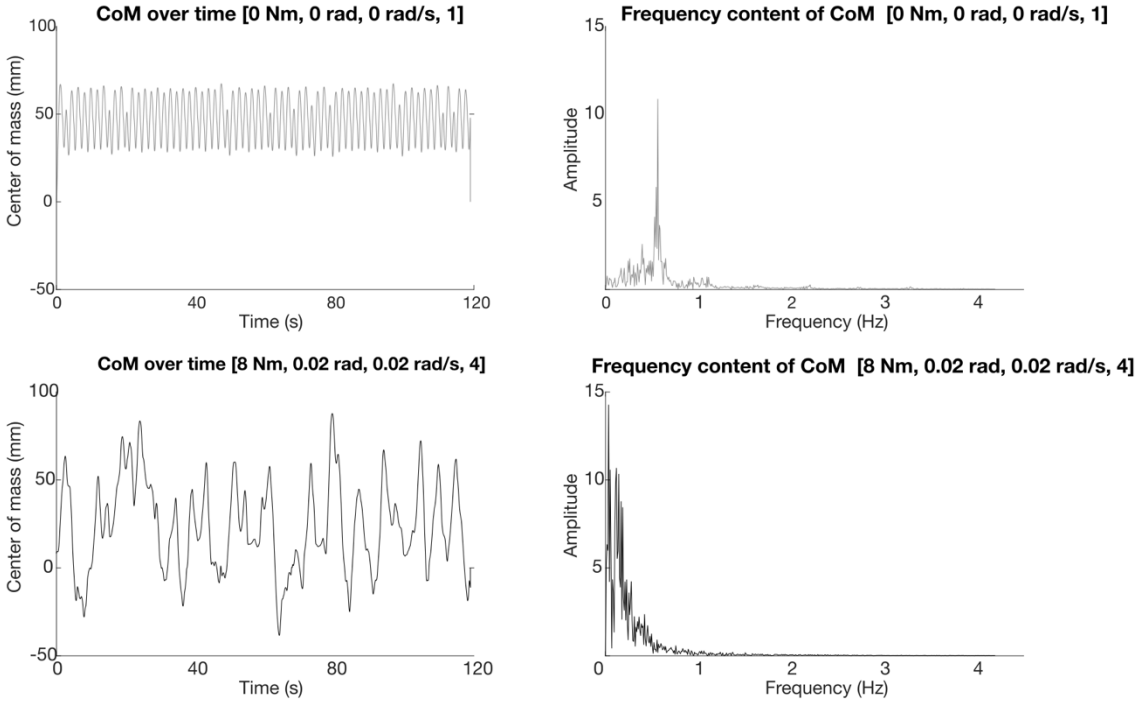


Figure 5. Example MDP simulated data (i.e., body angle) in time and frequency domain. The top row shows a simulation (0 Nm, 0 rad, 0 rad/s, 0) that differs from physiological control of balance and the bottom row shows the simulation (8 Nm, 0.02 rad, 0.02 rad/s, 4) that is closer to physiological balance control.

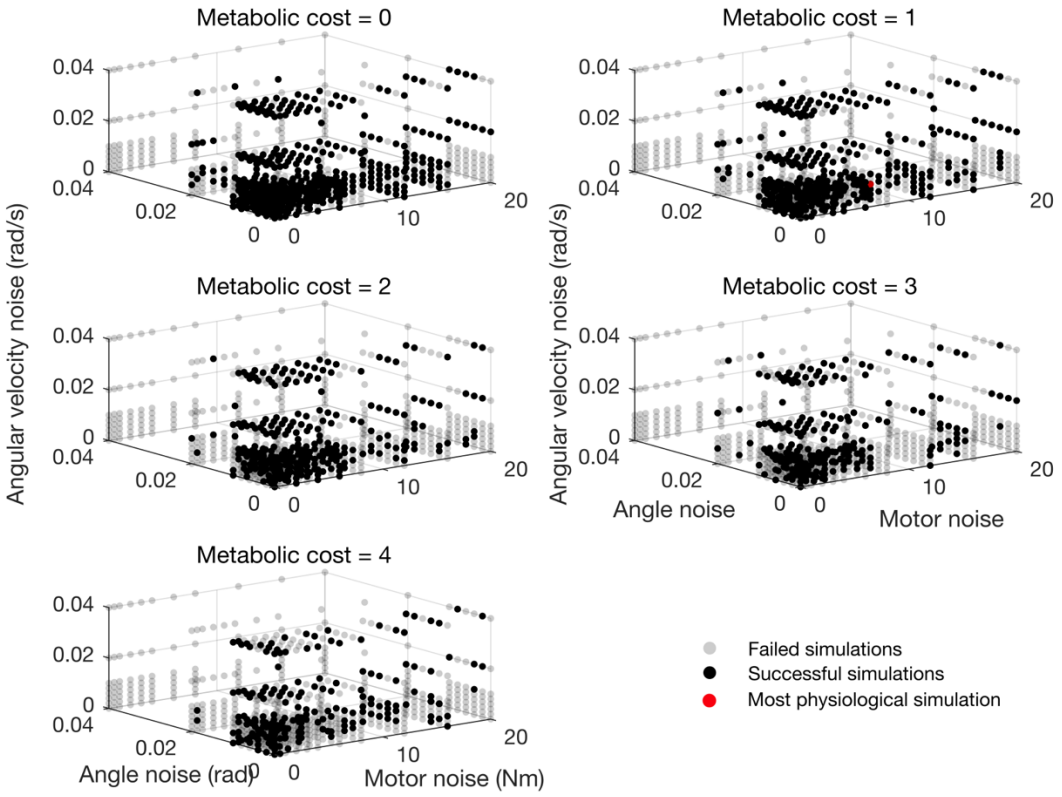


Figure 6. Distribution of successful and failed simulations. Gray represents failed simulations and black represents successful simulations. The red dot represents the simulation that yielded physiological behavior (8 Nm, 0.004 rad, 0.006 rad/s, 1).

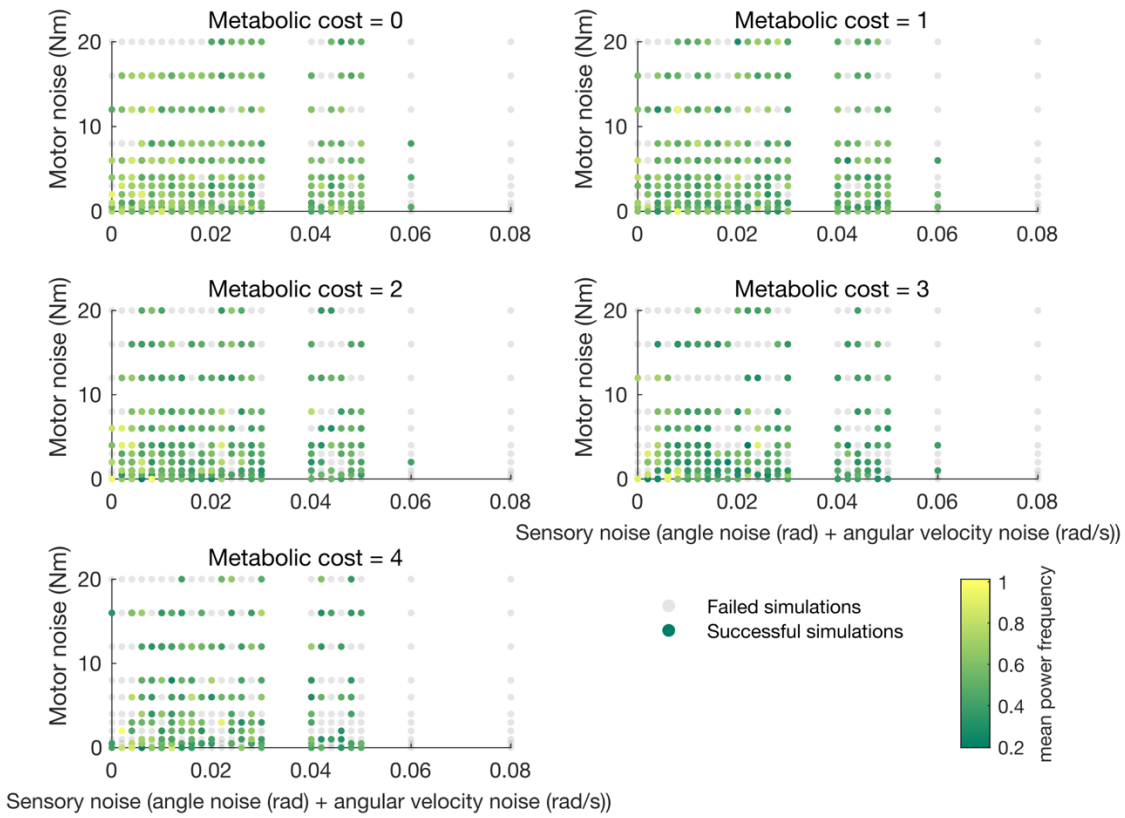


Figure 7. The effect of sensory and motor noises on MPF of the simulation from the MDP model. The x axis is the sensory noise, represented using the sum of the angle noise and the angular velocity noise and y axis is the motor noise. Gray represents failed simulations, and green represent successful simulations with different MPFs.

To further investigate the effect of each parameter on the frequency characteristics of simulated behavior, I conducted linear regression analysis between the parameter values (independent variables) and MPF (dependent variable). Results from multiple linear regression with different combinations of independent variables are presented in Table 3. All of the parameters exhibited negative coefficients after the fit, with high F-statistics, indicating their negative influence on the dependent variables and that the model fitted the data better than an intercept-only model. However, the r square values of the fits were low for all linear regressions

with the highest being 0.12 for a single parameter (metabolic cost), indicating weak linear relationships.

Variable included	Fitted coefficient	R^2	F statistics
Motor noise + intercept	-0.0043, 0.4444	0.013	18.74
Angle noise + intercept	-7.6589, 0.4657	0.028	41.52
Angular velocity noise + intercept	-2.63, 0.4564	0.027	41.17
Metabolic cost + intercept	-0.0526, 0.5036	0.1221	204.60
Motor noise, angle noise, angular velocity noise and metabolic cost noise + intercept	-0.0046, -8.4025, -2.0000, -0.053, 0.6036	0.1875	84.72

Table 3. Multiple regression analysis on MPF

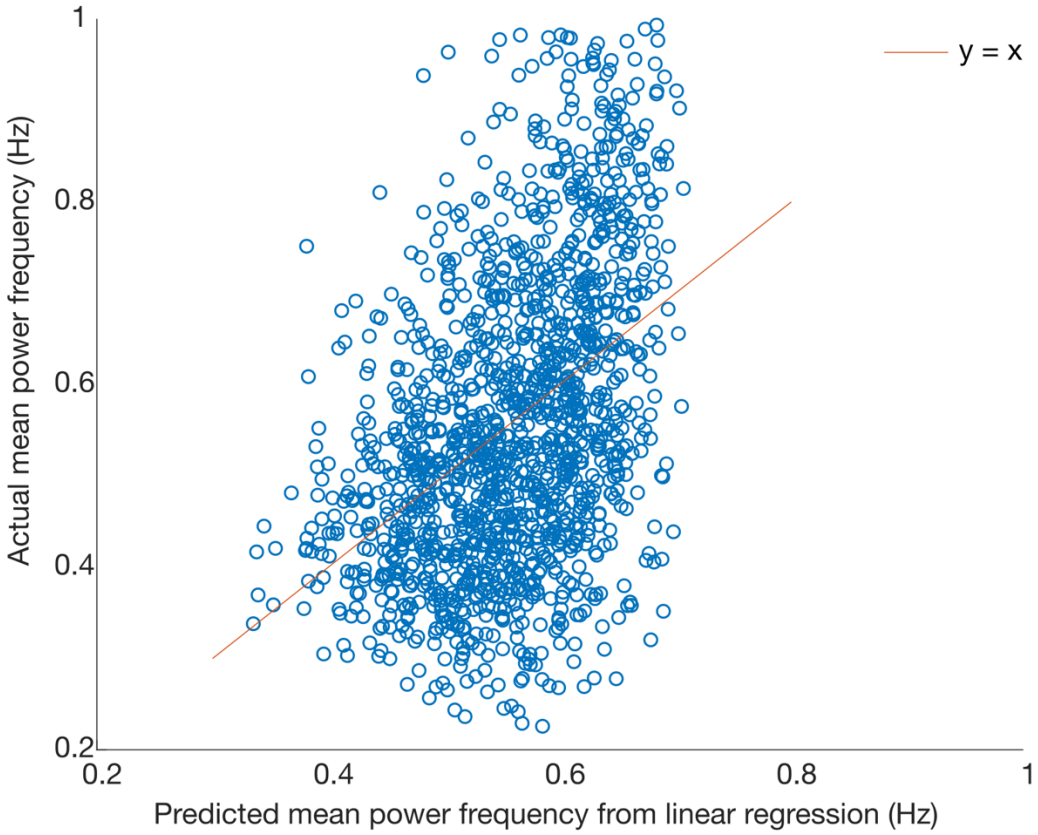


Figure 8. **T**he accuracy of the predicted **MPF** from the variables comparing against the actual **MPF**. The red line represents the line y = x.

Moreover, Figure 8 illustrates the actual MPF plotted against the predicted MPF for all simulations (using all four simulation parameters as predictors), and the linear regression prediction showed low accuracy as the data points were grouped in a cloud with variable distances from the red line (i.e., ideal relationship where the linear regression predicts the data perfectly).

4.1.2 Predictions from the MDP model to imposed torque perturbation

To address the second hypothesis in my thesis, I simulated the behavior of the MDP framework when the torque bias term was added to the dynamics of human standing balance using the parameter combination that generated simulation close to human balancing behavior

(see above: 8 Nm, 0.004 rad, 0.006 rad/s, 1). This was performed to replicate the experimental conditions and provide quantitative predictions regarding how an MDP controller with a penalty applied to the generated torque would react to applied torque biases.

The simulation results of the MDP model showed that when positive perturbations were added to the dynamics, the behavior of the MDP controller would shift the angle backward. This was consistent with my hypothesis. A negative perturbation $T_{pert} = -2.1\%\text{mgl}$ added to the dynamics shifted the body angle forward while $T_{pert} = -1.05\%\text{mgl}$ did not. Moreover, the size of the angle shift was not always proportional to the amplitude of the torque perturbation added (Table 4). Similar observations can be made from Figure 10, where it compares the body angle shift from the hypothesis and the MDP simulated behavior. The size of the angle shift when negative torque bias terms were added increased as the amplitude of torque bias term increased (Figure 9 and 10). However, the inverted pendulum leaned 1.82 and 1.63 degrees posteriorly when torque bias terms added were 1.05%mgl and 2.1%mgl respectively, while the predicted angle shift from the hypothesis is 0.6 and 1.2 degrees.

Torque perturbation added	Average posture angle without the perturbation (degree)	Average posture angle after the perturbation (degree)	Average difference ($\theta_{pert} - \theta_0$)
1.05%mgl	1.18 ± 0.37	-0.64 ± 0.35	-1.82 ± 0.19
-1.05%mgl	1.01 ± 0.87	0.72 ± 0.41	-0.29 ± 1.03
2.1%mgl	0.54 ± 0.29	-1.09 ± 0.23	-1.63 ± 0.32
-2.1%mgl	0.14 ± 0.99	1.32 ± 1.04	1.19 ± 1.91

Table 4. Descriptive statistics of the body angle shifts in MDP simulated conditions where torque bias terms were added to the standing balance dynamics

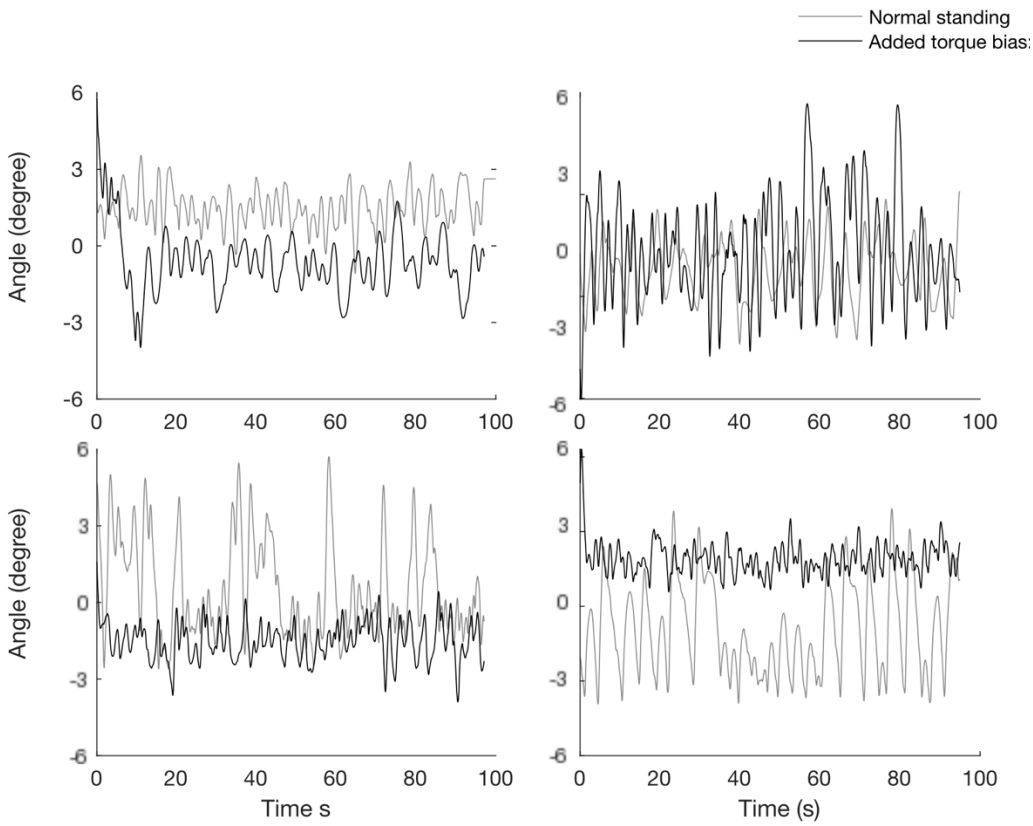


Figure 9. MDP simulated balancing behavior during normal and altered standing dynamics with added torque bias 1.05 %mgl (top left), -1.05 %mgl (top right), 2.1 %mgl (bottom left), and -2.1 %mgl (bottom right).

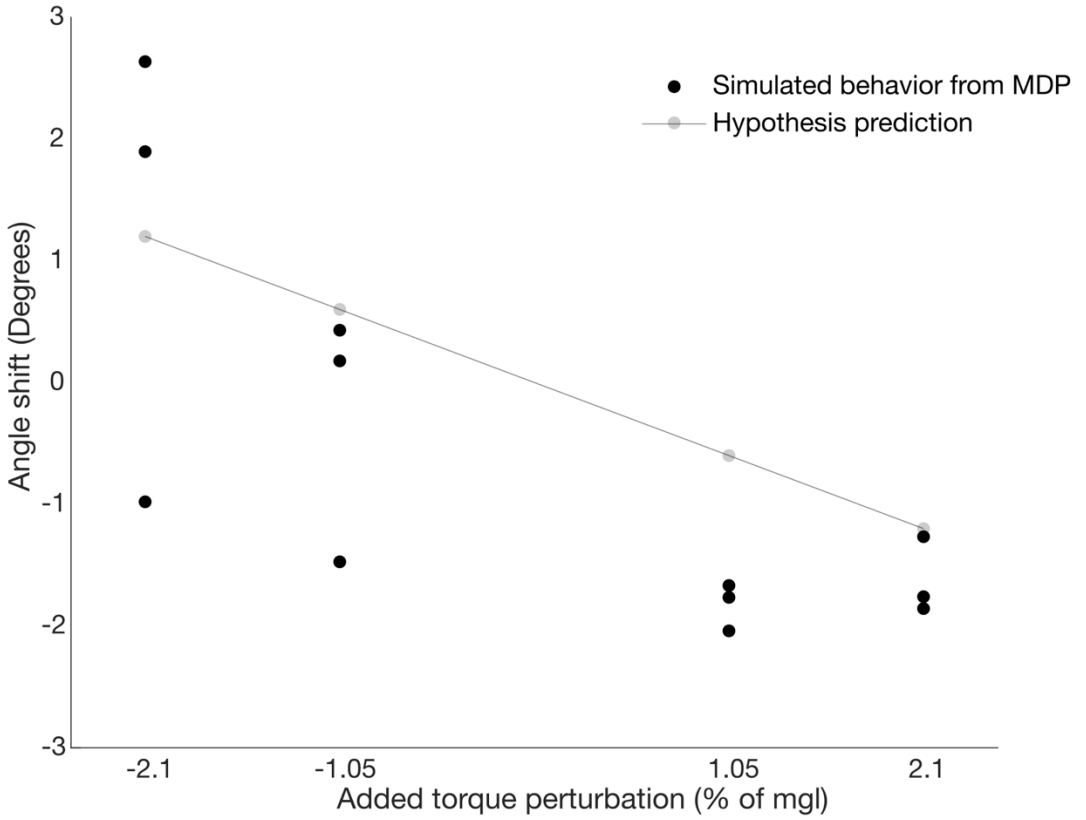


Figure 10. Body angle differences between the altered and normal dynamics from the hypothesis (gray) and from MDP model simulations (black) is plotted against the added torque bias term.

4.2 Experimental results

4.2.1 Quiet standing condition

Participants maintained their body angle at 0.55 ± 0.49, 0.63 ± 0.41 and 0.47 ± 0.28 degrees respectively, indicating that they leaned slightly forward when standing normally without any internal or external perturbations. The variability of body angles in quiet standing trials were similar across the first two participants, and lower for Participant 3 (Table 5).

	Average angle during quiet standing (degrees)		
	Participant 1	Participant 2	Participant 3
Trial 1	0.53 ± 0.88	0.63 ± 0.48	0.55 ± 0.23
Trial 2	0.57 ± 0.33	0.69 ± 0.43	0.57 ± 0.21
Trial 3	0.54 ± 0.25	0.57 ± 0.32	0.30 ± 0.40
Average angle for 3 trials (degrees)	0.55 ± 0.49	0.63 ± 0.41	0.47 ± 0.28

Table 5. Descriptive analysis of quiet standing data.

4.2.2 Perturbation condition: Control Trials

I performed exponential regression to the data obtained from the standing periods after the perturbations to obtain information about the duration of behavior shifts (i.e., time constant) and the predicted steady-state behavior level (i.e., asymptote) using standardized methods for all trials. Results from the exponential regression analysis were used to determine whether the behavior converged to a steady level for each condition. Among the experimental trials (including the control trials), 20 out of 72 of the perturbation trials did not converge and 3 out of 12 control trials did not converge for three Participants in total. In the discarded trials, the body angles did not stabilize to within one standard deviation of the body angle during quiet standing from the predicted asymptote value from the exponential regression.

I then calculated the differences of the preferred body angles between the first and second perturbations in the control trials (i.e., with no added torque bias). The difference between angles for three participants were 0.31 ± 0.14, -0.38 ± 0.53 and 0.05 ± 0.59 degrees respectively. I didn't observe a consistent trend across participants in the preferred body angle differences in the control trials. Participant 1 stabilized at a more forward angle after the 2nd perturbation compared with after the 1st perturbation. Participant 2 ended up at a more backward angle after the 2nd

perturbation compared with after the 1st perturbation, while Participant 3 maintained balance at similar angles after both perturbations. In addition, the variability of preferred body angle differences for Participant 2 and 3 were around three times higher than Participant 1 (Table 6).

	Difference of average angle after 1st and 2nd perturbations (degrees)		
	Participant 1	Participant 2	Participant 3
Mean	0.31	-0.38	0.05
Standard deviation	0.14	0.53	0.59

Table 6. Difference of preferred body angle after 1st and 2nd perturbations in control trials

4.2.3 Perturbation conditions: Experimental trials

After fitting an exponential curve to the body angles in the experimental trials, preferred body angles were extracted from the quiet standing periods after the two perturbations. Differences between the preferred body angles after the 1st and 2nd (with added torque bias) perturbation were calculated and presented along with the predicted body angle differences from the hypothesis in Table 7. The mean squared error (MSE) between the experimentally acquired preferred body angle differences and the prediction from the hypothesis for all experimental conditions for each participant were 0.35, 1.19 and 1.11 degrees respectively.

		Difference in preferred angles (degrees) post perturbation		
Perturbation conditions		Participant 1	Participant 2	Participant 3
	Predicted body angle shift	Mean, SD, and size N (number of trials available)	Mean, SD, and size N (number of trials available)	Mean, SD, and size N (number of trials available)
$mgl * 0.6, \theta = \theta_0$	-0.6	-0.54, 0.18, N=3	0.75, 0.51, N=3	-0.54, NA, N=1
$mgl * 0.6, \theta = \theta_0 - 0.6$	-0.6	-0.91, 0.3, N=3	0.19, NA, N=1	0.63, 0.38, N=2
$mgl * 1.2, \theta = \theta_0$	-1.2	0.07, NA, N=1	0.60, NA, N=1	0.67, 0.26, N=2
$mgl * 1.2, \theta = \theta_0 - 1.2$	-1.2	-1.34, 0.44, N=3	0.29, 0.46, N=3	NA, N = 0
$-mgl * 0.6, \theta = \theta_0$	0.6	0.23, 0.37, N=2	-0.19, 0.47, N=2	0.36, 1.02, N=3
$-mgl * 0.6, \theta = \theta_0 + 0.6$	0.6	1.47, NA, N=1	0.38, 0.67, N=2	0.6, 0.44, N=3
$-mgl * 1.2, \theta = \theta_0$	1.2	1.22, 0.41, N=2	0.62, 0.06, N=3	-0.11, 0.62, N=2
$-mgl * 1.2, \theta = \theta_0 + 1.2$	1.2	1.57, 0.52, N=2	0.44, 0.39, N=3	0.22, 0.54, N=3
Mean squared error (MSE) between actual difference and predicted difference		0.35	1.19	1.11

Table 7. Average angle differences in perturbation trials

To compare the whole-body angle observed experimentally with the predicted angles, I plotted the preferred angle differences in the control trials and experimental trials as open circles together with a line representing the hypothesis (Figure 11a). For Participant 1, the preferred angle differences were positive (negative) when the added torque is negative (positive) and the

size of the preferred angle differences increased as the amplitude of torque bias increased. For Participant 2 and 3, Figure 11(a) showed no clear trends of how the added torque bias terms affected the preferred angle difference.

Next, I used the preferred angle after the first perturbation to establish the baseline mean and standard deviation of the preferred body angles (results presented in Table 8). After the first perturbation where no torque bias term was added, participants maintained their body angles (i.e., preferred body angles) at 0.36 ± 0.35, 0.63 ± 0.55 and 0.39 ± 0.45 degrees respectively. The preferred body angle under normal dynamics of all participants were all positive, indicating a forward standing posture. The variability of the preferred angles for Participant 2 and 3 were higher than Participant 1, resulting in a wider 95% confidence interval.

	Participant 1	Participant 2	Participant 3
Mean	0.3647	0.6338	0.3942
Standard deviation	0.3472	0.5539	0.4535
95% confidence interval	0.2043, 0.5251	0.3779, 0.8897	0.1719, 0.6164

Table 8. Mean and variability of baseline measurement of preferred angle, calculated from the first perturbations of the experimental trials.

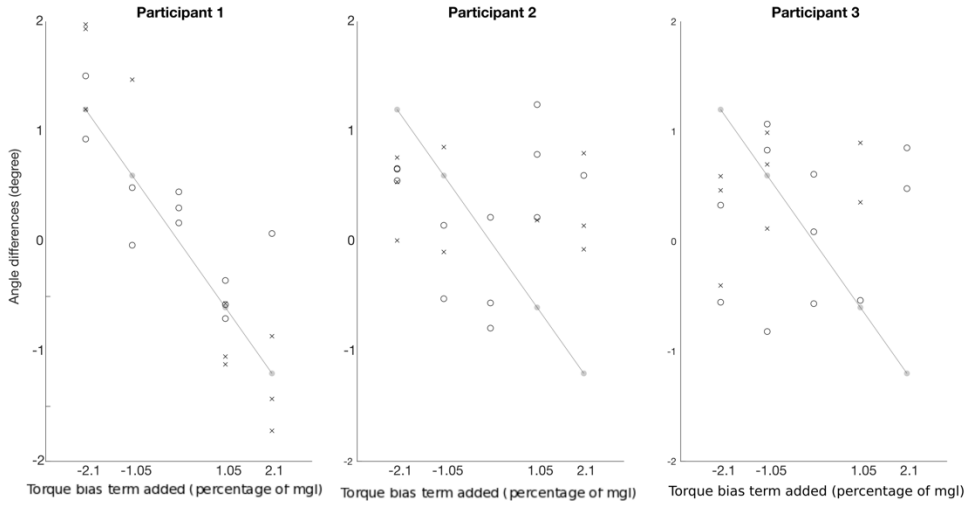
Then, I looked at the new preferred angles under altered dynamics in contrast to normal dynamics for all participants. To obtain a global relationship of how the added torque bias terms affected the preferred body angles, I fitted linear regressions with the preferred body angles from experimental conditions and compared them with the predicted outcomes from my second hypothesis (Figure 11b). Specifically, my second hypothesis states that participants will maintain their torque when torque bias terms are added to the control of balance, exhibiting a horizontal line that passes through the preferred angle during quiet standing on the torque-angle graph (See

Figure 4). Table 9 shows the linear regression results from all three participants using different training data (data from all experimental conditions, the subset of preferred angle trials and the subset of preferred torque trials). The slopes of the fitted lines using all experimental conditions were -2.23, -8.27, and -13.8 for all participants. The 95% confidence interval of the slopes (all experimental conditions) for participants 1 and 2 were mostly in the negative range and was in the negative range completely for participant 3. Finally, to examine potential behavioral differences when participants were brought back to their preferred angle or preferred torque, I compared the participants' responses to the added torque bias for the preferred angle and torque trials separately. In this analysis, participants 2 and 3 exhibited steeper slopes (more negative) of the regression when the regression used only the preferred torque trials compared with the preferred angle trials.

Regarding variability of the fitted linear regression, the root mean squared error (RMSE) value of the fit varied across individuals: participant 1 showed a lower RMSE value than participants 2 and 3 for all linear regression fits.

		Participant 1	Participant 2	Participant 3
LR with all experimental conditions	coefficients	y = -2.23x – 5.12	Y = -8.27x – 1.93	Y = -13.80x+4.65
	95% CI of slope	(-4.76, 0.29)	(-17.14, 0.61)	(-21.06, -6.54)
	RMSE	5.25	11.00	9.77
LR with preferred angle conditions	coefficients	Y = -3.00 x -5.27	Y = -4.68x-5.05	Y = -10.06x+3.03
	95% CI of slope	(-5.90, -0.10)	(-16.64, 7.28)	(-19.52, -0.60)
	RMSE	5.00	10.74	8.31
LR with preferred torque conditions	coefficients	Y = -3.48x-4.70	Y = -12.31x + 2.17	Y = -19.59x+4.15
	95% CI of slope	(-9.5124, 2.5587)	(-22.54, -2.08)	(-30.70, -8.49)
	RMSE	6.73	10.10	9.29

Table 9. Linear regression results using the preferred angles after the 2nd perturbation in the experimental trials.



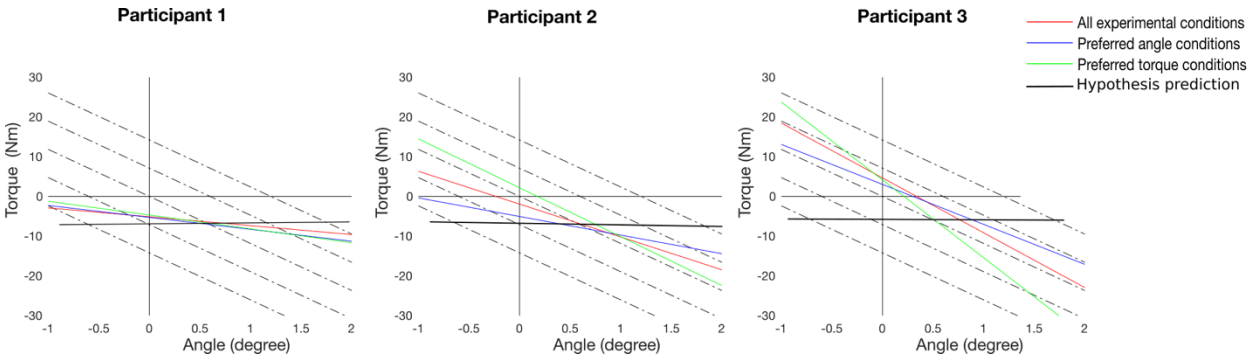


Figure 11 (Top): preferred angle differences are plotted against the added torque bias for participant 1, 2 and 3 from left to right. The dark line represents predicted angle differences from my second hypothesis. Figure 11(Bottom): fitted linear regressions using all of the experimental trials (red) (i.e., both the preferred angle condition as well as the preferred torque condition), only the preferred angle condition (blue) and the preferred torque condition (green) respectively. All the dotted lines represent the load-stiffness curved in the experimental conditions.

5 Discussion

The primary purposes of this thesis were to (a) model the physiological control of human standing balance with an MDP framework and identify the best parameter combinations of the environment that yields simulated behavior close to empirical evidence, (b) investigate how the MDP controller and human participants respond to changes in biomechanics of standing (i.e., added torque bias terms). In Chapter 3 and 4, I presented the MDP framework to model human standing balance and the simulated behavior from the MDP model. With carefully selected parameter combination (i.e., [8 Nm, 0.004 rad, 0.006 rad/s, 1]), the simulated behavior from the model resembled the frequency characteristics (i.e., MPF, 99%frequency bandwidth) reported from empirical studies, while the discrepancy in time-series characteristics (i.e., range and RMS) was minimized. When simulating the behavior to the addition of a torque bias term to the load-stiffness relationship with the MDP model, the predictions showed directional whole-body movements supporting my hypothesis but quantitatively differed from the hypothesis. The amplitude of the body angle shift when positive torque bias terms were added were larger than the predictions from hypothesis. I also observed participants' behavioral responses to the added torque bias terms experimentally. Experimental data from participant 1 showed a posterior (anterior) lean in body angle when positive (negative) torque bias terms were added, which aligned with the direction of body angle shift in my hypothesis. However, the regressed lines of the preferred body angles after the 2nd perturbation in the experimental conditions were not horizontal and had negative slopes. While the 95% confidence intervals of the slopes covered my hypothesis (i.e., slope = 0) for participant 1 and 2, the intervals were wide and were mostly in the negative range. These negative slopes did not fully match my hypothesis that they would

maintain their ankle torque the when the torque bias terms were added, as the hypothesis predicts a flat horizontal line (Figure 4).

5.1 Computational approach

Overall, the four parameters of the standing balance environment (i.e., motor noise, angle noise, angular noise and metabolic cost) shaped the simulated outcomes from the MDP controller. With selective parameters, the controller generated simulations that resembled human quiet standing behavior, as the frequency characteristics (i.e., MPF and 99% frequency bandwidth) fell within the physiological range reported from the literature. The range and RMS measure computed for all the simulations were out of range comparing with the reported data from the literature. The smallest range and RMS values obtained from the simulations were about three times that of the empirical data (Hasan et al., 1996). This indicates that simulated movement of CoM were more variable than natural human balancing behavior. Several potential mechanisms could be responsible for this: (1) the metabolic cost penalty on the system might not be high enough to restrict the generated torque. The metabolic cost places additional penalty on the amplitudes of torque, and higher metabolic cost forces the agent to select actions with lower amplitude in order to maximize the rewards; (2) lack of penalty placed on the jerkiness of actions (i.e., torque) introduces sudden changes to the movement trajectories, and thereby might affect the size of body sway; (3) the extent of exploratory behavior can also affect body movements as the agent was forced to select random actions at the probability of epsilon because of the epsilon-greedy algorithm. The random action could be completely different from the optimal action and the recent history of selected actions, causing sudden changes to the sway trajectories.

When investigating the influence of those parameters on simulated behavior, I found that model simulations failed over extreme parameter values (i.e., large values) more frequently. At

some extreme values such as when angle noise is 0.04 radians, the simulations always failed (Figure 6). It is reasonable as (1) increasing noises (sensory and motor noises) placed on the system increases the unpredictability and randomness (Bottaro et al., 2005) and (2) increased metabolic cost places more penalties on the amplitudes of torque, which forces the agent to select smaller torques to obtain higher rewards. Smaller torques are not sufficient to counteract gravitational pulls when body angles are large.

Successful simulations were denser in regions with lower parameter values and were sparser when parameter values increased. Interestingly, there was no fixed threshold below which simulation is guaranteed success. This is likely due to the stochastic nature of the decision-making process in Q learning. The action selection process in Q-learning uses the argmax function, which only selects one action with the highest value for each state (Sutton & Barto, 1998). However, in states where several actions have similar action values, a small update of Q value might change the value functions of those actions slightly but might affect the rankings of actions dramatically. This could be the underlying mechanism of why the MDP model failed over certain small parameter values (i.e., 1 Nm, 0.002 rad, 0.004 rad/s, 1).

The MDP simulations with physiological frequency characteristics were mostly obtained from non-zero sensory and motor parameters (Figure 6 and 7). This agrees with the commonly accepted notion that human central nervous system is inherently noisy (Faisal et al. 2008; Renart & Machens 2014; Stein et al. 2005). It also agrees with the PID control findings that variability placed on the torque input is necessary for generating low-frequency oscillations in the controller behavior (Bottaro et al., 2005). It is noticeable that there was no clear linear relationship between the MDP parameters and the frequency characteristics (i.e., MPF and 99% bandwidth). Two potential mechanisms might be responsible for this: (1) the system dynamics (i.e., interactions

with the environment and state propagation) was non-linear due to the delay and muscle dynamics (details explained in Appendix A) (2) the reward feedback was also non-linear as it consisted of a linear penalty on the torque and a non-linear step function of fall penalty.

5.2 Experimental study

Unlike goal-directed motor tasks such as reaching and sit-to-stand movements in which the task-specific goals are clearly defined, the goal of continuous tasks such as locomotion and standing is more implicit. Research evidence has suggested that during locomotion people aim to minimize over their overall energy consumption (Selinger et al., 2015; Selinger et al., 2019). In standing, it was also argued that participants' responses to perturbations during standing best matched the optimal feedback from minimizing joint torque (Kiemel et al., 2011). One of the reasons that the goal of standing is so hard to determine is that in normal standing the outcomes of minimizing angle and torque both lead to the same behavior: balancing near the upright posture. In other words, they cannot be disassociated. With the help of computational simulations and our custom-made robotic balancing platform, I was able to modify the landscape of torque-angle relationship and manipulate the internal control loop of standing balance. When torque bias terms were added to the standing dynamics, the MDP controller shifted its body angle in the same direction as the hypothesis prediction, except for the $-1.05\%\ mgl$ condition. However, when positive torque bias terms (1.05%mgl and 2.1%mgl) were added, the size of the angle shift (i.e., -1.82 and -1.62 degrees respectively) differed from the hypothesis (i.e., -0.6 and -1.2 degrees). Therefore, the MDP simulated data aligned with the predicted direction of body angle shift from the hypothesis, but did not agree with the prediction that torque levels would be maintained.

By adding an offset to the relationship, I could dissociate the strategy in which the angle was minimized or the torque was minimized, enabling examination of how people choose between these strategies, reflecting the underlying goal during standing. I hypothesized that humans would maintain their preferred torque when the standing balance dynamics was altered. Fitted linear regression using raw data from participants 1 and 2 agreed with the hypothesis partially as the direction of the angle shift when torque bias terms were added aligned with the hypothesis (Figure 11b). More specifically, their body angles shifted backward when positive added bias terms were added and shifted forward when negative added bias terms were added. The 95% confidence intervals of the fitted linear regression slopes were mostly in the negative range for participants 1 and 2 and completely negative for participant 3 while my hypothesis predicted a horizontal line (i.e., slope = 0), indicating the torques at the ankle joint were not always maintained (Figure 4). The observed trends of participants' behavior can be interpreted in two ways. First, if all participants have explored the new environment (i.e., altered dynamics) fully, then the results indicate that participants balanced at a different angle from the hypothesis prediction and did not maintain the same level of torque when balancing under the altered dynamics. Thus, the hypothesis that humans minimize their torque during standing would not be supported. Second, if the participants did not explore all of the state and action space under the new dynamics, then the results could be interpreted as suboptimal behaviors as participants could be still exploring at the end of the trials. This interpretation was partially supported by the observation that ~20% - 25% of the trials were considered not converged by my definition of convergence, as their body angles were still decreasing or increasing at the end of the trial. Given that participants only balanced for 90 seconds after each perturbation, it might not be enough for every participant to reach a steady level.

The baseline variability of preferred angles of participants 2 and 3 was higher than participant 1. Behavior after perturbations from participants 2 and 3 also showed higher variability, indicated by the RMSE values from the linear regressions and sparser distributions of preferred angle differences in Figure 11a. There are two interpretations for the differences in variabilities among participants. The first one is that participants 2 and 3 might have considered a wider range of angles equally comfortable to maintain their balance at. Thus, regardless of the changed torque-angle relationship their preferred body angles would be expected to exhibit wider ranges. The other potential mechanism is that participants 2 and 3 did not explore the consequences of their actions under the new dynamics fully at the end of trials and thus their behavior could not represent the steady-state behavior. This interpretation is supported by recent notions and evidence in the motor learning field that high behavioral variability has been linked to initial stage of learning where a larger number of exploratory actions take place (Wu, et al., 2014).

Another interesting observation is that the slopes of the linear regression were more negative when using data from the preferred torque trials than preferred angle trials for participants 2 and 3. This indicates that the weights participants place on maintaining torque and angle were different when they were brought to their preferred angle and their preferred torque. However, there were significant overlaps in the confidence intervals of the fitted slope. Therefore, I am uncertain whether this was due to chance because participants were still in the process of exploring the new dynamics or the preferred angle and preferred torque trials indeed affected participants' behavior differently.

5.3 Strengths of the approach

As stated before, not only are MDP models able to simulate behaviors in which goals are clearly defined (like traditional feedback-based controllers), they can also represent the exploratory behavior that those feedback-based controllers seem to lack. Another advantage of MDP is its ability to handle non-linear dynamics with non-linear reward function setup. To achieve biological fidelity of modeling, the representation of the system will inevitably include some non-linear components. In the computational project, the muscle dynamics (Appendix A) and sensorimotor delays are examples of non-linear components.

A physiological MDP model can generate testable predictions, which enable comparisons with the empirical approach. The experimental approach tested the predictions that target the underlying goal of standing balance directly. The robotic balance simulator created a unique opportunity that separated the balancing strategies of maintaining the same angle and maintaining the same torque, which is impossible in daily-life normal standing where standing perfectly upright is always associated with minimal ankle torque. This experimental setup creates a novel standing balance dynamics landscape that helps investigate how humans respond to the modified sensorimotor loop of balance.

5.4 Limitations

In the current simulated environment of human standing balance, I discretized the action and state spaces, as required by the Q-learning algorithms. As a result, the action selection process of Q-learning algorithm utilized the “hardmax” approach, where only one best action was selected without acknowledging the size of differences among action values. This could explain the stochasticity observed in the parameter space. Another limitation with the current MDP framework is that the learning was formalized as episodic tasks with clear definitions of

beginning and end. Each training episode ended after a fixed number of time steps (i.e., 2 minutes) despite of successful performance. Therefore, this prevented me from monitoring the system behavior continuously and thus I could only compare the individual training episodes when I examined how behavior shifts under changing environmental dynamics, instead of having continuous trends of behavior changes in a long episode.

One limitation of the experiment is that the duration of standing period after perturbations was short (i.e., 90 seconds). Some participants demonstrated variable behavior under the same experimental condition, which could be because they were still exploring the consequences of their actions at the end of the trials, and they haven't found the new behavioral optimum. Therefore, the demonstrated behavior pattern after the added torque bias term could not simply be the result of the altered dynamics. With the current experimental paradigm, I am unable to conclude whether their behavior reflects the true goal of standing, or their behavior is suboptimal because of the inability to explore the full dynamics. Another obvious limitation in the experimental approach is the small sample size of participants. I will address this problem by continuing with the experiment (with longer trial durations to deal with the first limitation) and recruiting more participants when the lab equipment is fixed.

In the experimental approach, I added two amplitudes of added torque bias terms: $1.05\%\ mgl$ and $2.10\%\ mgl$. Based on the hypothesis of maintaining torque, I expected these added torque terms would shift the body angle by 0.6 and 1.2 degrees respectively. With four levels of added torque bias terms (both positive and negative), I could obtain the global relationship (i.e., rule) of how body angle was affected when the torque bias changed. However, the effect of the added torque bias on participants perception at each level of amplitude was unknown. This raises a concern as whether the participants were able to consciously perceive the

torque bias and how this might affect their strategy of adjustment for their posture (i.e., conscious adjustment versus involuntary adjustment).

One limitation on the data analysis is that I used one linear regression to fit the data for both the positive and negative added bias terms. I was limited by the number of conditions for positive and negative added torque bias terms each (i.e., 2 conditions for each), therefore I could not perform separate linear regression analysis for them. With one linear regression, I assumed by default that behavior shifts would be consistent for positive and negative added torques. The hypothesis of the experiment states that positive and negative added torque bias terms will move the body angle backward and forward respectively. However, humans naturally favor and stand in a slightly forward posture (Cotton, 1931; Smith, 1957). Therefore, this assumption might not be accounting for the differences of goals when the posture is forward or backward. Future studies can address this issue by including more torque bias terms and fitting separate linear regressions with data from conditions with positive and negative bias terms.

5.5 Future directions

Robustness of the Q-learning model needs to be examined and quantified over a wider range of parameter. Example robustness check is to repeatedly run the same parameters and compare whether they all yielded similar outputs. There are two purposes for this. First, it will show whether the selected physiological parameters can yield consistent behavior. Second, the robustness check might provide answers regarding why some small parameters combinations failed while relatively large parameters yielded successful simulation results. This is crucial for two reasons: (1) it is necessary to determine whether the obtained policy is optimal in terms of the defined reward function and (2) sometimes an MDP problem will have several equally optimal but distinct policies.

Future work on the model can be directed at four main directions. First, one of the limitations of the model presented in this thesis is that the range and RMS of body angle exceeded the physiological limits of human quiet standing. The next step on the model could be to investigate physiologically feasible approaches to restrict the variability of body angle (i.e., manipulating the metabolic cost and exploration factors or incorporating additional penalties such as smoothness/continuity of actions). Second, the sensory system dynamics (i.e., somatosensory, visual, vestibular and auditory systems) can be incorporated into the model to replace the current state variables (i.e., angular displacement and velocity). This will lead to a direct mapping between the sensory code and motor commands. Third, the current model represents actions as one-dimensional variables at the biomechanical level (i.e., torque) as the net output for the muscles. The physiological control of standing involves controlling muscle activations in the lower limbs (i.e., plantar flexors and dorsiflexors) in the AP direction. Therefore, it would be more physiological to replace the torque control with activation patterns of the plantar and dorsiflexors. Both the second and third directions aim to improve the physiological resemblance to human balancing behavior. Finally, to address the limitations of the Q-learning algorithm (i.e., argmax function and discretized action and state spaces), policy gradient methods can be used to solve for the optimal policy of the standing balance environment, as it operates on continuous and high-dimensional state and action spaces, which are inevitable results if the second and third point are implemented (Sutton & Barto, 1998).

Follow-up experimental studies can address the issue of degrees of exploration when participants were exposed to altered dynamics mentioned in the above section. One possible approach is to force the participants to explore the action space. For example, participants can be instructed and cued to maintain balance at several previously determined body angles to

voluntarily explore the new dynamics landscape. Then they can be asked to select the preferred posture to balance. This is similar to the experiment performed by Selinger and colleagues (2019) to investigate the new optima when the energy landscape of walking was changed. They asked participants to experience walking at different speeds for some time, and then asked them to select the walking speed they were comfortable with. With forced exploration, they would be expected to gain knowledge of the new dynamics equally. The final preferred posture identified by them may reflect the underlying goal of standing, and not affected by the degree of exploration.

5.6 Conclusion

In conclusion, the MDP model in my thesis simulated standing behavior similar to the human quiet standing data with some differences. With the novel robotic balancing platform, I dissociated the strategy of maintaining body angle and ankle torque in standing. The experimental findings did not fully support my hypothesis that humans maintain their torque at the same level when standing balance dynamics is altered. Overall, my thesis provides a physiological controller that replicates characteristics of standing balance which, coupled with a novel robotic balance simulator, show promise for direct investigation on the goal of standing.

Works Cited

Asai, Y., Tasaka, Y., Nomura, K., Nomura, T., Casadio, M., & Morasso, P. (2009). A model of postural control in quiet standing: Robust compensation of delay-induced instability using intermittent activation of feedback control. *PLoS ONE*, *4*(7). https://doi.org/10.1371/journal.pone.0006169

Carpenter, M. G., Murnaghan, C. D., & Inglis, J. T. (2010). Shifting the balance: Evidence of an exploratory role for postural sway. *Neuroscience*, *171*(1), 196–204. https://doi.org/10.1016/j.neuroscience.2010.08.030

Fitzpatrick, R. C., Taylor, J. L., & McCloskey, D. I. (1992). Ankle stiffness of standing humans in response to imperceptible perturbation: reflex and task-dependent components. *The Journal of Physiology*, *454*(1), 533–547. https://doi.org/10.1113/jphysiol.1992.sp019278

Fitzpatrick, R., & McCloskey, D. I. (1994). Proprioceptive, visual and vestibular thresholds for the perception of sway during standing in humans. In *Journal of Physiology*.

Gawthrop, P., Loram, I., Lakie, M., & Gollee, H. (2011). Intermittent control: A computational theory of human control. *Biological Cybernetics*, *104*(1–2), 31–51. https://doi.org/10.1007/s00422-010-0416-4

Hasan, S. S., Robin, D. W., Szurkus, D. C., Ashmead, D. H., Peterson, S. W., & Shiavi, R. G. (1996). Simultaneous measurement of body center of pressure and center of gravity during upright stance. Part I: Methods. *Gait and Posture*, *4*(1), 1–10. https://doi.org/10.1016/0966-6362(95)01030-0

Hidenori, K., & Jiang, Y. (2006). A PID model of human balance keeping. *IEEE Control Systems*, *26*(6), 18–23. https://doi.org/10.1109/MCS.2006.252809

Huryn, T. P., Blouin, J. S., Croft, E. A., Koehle, M. S., & Van Der Loos, H. F. M. (2014).

Experimental performance evaluation of human balance control models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *22*(6), 1115–1127. https://doi.org/10.1109/TNSRE.2014.2318351

Jamali, S., Taghvaei, S., & Haghpanah, S. A. (2018). *Optimal Strategy for Sit-to-Stand Movement Using Reinforcement Learning*. *3*(2017), 70–75.

Kiemel, T., Zhang, Y., & Jeka, J. J. (2011). Identification of neural feedback for upright stance in humans: Stabilization rather than sway minimization. *Journal of Neuroscience*, *31*(42), 15144–15153. https://doi.org/10.1523/JNEUROSCI.1013-11.2011

Kuo, A. D. (1995). An Optimal Control Model for Analyzing Human Postural Balance. *IEEE Transactions on Biomedical Engineering*, *42*(1), 87–101. https://doi.org/10.1109/10.362914

Li, Y., Levine, W. S., & Loeb, G. E. (2012). A two-joint human posture control model with realistic neural delays. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *20*(5), 738–748. https://doi.org/10.1109/TNSRE.2012.2199333

Lockhart, D. B., & Ting, L. H. (2007). Optimal sensorimotor transformations for balance. *Nature Neuroscience*, *10*(10), 1329–1336. https://doi.org/10.1038/nn1986

Loram, I. D., Gollee, H., Lakie, M., & Gawthrop, P. J. (2011). Human control of an inverted pendulum: Is continuous control necessary? Is intermittent control effective? Is intermittent control physiological? *Journal of Physiology*, *589*(2), 307–324. https://doi.org/10.1113/jphysiol.2010.194712

Loram, I. D., & Lakie, M. (2002). Human balancing of an inverted pendulum: Position control by small, ballistic-like, throw and catch movements. *Journal of Physiology*, *540*(3), 1111–1124. https://doi.org/10.1113/jphysiol.2001.013077

McClenaghan, B. A., Williams, H. G., Dickerson, J., Dowda, M., Thombs, L., & Eleazer, P.

(1996). Spectral characteristics of aging postural control. *Gait and Posture*, *4*(2), 112–121. https://doi.org/10.1016/0966-6362(95)01040-8

Michimoto, K., Suzuki, Y., Kiyono, K., Kobayashi, Y., Morasso, P., & Nomura, T. (2016). Reinforcement learning for stabilizing an inverted pendulum naturally leads to intermittent feedback control as in human quiet standing. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2016-Octob*(1), 37–40. https://doi.org/10.1109/EMBC.2016.7590634

Peterka, R. J. (2002). Sensorimotor integration in human postural control. *Journal of Neurophysiology*, *88*(3), 1097–1118. https://doi.org/10.1152/jn.2002.88.3.1097

Popovych, O. V., Manos, T., Hoffstaedter, F., & Eickhoff, S. B. (2019). What can computational models contribute to neuroimaging data analytics? *Frontiers in Systems Neuroscience*, *12*(January), 1–9. https://doi.org/10.3389/fnsys.2018.00068

Rescorla, M. (2016). Bayesian Sensorimotor Psychology. *Mind and Language*, *31*(1), 3–36. https://doi.org/10.1111/mila.12093

Robinson, D. A. (2011). Control of Eye Movements. In *Handbook of physiology* (pp. 419–430). https://doi.org/10.1093/med/9780190228958.003.0025

Selinger, J. C., Wong, J. D., Simha, S. N., & Donelan, J. M. (2019). How humans initiate energy optimization and converge on their optimal gaits. *Journal of Experimental Biology*, *222*(19). https://doi.org/10.1242/jeb.198234

Soames, R. W., & Atha, J. (1982). The spectral characteristics of postural sway behaviour. *European Journal of Applied Physiology and Occupational Physiology*, *49*(2), 169–177. https://doi.org/10.1007/BF02334065

Van Der Kooij, H., Jacobs, R., Koopman, B., & Van Der Helm, F. (2001). An adaptive model of

sensory integration in a dynamic environment applied to human stance control. *Biological Cybernetics*, *84*(2), 103–115. https://doi.org/10.1007/s004220000196

Van Der Kooij, H., & Peterka, R. J. (2011). Non-linear stimulus-response behavior of the human stance control system is predicted by optimization of a system with sensory and motor noise. *Journal of Computational Neuroscience*, *30*(3), 759–778. https://doi.org/10.1007/s10827-010-0291-y

Welch, T. D. J., & Ting, L. H. (2008). A feedback model reproduces muscle activity during human postural responses to support-surface translations. *Journal of Neurophysiology*, *99*(2), 1032–1038. https://doi.org/10.1152/jn.01110.2007

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience*, *17*(2), 312–321. https://doi.org/10.1038/nn.3616

Yoshikawa, N., Suzuki, Y., Kiyono, K., & Nomura, T. (2016). Intermittent feedback-control strategy for stabilizing inverted pendulum on manually controlled cart as analogy to human stick balancing. *Frontiers in Computational Neuroscience*, *10*(APR). https://doi.org/10.3389/fncom.2016.00034

Appendices

A. Q-learning algorithm – pseudo code

1. Initiate an action value vector $Q(s, a)$ for all state action pairs
2. Set up epsilon vector for the total number of trials. This dictates the amount of exploration during the training phase
3. Repeat for each trial (episode):
   (a) Choose a random number between 0 and 1. If the number is below epsilon, then select a random action. Otherwise, select the action with the highest action value in the current state.
   (b) Simulate the next state s' using the dynamics of the inverted pendulum given the action selected and the current state
   (c) Receive the reward for the current state
   (d) Update the action value vector $Q(s, a) = Q(s, a) + \alpha[r + \gamma max Q(s', a') - Q(s, a)]$

   One training episode ends when failure happens (i.e. the inverted pendulum falls) or it has remained upright and not fallen for a certain number of time steps (in my code, the number is 30000 (i.e., 120 seconds))

Until the changes in action value vectors plateau.

B. Simulation of environment

The MDP model was trained and simulated in Python (version: 3.7; open source) using the Jupyter Notebook, Anaconda (Anaconda Inc, USA). The simulation data from the MDP model were analyzed in Jupyter Notebook, Anaconda (Anaconda Inc, USA) as well.

(1) Dynamics of the human plant (physics law of standing)

The dynamics of the inverted pendulum was programmed in the Jupyter Notebook environment. The function simulate() yielded the next state based on the current state and action by numerically integrating the differential equation using Forward Euler Method.

$$I\ddot{\theta} + b\dot{\theta} - mgLsin\theta = T$$

(2) Muscle dynamics

The muscle dynamics block was represented as a combination of a first-order dynamics equation (Thelen, 2003). This equation relates the rate of change in muscle activation to muscle excitation. In the muscle dynamics block, the time delay is introduced to the system depending on the frequency of muscle activation. The dynamics resembles two low pass filters with different time constants (i.e., 10 ms and 40 ms) for muscle activation and deactivation respectively.

$$\frac{da}{dt} = \frac{u - a}{T(a,u)}$$

$$T(a,u) = \begin{cases} 0.01(0.5 + 1.5a), if u > a \\ \dfrac{0.04}{1.5 + 1.5a}, if u \leq a \end{cases}$$

Here, u and a are the excitation and activation of the muscles respectively.