Evaluating Optimality in the Control of Wearable Robotic Devices

by

Kimberly A. Ingraham

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering)
in the University of Michigan
2021

Doctoral Committee:

Professor C. David Remy, Co-Chair
Assistant Professor Elliott Rouse, Co-Chair
Associate Professor Deanna Gates
Professor Noel Perkins

Kimberly A. Ingraham

kaingr@umich.edu

ORCID iD: 0000-0001-7695-0633

This dissertation is dedicated to my unmet expectations, and the beautiful reality that emerged instead.

ACKNOWLEDGMENTS

Reflecting on my career so far, it genuinely shocks me how many incredible people I have been fortunate enough to cross paths with. I have had the opportunity to work alongside so many brilliant scientists, dedicated educators, and kind and genuine humans. My space here is limited, so I will never be able to fully acknowledge all the individuals who inspired me along the way. But I know that none of my success would have been possible without the unending support from mentors, friends, and family.

I must start by acknowledging the advisors who have guided me and supported me throughout my graduate career. First, I would like to thank David Remy for taking a chance all those years ago and accepting me into his already "full" lab. At the time, we couldn't have anticipated how things would change, but it has been an honor to work with you and learn from you during each phase of our time together. You always seemed to know when to push, when to support, and when to let me take the reins. Thank you for your continued dedication to my academic training and personal and professional development, despite the time zone difference. Next, to Elliott Rouse, thank you for graciously offering me a new lab home when I needed it most. The end of my graduate career would not have been nearly as fulfilling or productive if you hadn't welcomed me into the Neurobionics group. I have learned so much from you, both technically and professionally (including how to win at Blackjack in Vegas). Thank you for always advocating for your students, we are lucky to have you in our corner.

Although our formal advisor/advisee relationship was relatively short, I am incredibly grateful to Dan Ferris for always having ten minutes to chat. Dan, your perspective and advice are invaluable to me, even if your taste in volleyball teams is questionable. To Deanna Gates, thank you for supporting me through the transitions and for always being so flexible. Thank you for serving on my committee and for sharing your commitment to always center the patient in the work that we

do. And finally, I am very appreciative to my teaching mentors, Alex Shorter and Noel Perkins. Alex, thank you for giving me the freedom to hone my teaching skills and reminding me to always have fun with it. Noel, it was a gift to watch the master at work. Thank you for taking the time to serve on my committee before you move into the well-earned retirement phase of your life.

I would not be where I am today without the experiences I had during my time with the Center for Bionic Medicine. First and foremost, I owe much of my success to Annie Simon, my mentor, role model, and friend. Annie, thank you for giving me my dream job—working for you was an absolute joy during a formative time in my life. Thank you for managing with empathy and teaching me to always do what needed to be done. I am also deeply appreciative to Levi Hargrove for leading by example and giving me the opportunity to grow as a researcher and an individual. To all my friends at RIC, I have the fondest memories of our multilingual lunches, lasagna dinners, and late nights out together. What an unbelievable blessing to have a community of friends and researchers from all over the world. I look forward to many more years of conference reunions.

I am very grateful for the collaborators that helped move this research forward, because I couldn't have done it alone. The work presented in Chapter 2 was an enormous group effort, and I wish to acknowledge Hwan Choi, Emily Gardinier, Kelsey Ebbs, Jeffrey Wensman, the members of the RBL who assisted with data collections, and the individuals who dedicated their time to participate in the study. Thank you to Bryan Schlink for his expertise and help collecting the experimental dataset presented in Chapters 3 and 4. The exoskeleton work presented in Chapter 5 would not have been possible without the support of Dephy. Thank you especially to Luke Mooney and Jean-François Duval, who took calls and responded to emails at all hours of the day. I must also acknowledge Leo Medrano for his collaboration on this project and the long hours of debugging (and his 20-year-old optimism). Finally, I am very thankful for the support of the National Science Foundation Graduate Research Fellowship, which helped me to achieve my research goals.

I am incredibly grateful for the friends and community I have found at the University of Michigan. I would like to give a huge shout-out to all the members of the RAMlab, HNL, RBL, and Neurobionics groups (and an extra thank you to all those who participated in my data collections!).

Grad school is infinitely more fun when you have a wonderful crew of labmates. There are a few friends that I would especially like to acknowledge for their support and friendship throughout my PhD. Misaki Nozawa, my very first friend at U of M, your unyielding support and the copious amounts of caffeine from Songbird and Mighty Good were directly responsible for getting me through classes (and some major real life stuff). Rachel Vitali, thank you for always having my back, no matter what—and for knowing that I always have yours. Jeffrey Koller and Yevgeniy Yesilevskiy, through our conference adventures, lab meeting shenanigans (sorry, David), and support-group chat, you have both been a source of true friendship and constant laughter. Max Shepherd, thank you for being a statistics advisor and a sounding board for research and real life matters alike. And last but not least, Tyler Clites, thank you for being my number one champion and for reading my mind. I look forward to many more years of friendship and collaboration.

Thank you to my partner, Casey Bantle, for making Ann Arbor my home. Thank you even more for reminding me that home is wherever we're together with Maizey. I am so grateful to the entire Bantle family (especially Pat, Ray, and Marge) for welcoming me with open arms, and for supporting me over the years with words, actions, and home-cooked meals. And of course, my time here would not have been the same without the amazing Ann Arbor volleyball community. Thank you to my friends from near and far whose visits and continued love and support made all the difference when transitioning to a new place.

And finally, the biggest thank you of all is to my parents, Dave and Gina Ingraham. My entire life, you have encouraged and supported me to pursue my dreams, and never gave me a reason to think they weren't achieveable. There will never be adequate words to acknowledge the sacrifices you have made that helped me get to where I am today, so all I can say here is thank you, and I love you both deeply. Thank you to my brother Kyle and the rest of my extended family, particularly my grandparents, for supporting me through this process. Without the unyielding love and support of my family, I would not be the person I am today. My great-grandfather Raymond was a Professor of Physiology and the first Dr. Ingraham, and it is an honor to carry on that legacy.

PREFACE

Chapters 2-5 of this dissertation were each written as separate manuscripts. Given this, there may be some repetition between them, particularly between Chapters 3 and 4.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ACRONYMS

AIC Akaike information criterion

COT cost of transport

CV coefficient of variation

EDA electrodermal activity

EMG electromyography

HR heart rate

IMU inertial measurement unit

LMEM linear mixed effects model

MAE mean absolute error

MDC minimum detectable change

RMSE root mean squared error

SD standard deviation

SEM standard error of the mean

ABSTRACT

Lower-limb wearable robots, such as prostheses and exoskeletons, have the potential to fundamentally transform the mobility of millions of able-bodied and disabled individuals during work, recreation, and/or rehabilitation. Researchers have developed high-performance robotic devices that are lightweight and autonomous, yet we have not seen their widespread adoption as clinical or commercial solutions. One major factor delaying the dissemination of this technology is that our control of these devices remains limited, especially outside controlled laboratory settings. While many adaptive control strategies have been proposed and tested for wearable robotic systems, it is still an open question as to how to provide users with *optimal* assistance. This is a challenge because a clear definition of 'optimal assistance' has not yet been established, due to the various goals of assistive devices and the complex interactions between robotic assistance and human physiology. Furthermore, it has been difficult to move these systems outside the laboratory environment due to constraints imposed by tethered hardware and sensing equipment, as well as the difficulty associated with quantifying non-steady-state activities.

To address this challenge, the goal of the work presented in this dissertation is to further our understanding of how to provide users with optimal assistance from robotic exoskeletons and prostheses outside the laboratory environment. The first theme of this dissertation is the translation of experimental methods (*e.g.,* human-in-the-loop optimization) and measurement tools outside the laboratory environment. The second theme is the evaluation of context-specific objectives for the optimization of wearable robotic devices (*e.g.,* metabolic cost, user preference) in clinical and research settings. This dissertation comprises four primary projects.

First, I evaluated the impact that changing the power setting of the BiOM commercial powered ankle prosthesis has on the metabolic cost of transport of individuals with transtibial am-

putation [74]. This work informs clinical tuning and prescription practices by revealing that, to minimize their metabolic energy consumption, individuals require a higher power setting than the setting chosen by the prosthetist to approximate biological ankle kinetics. Second, I investigated an alternative method for estimating instantaneous metabolic cost using portable wearable sensors. The goal of this project was to accurately estimate instantaneous metabolic cost without relying on indirect calorimetry, and thus enable the use of human-in-the-loop optimization algorithms outside the laboratory environment. I utilized linear regression algorithms to predict instantaneous metabolic cost from physiological signals, and systematically compared a large set of signals to determine which sensor signals contain the highest predictive ability, robust to unknown subjects or tasks [75–77]. In the third project, I built upon this work and demonstrated that including the derivatives of physiological signals in the linear regression algorithm could improve both the speed and accuracy of instantaneous metabolic cost prediction from portable sensors [81]. Finally, I investigated user preference as an objective for the control of lower-limb robotic exoskeletons [80]. In this work, I demonstrated that exoskeleton users can quickly and precisely identify unique preferences in the characteristics of their ankle exoskeleton assistance in two dimensions simultaneously, and that preference changes with speed, exposure, and prior experience. These results provide insight into how users interact with exoskeletons, and establish important benchmarks for researchers, designers, and future consumers. Together, these four projects evaluate the notion of optimality in the control of wearable robotic systems and lay the foundation for translating optimization protocols outside the laboratory environment.

CHAPTER 1

Introduction

1.1 Research Motivation

Bionic exoskeletons and prostheses have long-been anticipated as regular installments in our daily lives [46,62]. For able-bodied individuals, lower-limb robotic exoskeletons have the potential to increase strength or improve walking or running endurance by augmenting the wearer's existing neuromotor capability [151]. This type of exoskeleton may assist individuals while lifting heavy objects or carrying heavy loads over long distances in settings such as warehouses, construction sites, or military excursions [185]. For individuals with disabilities, lower-limb robotic prostheses and exoskeletons have the potential to make a large impact as medical devices. In the United States in 2019, an estimated 18.6% of adults over the age of 18 reported difficulty walking or climbing steps [122] due to conditions such as stroke, spinal cord injury, osteoarthritis, multiple sclerosis, or limb loss, among others [108]. Many of these individuals (an estimated 6.8 million as of 2000) reported using some sort of assistive device to help them with mobility [85]. However, the vast majority of assistive devices prescribed today are mechanically passive and can not add net positive work to the wearer during ambulation, which may still result in impaired mobility. For example, compared to non-amputees, individuals with transtibial amputation wearing passive-elastic prosthetic feet demonstrate higher kinematic asymmetry [150, 160, 183], expend 10–30% more energy while walking [68, 179], and develop increased incidences of chronic pain and osteoarthritis [97, 125]. Similarly, while mechanically-passive ankle-foot orthoses (AFOs) have been

shown to improve the energetic and spatiotemporal parameters of gait in individuals with hemiplegia [49] and muskuloskeletal weakness [130], gait performance may still be limited due to the lack of positive power generation from the device [11, 134, 147]. Alternatively, robotic lower-limb assistive devices use actuators to replace the work from missing or weakened musculature. As such, many individuals with mobility challenges could greatly benefit from the use powered lower-limb assistive robots to supplement or restore lost limb functionality.

Lower-limb wearable robots, such as powered prostheses and exoskeletons, have the potential to fundamentally transform the mobility of millions of able-bodied and disabled individuals during work, recreation, and/or rehabilitation. While significant strides towards this goal have been made over the last several decades, we have not yet seen the widespread adoption of these systems as clinical or commercial solutions. The work presented in this dissertation advances the field by proposing new tools for the design and evaluation of controllers for wearable robotic systems that may be used outside the laboratory environment.

1.2 Background

1.2.1 Lower-Limb Wearable Robotic Hardware

Today, there is a small, yet growing, number of commercially-available robotic lower-limb prostheses and exoskeletons. As of this writing, there are currently two commercially-available powered lower-limb prostheses—the Ossür Power Knee, and the Ottobock Empower Ankle (formerly the BiOM[1]). Some commercially-available powered lower-limb exoskeletons have been designed to augment the performance of able bodied users during activities such as walking (Honda Walking Assist Device), skiing (Roam Robotics Elevate), or repetitive industrial tasks (Sarcos Guardian). However, the largest number of commercial robotic lower-limb exoskeletons are currently employed as rehabilitation devices or assistive technology solutions for individuals who

[1]The BiOM ankle is currently commercially available as the Ottobock Empower [128], but will henceforth be referred to as the BiOM in this dissertation.

have sustained a spinal cord injury (*e.g.,* Parker Hannifin Indego, ReWalk, EksoBionics EksoNR, SuitX Phoenix) or stroke (ReWalk ReStore). Despite the relatively sparse commercial market for powered assistive devices, numerous research prototypes are under development. For prostheses, several robotic knees, ankles, and combined knee-ankle systems have been developed [8, 138, 144, 165, 169], with the goal of creating devices that are lighter and more compact, with longer battery lives. In exoskeletons, researchers have developed innovative systems such as a multi-joint, cable-driven exosuit made from flexible textiles [6], or an autonomous, lightweight ankle exoskeleton [119–121], which have the potential to be used outside the laboratory environment. New advents in mechatronic design have resulted in lightweight, portable, and high performance exoskeleton and prosthetic hardware, yet these devices have not yet achieved the anticipated adoption as widespread commercial solutions.

1.2.2 Lower-Limb Wearable Robotic Control

One major factor that has delayed the dissemination of robotic prosthetic and exoskeleton technology is that our control of these devices remains limited, particularly outside controlled laboratory settings. Control of robotic lower-limb prostheses and exoskeletons is of paramount importance to ensure that safe and effective robotic assistance can be delivered to the user in all environments. Across all lower-limb robotic devices, controllers often leverage the periodicity of ambulation and generate cyclical locomotion patterns for a variety of ambulation modes (*e.g.,* walking, stair ascent, ramp descent). High-level control laws often dictate the desired ambulation mode, while mid-level control laws are defined to generate the appropriate assistance trajectories that produce coordinated locomotion as a function of time or gait phase. In robotic lower-limb prostheses, several mid-level control paradigms have been implemented, including impedance control [79, 101, 161, 164, 165], myoelectric control [59, 71, 72], position control [138, 176], and neuromuscular model-based control [38], among others. These trajectories have be generated using finite state machines [38, 79, 101, 161, 164, 165], a continuous gait phase variable [138, 176], or the user's biological muscle activity [59, 71, 72]. For lower-limb exoskeletons (in particular, those

designed for able-bodied augmentation), many controllers are feedforward, with parameterized assistance profiles generated as a function of time or a presumed location in the gait cycle. Examples of various control trajectories include actuation force in Bowden cables [102, 132], or artificial pneumatic muscles [20, 47, 94, 95, 111, 152], joint torque [82, 189], or joint power [119–121]. These assistance trajectories have been generated in a variety of ways, using time between estimated gait events (*e.g.,* heel contact, toe-off) [20, 82, 95, 111, 119–121, 189], or biological muscle activity [20, 47, 94, 152] as triggers for assistance generation.

Within a particular control paradigm, the assistance that a user receives from the device is often defined by a set of *parameters* that shape the assistance trajectory. How an assistance profile is parameterized and how the parameters are individually tuned dictate the device's behavior. For example, in an impedance control scheme for a knee-ankle prosthesis, joint torque (τ) is generated according to the mechanics of a spring-damper system ($\tau = k(\theta - \theta_e) + b\dot{\theta}$), where θ and $\dot{\theta}$ are the joint's position and velocity, respectively. The parameters of this system are joint stiffness, k, damping coefficient, b, and equilibrium angle, θ_e. Within a state machine architecture, the selection of these three parameters can produce distinct ambulation patterns, such as level ground walking, sit-to-stand, incline/decline walking, and stair ascent/descent—but they must be carefully configured in order to achieve the desired behavior [79, 101, 161, 164, 165]. Parameters are usually both device- and controller-specific; additional examples of parameters in assistive device controllers include assistance timing [111] or assistance magnitude [131].

Numerous studies have demonstrated that a variety of control strategies and parameterizations are capable of producing coordinated locomotion patterns. Yet, it is still an open question how to design a controller that is 'best' for an individual user. Recent studies have demonstrated that the tuning of control parameters not only impacts the device performance, but can also dramatically affect the user's physiological response to the device. In a study performed by Malcolm *et al.* using bilateral pneumatic ankle exoskeletons, a simple 'bang-bang' control paradigm was implemented in which the onset of the artificial muscle's contraction occurred at a predefined percentage of the gait cycle [111]. This study demonstrated that incrementally changing one control parameter (the

onset of actuation) from 13% of the gait cycle to 54% produced changes in the user's metabolic cost on the order of 0.5 $\mathrm{W/kg}$, or 35 W for an average adult. In addition, recent research has indicated that individual subjects do not respond in the same way to the same parameter settings. Quesada *et al.* varied the net ankle work on a powered ankle prosthesis emulator, and reported unique relationships between net prosthesis work and metabolic cost for each individual subject [131]. Thus, in the context of recent results, there is a need to forgo static, 'one-size-fits-all' control laws and design *optimal* controllers for wearable robotic devices in order to maximize the benefits to individual users.

1.2.3 Identifying a Need for Research

While many control strategies have been proposed and tested for wearable robotic systems, a *fundamental gap in knowledge* still exists regarding how to provide users with optimal assistance, particularly outside the laboratory environment. This is a difficult problem because a clear definition of 'optimal assistance' has not yet been established, due to the various goals of assistive devices and the complex interactions between robotic assistance and human physiology. Furthermore, the vast majority of prosthetic and exoskeleton research has been conducted using steady-state activities performed under tightly-regulated experimental conditions. Thus far, it has been challenging to move these systems outside the laboratory environment due to constraints imposed by tethered hardware and sensing equipment, as well as the difficulty associated with accurately quantifying non-steady-state activities.

To close this gap, in this dissertation I evaluate the notion of optimality in the control of wearable robotic systems by broadly asking, "what does it mean to be optimal?" and more specifically, "how do we provide users with optimal assistance from robotic exoskeletons and prostheses outside the laboratory environment?" To ground my research approach, I propose a general framework for addressing these questions, which is presented in Fig. 1.1. This framework is divided into four stages: 1) define the *objective* of the robotic assistance (*i.e.,* what is the goal of the device?), 2) robustly *measure* or quantify this objective (particularly outside the laboratory), 3) *design* and

customize or *tune* a controller that achieves this objective, and 4) *evaluate* the human-robot system to determine if the objective has been achieved, as well as to broadly investigate how the assistance affects the user's physiology and perception. Thus, in this context, *optimal* assistance refers to the controller which maximizes the objective of the device or assistance. Importantly, this framework feeds back upon itself, with results from the evaluation offering critical insight into each preceding stage. The central tenant of this thesis is that by specifically targeting each phase of the optimization and evaluation framework, as well as considering the framework as a whole, we can design optimal and translational controllers for wearable robots. Evaluating optimality in the control of wearable robotic systems is a critical step towards the translation of such systems into viable clinical and commercial devices.

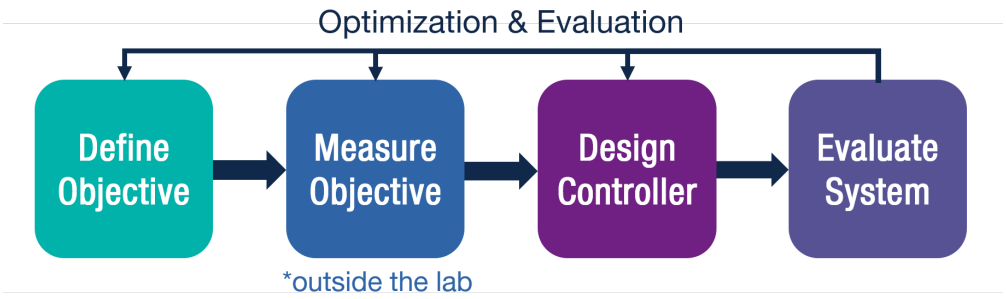

Figure 1.1: A proposed framework for the optimization and evaluation of wearable robotic devices.

1.3 State of the Art

1.3.1 Customized Control for Lower-Limb Wearable Robotic Devices

To date, many different methods have been employed to tune device controllers for individual users or multiple ambulation modes. This section summarizes the state-of-the-art approaches for the customization and optimization of controllers for wearable robotic systems.

Biologically-Inspired Tuning

One intuitive method of designing a controller for a wearable robot is to reproduce a pre-defined desired kinematic (position) or kinetic (force/torque/power) trajectory as a function of time or gait phase. For example, in exoskeletons, Mooney *et al.*, employed an ankle exoskeleton controller that mimicked biological ankle power generation over the gait cycle [119–121]. Similarly, Panizzolo *et al.* and Kim *et al.* developed 'biologically-inspired' exosuit controllers, which controlled the force in the Bowden cables to reproduce biological hip and ankle torques throughout the gait cycle [88, 129]. In a powered knee-ankle prosthesis, Rezazadeh *et al.* commanded biological kinematic trajectories using a phase variable and virtual constraints [138]. In these implementations, the *optimal* controller was that which best reproduced the desired trajectories. While commanding scaled versions of biological joints' kinematics and kinetics is an intuitive starting place for lower-limb robotic controllers, it is not clear if they produce the 'best' assistance for an individual user, given the static control trajectories and lack of a defined physiological objective.

Expert Tuning

In a clinical environment, the optimization and evaluation of robotic prostheses often relies on *expert tuning*, in which a clinician (*e.g.,* a prosthetist) or trained device expert is employed to hand-tune the device parameters for an individual patient. This method has been demonstrated in studies using the BiOM powered prosthesis [45, 52, 58, 63, 145], and a powered combined knee-ankle prosthesis [79, 161]. In this expert tuning paradigm, the objective of device tuning is often to enable safe and comfortable ambulation, to produce 'satisfactory' kinematics and kinetics for a variety of ambulation modes, or to replicate able-bodied gait parameters (*e.g.,* net ankle work as a function of walking speed). These objectives are measured in the lab using a combination of patient feedback, expert observation, and analysis of limited measurements from the prosthesis, such as joint angles or torques. Although some methods have been proposed to streamline this empirical process, such as the use of an intuitive clinical tuning interface [133], or parameter reduction [161], expert tuning is time-intensive and imprecise. This method requires the expensive services of

a trained clinician or device expert, lacks rigid tuning methodology (*i.e.,* uses trial-and-error), does not have clear objective metrics for tuning success, and quickly becomes infeasible with an increasing number of parameters. Despite these drawbacks, expert tuning remains standard clinical practice for prescribing powered prostheses, and expediting and improving the tuning process by understanding how tuning choices impact clinical outcomes is a worthwhile avenue of study.

Parameter Studies

In a non-clinical experimental setting, researchers have leveraged new tools to investigate alternative objectives and controller designs for lower-limb robotic prostheses. Caputo and Collins developed a 'universal ankle-foot prosthesis emulator,' which enables researchers to rapidly explore a wide range of prosthesis controllers and evaluate their effects on human subjects [24]. Using this tool, several studies have investigated the effect of various prosthesis control architectures on the metabolic cost of able-bodied individuals wearing an adapter [23, 112], and individuals with transtibial amputation [131]. In these studies, the defined physiological objective of the prosthesis controller was to reduce the metabolic cost of walking; reducing energy cost has clinical significance as individuals with transtibial amputation spend 10-30% more energy walking compared to able-bodied individuals [68, 179]. To achieve this objective, these studies performed a brute force parameter sweep, measuring metabolic cost via indirect calorimetry across a range of parameter settings (*e.g.,* push-off work [23, 131], and push-off timing [113]). In this tuning paradigm, the optimal parameter settings were identified post-hoc, as those which minimized the user's energy expenditure.

Similar brute force parameter sweep methods have been applied to robotic lower-limb exoskeletons. Studies have identified parameter settings that minimize the user's metabolic cost in pneumatic ankle exoskeletons [50, 51, 111], a soft exosuit [112, 132], and an ankle exoskeleton emulator [82]. Thus far, for robotic exoskeletons, metabolic cost has served as the near-universal physiological objective to minimize—the less metabolic energy a person consumes, the longer they can continue performing said task. A confounding factor with this method is that the *measurement*

of metabolic cost via indirect calorimetry requires 3-6 minutes of steady-state data acquisition to accurately quantify the corresponding metabolic cost[2]. While measurement time can be slightly reduced using a first-order dynamic model of respiratory kinetics [154] and more efficient parameter-search algorithms [44,93], the brute force parameter sweep remains time-consuming and infeasible for large numbers of parameters. Despite its drawbacks as a controller tuning method, the brute force parameter sweep does allow for controlled scientific investigation of how one or two aspects of controller design affect the human-robot mechanical system (*e.g.,* center-of-mass mechanics, joint mechanics, muscle activity, and energy consumption). This type of information is crucial to understanding the complex interactions between control choices and human's biomechanics and physiology.

Automated Tuning

Instead of brute force search methods, several groups have investigated *automated tuning* methods for the customization of lower-limb wearable robots, in which an algorithm chooses the appropriate controller configuration in order to achieve a defined objective. For tuning control parameters of powered prostheses, methods such as adaptive dynamic programming [180], rule-based fuzzy logic [70], and extremum-seeking control [98] have been explored. For these systems, defining an explicit cost function (*i.e.,* objective) for automatic parameter tuning can be challenging, and may still rely on experts to pre-define the tuning rules, as in the fuzzy logic paradigm presented by Huang *et al.* [70]. In [98], Kumar *et al.* presented a convex optimization function for powered prosthetic legs, which incorporated both kinematic tracking error and user discomfort. Yet, the authors acknowledged that this cost function may not capture optimal gait biomechanics or favorable energetic conditions.

As an alternative, 'body-in-the-loop' or 'human-in-the-loop' optimization algorithms utilize a physiologically-relevant cost function, such as metabolic energy expenditure [44]. In body-in-the-loop optimization (illustrated in Figure 1.2), individuals walk with a robotic device providing

[2]For state of the art metabolic cost measurement techniques, please refer to Section 1.3.2.

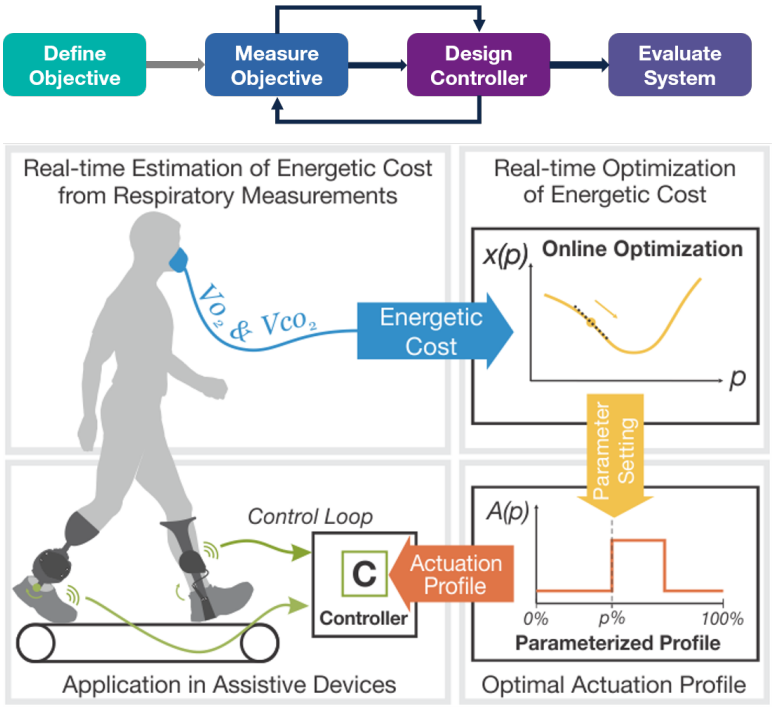

Figure 1.2: Illustration of body-in-the-loop optimization. This figure has been adapted with permission from Koller *et al.* [92].

assistance according to a parameterized actuation profile, $A(p)$, shaped by one or more parameters p. Once defined, the actuation profile is enforced by the device's low-level control loop, which often incorporates device feedback in a closed-loop controller. While the user walks with the current actuation profile, their metabolic cost is measured using indirect calorimetry (*i.e.,* respiratory measurements). Given the metabolic cost associated with the current parameter setting, $x(p)$, an online optimization algorithm (*e.g.,* gradient descent [92], CMA-ES [189], or Bayesian [35]) chooses the next parameter setting to test. This process continues, eventually converging on the optimal actuation profile that minimizes the metabolic cost of the wearer. In the proposed general framework (Fig. 1.1), body-in-the-loop optimization represents an internal feedback loop (Fig. 1.2, top); this algorithm continuously measures the objective and updates the controller parameters until the physiological objective has been met.

This paradigm has been successfully implemented to identify able-bodied individuals' optimal step frequency [44, 90], as well as optimize push-off timing of pneumatic ankle exoskeletons [92], four-torque-profile shaping parameters on an ankle exoskeleton emulator [189], and three force-

profile-shaping parameters on a hip exosuit [35]. Utilizing human-in-the-loop optimization to design a device controller has resulted in metabolic cost reductions of 24 % (compared to wearing the exoskeletons in a 'zero-torque' mode) [189] and 17 % (compared to not wearing the device) [35]. It is important to note, however, that since both these studies utilized an emulator system, in which the actuation hardware was located off-board, the unilateral ankle exoskeleton [189] and bilateral hip exosuit [35] hardware were very lightweight; extrapolation of the energetic savings that this method may incur with a fully autonomous device is a challenge. One of the most promising features of human-in-the-loop optimization techniques is the ability for the controller to automatically adapt to different users, different ambulation tasks, or environmental conditions in real time. However, current implementations of these algorithms are unable to be used outside a laboratory environment, due to tethered experimental hardware and the challenges associated with measuring metabolic energy cost. As with the brute force parameter sweep method, measuring metabolic cost from respiratory measurements requires that the user wears a bulky rubber mask that covers their nose and mouth, which renders it impractical for daily use applications. Furthermore, data collected from indirect calorimetry are sparesly sampled, noisy, and dynamically delayed, and thus it requires 3-6 minutes of constant-intensity activity in order to obtain one estimate of steady-state energetic cost. To mitigate some of these challenges, all body-in-the-loop optimization studies to date have leveraged the work of Selinger and Donelan [154], modeling and inverting the kinetics of the respiratory response as a first order linear system[3]. While this model enables researchers to predict steady-state energy cost from non-steady-state measurements [35, 90, 189], or to directly estimate the cost landscape (*i.e.,* energy cost as a function of parameter setting) [44, 92], current protocols still require 2-4 minutes to estimate steady-state energy cost at one parameter setting—this results in optimization routines that take 40-60 minutes for one individual. For human-in-the-loop optimization to reach its full potential as a real-time automatic tuning algorithm outside the laboratory, we must be able to *measure* metabolic energy expenditure without using indirect calorimetry or define a cost function that can be more easily measured.

[3]For state of the art metabolic cost measurement techniques, please refer to Section 1.3.2.

Self-Tuning

As discussed previously, pre-defining an objective function that captures the user's optimal biomechanical and physiological state is a major challenge, since we don't know exactly which objectives a human optimizes for and in which situations. As a result, researchers have begun to investigate different objectives for the customization of wearable robots, including more subjective metrics such as *preference* or *comfort*. Yet, with these subjective metrics, it is critical to be able to robustly measure the objective you wish to evaluate. To measure preference, some studies have utilized a two-alternative forced choice paradigm, in which subjects were presented with pairwise comparisons (*i.e.,* A-B testing) [105, 168, 172, 173], or asked to compare a condition to their own internalized preference [157, 158]. Data obtained from forced choice methods can be used to identify or learn a user's preference, but it may require extended experimental time to obtain a sufficient number of comparisons. To mitigate this challenge and accelerate the estimation of a user's preference, one study implemented a coactive learning paradigm in which users not only chose between paired conditions but also directly suggested improvements for the assistance each trial (*e.g.,* longer or shorter step length) [173]. As an alternative to forced choice methods, some studies have employed a method of adjustment, in which individual users (or experimenters directed by users) adjusted control parameters until they achieved their preferred assistance [22, 28, 155]. Such *self-tuning* methods are advantageous because they can quickly yield an individual's preference and are intuitive to the subject——it is easy to imagine someone using a smart phone or smart watch to adjust their settings during activities of daily living. In his doctoral thesis, Caputo describes a method for optimizing three parameters of a ankle prosthesis emulator (programmed to mimic the behavior of a passive-elastic foot) based on subjective user feedback [22]; in this study, the experimenter incrementally adjusted each parameter until the subjects became 'dissatisfied' with the device behavior. Shepherd *et al.* gave the tuning power directly to the individual subject, and developed a method to robustly measure subject preference in the laboratory environment [155]. Subjects wore a variable-stiffness quasi-passive ankle prosthesis [156], and were instructed to use a dial to modify their prosthetic foot stiffness until they arrived at the stiffness they preferred; this

was repeated multiple times from random initial conditions. In this study, participants repeatably and precisely determined their preferred ankle stiffness, with an inter-subject mean coefficient of variation of 14 %. Finally, while Young *et al.* did not directly incorporate user feedback into the parameter tuning of a bilateral powered hip exoskeleton, they did ask individuals to indicate their preferences between two subsequent trials with different assistance patterns [186]. Results from this study showed that subject preference did not correspond to conditions which minimized the user's energetic cost, highlighting the need to explore additional metrics for exoskeleton evaluation and tuning. Given the high variability between users and the difficulty in defining one specific objective function for all situations, incorporating subject preference into tuning algorithms may encode multiple subjective costs at once, thereby allowing us to design robust, customized assistive device controllers.

1.3.2 Metabolic Cost Measurement

Metabolic cost (herein used interchangeably with metabolic rate, energetic cost, energy cost, or energy expenditure) is an important physiological quantity used in the design and evaluation of control systems for wearable robots. Most often, measurements of whole-body energetic cost are obtained using indirect calorimetry. With indirect calorimetry, energy production is determined by measuring oxygen (O_2) consumption and carbon dioxide (CO_2) production, rather than directly measuring heat transfer (as in direct calorimetry) [162]. Modern indirect calorimetry systems (*e.g.,* Cosmed K5) are portable, and the individual wears the processing unit in a harness on their chest or back (Fig. 1.3). The person also wears a rubber mask over their nose and mouth, which contains an embedded flowmeter and oxygen/carbon dioxide sensors to measure the volumetric flow rate of each of the respiratory gases ($\dot{V}_{O_2}$, $\dot{V}_{CO_2}$) breath-by-breath [109]. Using the Brockway equation (a weighted sum of $\dot{V}_{O_2}$ and $\dot{V}_{CO_2}$), respiratory gas measurements can be converted into a measurement of whole body metabolic rate (unit: Watts) [18].

Yet, there are significant challenges associated with using indirect calorimetry to measure metabolic cost. From a practical perspective, although modern equipment is relatively lightweight

Figure 1.3: Cosmed K5 indirect calorimetry unit. This photo was obtained on from the Cosmed K5 brochure [31].

and portable, all indirectly calorimetry systems require that the user wear a rubber mask over their nose and mouth (Fig. 1.3), which is unsuitable for long-term data collection outside the laboratory environment. Furthermore, there are several challenges related to the data obtained from indirect calorimetry systems. First, measurements of metabolic cost are sparsely sampled, with one data point recorded every breath. With the typical breathing rate of 20 breaths per minute during light exercise, this corresponds to sampling rate of just 0.3 Hz [154]. Second, measurements of metabolic cost are extremely noisy, with a signal-to-noise ratio of approximately four [44, 99]. And finally, metabolic cost measurements obtained from the respiratory gases are dynamically delayed from the instantaneous metabolic demands of the body. To illustrate this relationship, consider an individual performing a step-change in activity level (*e.g.,* changing from walking at 1.0 m/s to 1.5 m/s). On a chemical level, the muscles *instantaneously* consume more energy to meet the demands of the new task. Yet, there are a number of physiological processes (*e.g.,* mitochondrial dynamics, blood circulation from the muscles to the lungs, lung tidal volume) that

delay measuring this change in energy expenditure at the mouth through the analysis of respiratory gases [99, 154]. For all these reasons, metabolic cost is traditionally measured during long bouts of constant-intensity activity.

One method to obtain an accurate measurement of the body's instantaneous metabolic cost for a given activity is to have an individual perform that activity for 2–3 minutes to reach a steady-state oxygen consumption level [116,136,154]. Once steady state has been reached, another 2–3 minutes of breath measurements are recorded and then subsequently averaged into one estimate of steady-state (*i.e.,* instantaneous) energetic cost for that activity (Fig. 1.4A). Averaging over the last several minutes of steady-state data minimizes the impact of the measurement noise and thus produces an accurate estimate of the body's instantaneous energetic cost. As we assume the instantaneous energy cost is constant for a given activity, this steady-state value can be extrapolated backwards in time for the entire activity. While this traditional method has been employed for decades, it requires 5–6 minutes to obtain *one* estimate of energetic cost for a particular activity, and can not capture transient changes.

As an alternative, Selinger and Donelan proposed a method to estimate instantaneous metabolic cost that leverages a dynamic model of respiratory kinetics [99, 154, 181]. In their work, they demonstrated that, in response to a step-change in activity intensity, the dynamics of the breath measurements can be characterized by a first-order linear differential equation:

$$\dot{y}(t) = \frac{1}{\tau}\Big(x(t) - y(t)\Big) \tag{1.1}$$

where $x(t)$ is the energetic cost that will be reached in *steady state* when the activity performed at time t is continued at the current intensity, and where $y(t)$ is the *transient* respiratory measurement at time t. The subject-specific time constant τ represents the speed with which the breath measurements can respond to a step change in activity intensity. In the study performed by Selinger and Donelan, they experimentally determined that τ varied between individual subjects, with a mean (SD) of 41.8 (12.1) seconds [154].

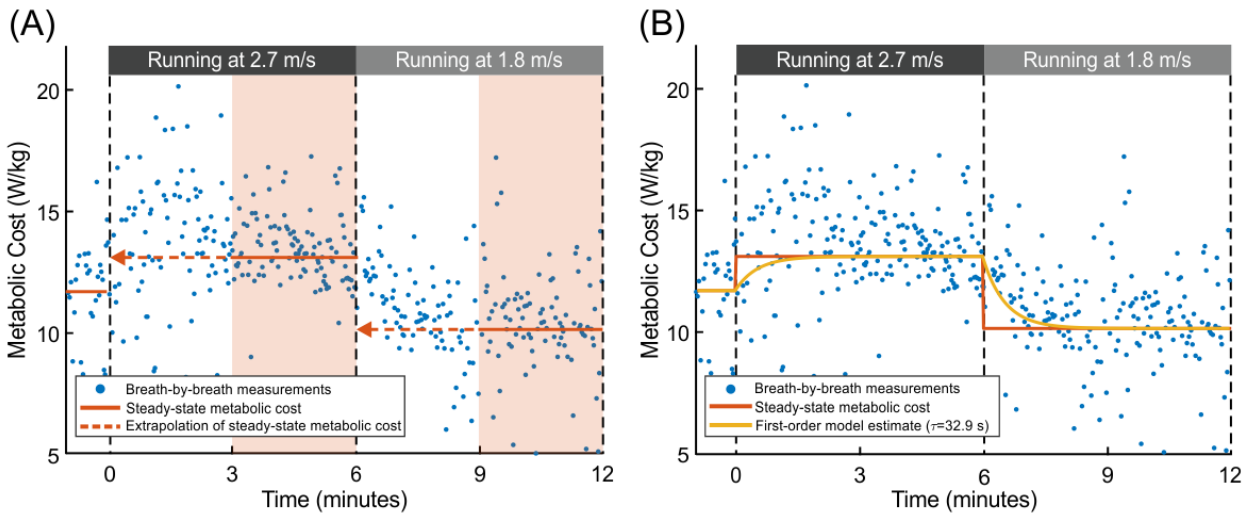

Figure 1.4: Methods for calculating steady-state metabolic cost from breath measurements. (A) For a given activity, the steady-state metabolic cost (solid red line) is the average of the last three minutes (shaded red area) of breath measurements (blue dots). Because instantaneous energetic cost is constant for a given activity, this value is extrapolated backwards in time for the entire duration of activity (red dashed line). This approach requires 5–6 minutes to obtain an estimate of instantaneous energetic cost. **(B)** A first order model (yellow line) can be fit to the breath measurements (blue dots). For this individual, the subject-specific time constant τ was determined to be 32.3 seconds. The asymptote (*i.e.,* steady-state value) of this model determines the instantaneous metabolic cost for that activity. This approach requires 2–4 minutes to obtain an estimate of instantaneous energetic cost. These metabolic cost data were obtained from the experimental data collection presented in Chapters 3 and 4.

Given this model structure, to obtain an estimate of the steady-state value $\hat{x}(t)$ from transient breath measurements $y(t)$, one must simply rearrange the terms:

$$\hat{x}(t) = y(t) + \tau \dot{y}(t). \tag{1.2}$$

Thus, the estimated steady-state value, $\hat{x}(t)$ corresponds to the instantaneous energetic cost for that given activity (Fig. 1.4B). One approach for calculating $\hat{x}(t)$ is to fit a first-order model to transient breath measurements and identify the asymptote of the model [35, 90, 189]. Alternatively, Felt *et al.* developed a mathematical approach to directly determine the cost landscape (*i.e.,* metabolic cost as a function of the various conditions tested) [44], and Koller *et al.* expanded this approach by introducing confidence analyses of the cost landscapes [93]. Regardless of the chosen analysis method, leveraging the first-order dynamic model allows researchers to use non-steady-

state breaths to predict instantaneous energetic cost instead of waiting for the respiratory kinetics to reach steady state. Therefore, current implementations of instantaneous energetic cost estimation using the dynamic model require 2–4 minutes of data collection per activity, which is faster than the traditional approach of averaging steady-state measurements.

1.4 Contributions

The goal of the work presented in this dissertation is to further understand how to provide users with optimal assistance from robotic exoskeletons and prostheses outside the laboratory environment. To achieve this goal, I experimentally evaluate context-specific objective functions for the optimization of assistive robots in clinical, laboratory, and real-world environments. Additionally, I develop versatile tools to measure and/or estimate relevant physiological and biomechanical quantities and subjective criteria (*e.g.,* user preference). As such, my contributions are relevant to two main themes: first, the translation of experimental methods and measurement tools outside the laboratory environment and second, the evaluation of context-specific objectives for the optimization of wearable robotic devices. The content of this dissertation is described over four chapters.

In Chapter 2, I evaluate the impact that changing the power setting of the BiOM commercial powered ankle prosthesis has on the metabolic cost of transport of individuals with transtibial amputation. Additionally, I compare the power setting that minimizes an individual's cost of transport with the setting that was chosen by a prosthetist during a standard clinical fitting session (*i.e.,* the setting that approximates biological ankle kinetics). This work experimentally demonstrates that on average, subjects exhibit lower cost of transport with increasing power setting. Yet, the relationship between BiOM power setting and cost of transport varies significantly between individuals and thus reinforces the need for user-specific tuning. This observed variability led us to further investigate the relationship between the prosthetic ankle's mechanics and individuals' COT. Furthermore, this work informs clinical tuning and prescription practices by revealing that, to minimize their metabolic energy consumption, individuals require a higher power setting than

the setting chosen by the prosthetist to approximate biological ankle kinetics. This chapter has been published in full in *Scientific Reports* [74].

In Chapter 3, I investigate an alternative method for estimating instantaneous metabolic cost using portable wearable sensors. The goal of this work is to accurately estimate instantaneous metabolic cost without relying on indirect calorimetry, and thus enable the use of body-in-the-loop optimization algorithms outside the laboratory environment. I present a black-box approach (*i.e.,* linear regression algorithms) for estimating instantaneous metabolic cost quickly and accurately using physiological signals collected from portable, wearable sensors. In this chapter, my contributions include systematically comparing a large set of physiological signals and determining which sensor signals contain the most salient information for predicting instantaneous metabolic energy cost, robust to unknown subjects or tasks. Additionally, I discuss the 'best' physiological signals to use in a regression algorithm to maximize predictive capability when only a small number of signals (fewer than 8) are included. Alongside the analyses presented in this chapter, I describe the large experimental dataset that I collected for this study, which is published open-source to enable other researchers to build their own predictive models [78]. Preliminary findings from this work have been published as peer-reviewed conference proceedings and presented at the *2017 IEEE International Conference on Rehabilitation Robotics* [76] and the *2017 IEEE Global Conference on Signal and Information Processing* [75]. Chapter 3 has been published in full in the *Journal of Applied Physiology* [77].

In Chapter 4, I build upon the work presented in Chapter 3, and demonstrate that including the *derivatives* of physiological signals in the linear regression algorithm can improve the speed and accuracy of instantaneous metabolic cost predictions. In particular, I show that adding derivatives improves predictions during the transient phase of an activity and specifically when signals with delayed dynamics are included in the regression (*e.g.,* heart rate). To support these analyses, a major contribution of this work is the reinterpretation of a first-order dynamic model as a linear regression model with a signal and its derivative and the formal comparison between the coefficients of these two models. This work is significant to the optimization of wearable robots because

the faster we can obtain accurate estimates of instantaneous metabolic cost, the more quickly we can iterate through control parameters in a body-in-the-loop optimization routine. This chapter has been published in full in the *IEEE Robotics and Automation Magazine* [81].

In Chapter 5, I investigate user preference as an objective for the control of lower-limb robotic exoskeletons. The goal of this work is to understand the characteristics of preference in the context of exoskeleton control and discover how individuals perceive, interact with, and learn from the assistance provided. To this end, I developed a novel two-dimensional self-tuning touch screen interface to enable ankle exoskeleton users to simultaneously adjust the timing and magnitude of peak torque delivery in real time. In this chapter, I systematically analyze the control settings users prefer, how precisely individuals are able to identify their preferences, and how they interact with the touch screen to find their preferred settings. I discover how these outcomes vary as a function of walking speed, as a function of time spent wearing and interacting with the robotic device, and between users with different technical backgrounds. To my knowledge, this work is the first to demonstrate that users have unique preferences in their robotic ankle exoskeleton assistance, and that they can precisely and quickly identify these preferences using a two-dimensional self-tuning interface. This work motivates new strategies for the control of lower-limb exoskeletons in which users customize assistance according to their unique preferences, leveraging their internalized representation of optimality. Preliminary findings from this work have been published as peer-reviewed conference proceedings and presented at the *2020 IEEE Conference on Biomedical Robotics & Biomechatronics* [80]. As of this writing, Chapter 5 has been submitted in full for publication and is currently under review.

As a whole, this dissertation expands the notion of optimality in the control of wearable robotic systems and lays the foundation for translating optimization protocols outside the laboratory environment. This dissertation contributes experimental results, robust measurement tools, theoretical framework, and scientific evaluation that each advance the field of wearable robotics closer towards the goal of achieving clinical and commercial ubiquity in the future.

CHAPTER 2

Choosing Appropriate Prosthetic Ankle Work to Reduce the Metabolic Cost of Individuals with Transtibial Amputation

2.1 Background

Individuals with transtibial amputation spend 10–30 % more metabolic energy when walking compared to able-bodied individuals [68, 69, 171]. This observed increase in energy expenditure may be due to the fact that most ankle prostheses are passive-elastic devices, which store and release energy when in contact with the ground but cannot perform positive net work. In fact, passive-elastic prostheses only produce about an eighth the power of the intact gastrocnemius and soleus muscles [190]. This deficit has a significant impact on walking as the majority of the total mechanical power generated during the gait cycle comes from the ankle-foot complex [123, 183, 188]. As a result of decreased ankle power generation, people with amputation may put forth additional muscular effort from their residual limb or compensate with their intact limb to walk with passive prostheses [148]. Additionally, physical fitness may play an important role in determining the metabolic demands of individuals with amputation [146].

To overcome these limitations of passive devices, various types of powered ankle prostheses have been developed [27, 63, 65]. These powered devices use actuators to deliver positive work to

This chapter has been previously published in *Nature Scientific Reports* [74]

the user during the push-off phase, and can potentially alleviate the increased energetic demand that people with amputation experience during walking. Of these devices, only the BiOM[1] powered ankle prosthesis (BionX, Bedford, MA) is currently commercially available. Investigating the efficacy of the BiOM in clinical trials has resulted in mixed outcomes. For walking speeds faster than 0.75 $\frac{m}{s}$, Herr *et al.* [63] found a significant reduction in metabolic cost when individuals with transtibial amputation walked with a powered ankle prosthesis compared to passive-elastic feet. Similarly, Esposito *et al.* [145] found a 16 % decrease in metabolic rate in highly active individuals using a powered ankle compared to their prescribed passive feet. In contrast, Gardinier *et al.* [52] found no significant differences in metabolic cost between individuals using a powered ankle and a non-powered prosthesis. The mixed results of studies evaluating the efficacy of powered prostheses are likely related to a variety of factors, including different subject populations. In the study by Gardinier *et al.* [52] individuals with the maximum functional classification level (K4) received an metabolic benefit from using the BiOM, while those with a lower functional classification (K3) did not. Correspondingly, the studies which demonstrated the largest reductions in metabolic cost were those that tested high-functioning active-duty military members [45, 145].

Another important factor when evaluating the BiOM's impact on reducing metabolic energy expenditure is how the device is tuned; Esposito *et al.* [145] suggested that near-optimal tuning of the BiOM is required to positively impact an individual's metabolic cost. Of particular importance when tuning the BiOM are the *power settings*, which determine the amount of ankle work delivered by the device. The BiOM provides prothetists with a visual display to tune the power settings on the device. The prosthetist modifies the power settings until the ankle work delivered by the prosthesis approximates normative data of healthy ankle work at the subject's preferred walking speed. Yet, it is still uncertain whether that amount of prosthetic ankle work is optimal to lower the user's metabolic cost. In fact, a number of arguments could be made against this assumption: 1) the work produced by a prosthesis might not be delivered in full to the user's center of mass, 2) a uniarticular prosthesis can never fully replicate the biarticular muscles in a human ankle, and 3) excess ankle

[1]The BiOM ankle is currently commercially available as the Ottobock Empower, but will henceforth be referred to as the BiOM in this dissertation.

work could be beneficial to compensate for other losses.

First of all, there is an open question regarding how much of the push-off work produced by a powered prosthesis is effectively delivered to the user. Prosthetic power may be lost in transmission due to residual limb deformation or the relative movement between the residual limb and the socket (*i.e.,* pistoning) [187]. Compliant characteristics of foot cosmeses and shoes may also contribute to dissipating mechanical work of powered prostheses [110]. These additional losses might render the power delivered to the center of mass smaller than anticipated.

Furthermore, since all currently existing powered ankle prostheses are uniarticular, they can never fully replicate an intact human ankle. In particular, they cannot replace the function of the gastrocnemius, a biarticular muscle, which contributes to work at both the ankle and knee. This is important since both ankle plantarflexors (soleus and gastrocnemius) deliver energy to support the body and propel the center of mass forward [123]. In the absence of functional plantarflexor muscles, trunk support can be accomplished by energy return from a carbon fiber foot, but forward propulsion requirements may not be fully met [190]. Uniarticular actuation might further lead to insufficient energy transmission to initiate knee flexion during pre-swing. In order to propel the body and leg forward, other leg or hip muscles may compensate and thereby drive up metabolic cost, even though the supplied prosthetic plantarflexor power approximates biological norms. Thus, walking with prosthetic push-off work in excess of biological norms may provide not only body support but also propulsion, which more closely matches the function of intact plantarflexors.

Finally, there is some general evidence that walking with ankle work in excess of biological norms may reduce metabolic effort. Providing ankle power greater than the biological ankle has been shown to reduce metabolic cost in able-bodied subjects walking with exoskeletons [94, 119] and an ankle prosthesis adapter [23]. Simplified gait models suggest that ankle push-off work from the trailing limb can reduce the negative work performed through dissipation at heel contact by the leading limb [37], yet recent human experiments with an ankle prosthesis adapter have not observed this effect [23, 131]. A reduction in negative work and/or increase in push-off work may lower metabolic cost by reducing the required hip work of the stance leg during single support

stance phase [68, 104, 153]. Although the exact biological mechanisms are not yet known, supplying prosthetic ankle work in excess of biological norms may be a useful strategy to compensate for increased effort in other places.

While this prior work suggests that increasing the amount of ankle work beyond the biological norms may be beneficial, we also believe that there may exist an upper bound with diminishing or negative return. For example, an excessive ankle plantarflexion moment can induce knee hyperextension [142]. In the non-amputated limb, the gastrocnemius and Achilles tendon restrain excessive knee extension during stance. In the amputated limb, however, the functional absence of these restraints may result in knee hyperextension with excessive prosthetic ankle push-off, thereby requiring other knee flexors to compensate and potentially increasing the user's metabolic effort [45, 145]. Additionally, the uniarticular function of the prosthesis may cause excess plantarflexion power to be re-distributed to the leg via the trunk and lead to increased hip power generation and increased knee power absorption [45]. As a result, these compensatory muscular efforts may contribute to increased metabolic cost. Due to the complexity of the human musculoskeletal system, there will likely be other examples of compensations that increase metabolic cost in response to excessive ankle power.

The goal of this study was to determine the effect of the BiOM's ankle power setting on the metabolic cost of walking. We performed an experiment in which ten individuals with transtibial amputation walked on a treadmill wearing the BiOM powered ankle prosthesis, tuned to six different power settings. The power settings ranged from no power (0 %) to the device's maximum power setting (100 %) in increments of 25 %, and included the power setting chosen by the prosthetist during a clinical fitting session. We calculated subjects' cost of transport (COT) during the steady-state portion of each condition. The BiOM records its net ankle work during each step, and we averaged this value over the last 30 steps of each condition to obtain the average net work. Our first hypothesis was that a subject's energetically optimal power setting would be higher than the prosthetist-chosen power setting (*i.e.,* the power setting chosen to approximate biological ankle kinetics). Our second hypothesis was that a subject's energetically optimal power setting would

not fall at 100 % power (*i.e.,* the maximum). We expected that the highest power setting would increase metabolic cost as the person is forced to absorb excess work with their musculoskeletal system.

2.2 Methods

2.2.1 Participants

Ten adult males with unilateral transtibial amputation participated in this study (Table 2.1). Subjects self reported their K-level and it was confirmed by the clinician. Potential subjects were excluded if they had a history of serious cardiovascular, neurological, respiratory, or visual problems, or were taking medications that might interfere with walking ability. Subjects were required to have the ability to walk for 30 minutes without a walking aid. Accordingly, all participants had a Medicare functional classification level of K3 or K4 [26], which corresponds to a moderate to high level of ambulatory function. Study procedures were reviewed and approved by the University of Michigan's Medical School Institutional Review Board. The study was carried out in accordance with the approved protocol. All participants provided written informed consent prior to participation.

Participant	Age (years)	Sex	Weight† (kg)	Height† (m)	Leg Length† (m)	K-Level
1	59	M	83.7	1.70	0.88	K3
2	24	M	88.2	1.81	0.97	K3
3	26	M	79.9	1.88	1.04	K4
4	60	M	123.8	1.84	0.92	K4
5	55	M	99.8	1.80	0.94	K3
6	32	M	87.3	1.85	0.98	K4
7*	27	M	120.2	1.78	0.96	K4
8	54	M	107.5	1.82	1.00	K4
9	53	M	73.9	1.70	0.89	K4
10	27	M	64.9	1.89	1.00	K4
Mean (SD)	41.7 (15.5)	-	92.9 (19.5)	1.80 (0.07)	0.96 (0.05)	-

†Measured while wearing the BiOM prosthesis.

*Subject 7 was excluded from analysis due to prosthetic battery failures during collection.

Table 2.1: Participant demographics.

2.2.2 Prosthetic Fitting and Tuning

At the start of each session, all participants were fitted with the BiOM powered ankle prosthesis (BiOM T2 Ankle, BionX Medical Technologies Inc., Cambridge, MA), except for Subject 9 who was already a regular user of the BiOM. The same certified prosthetist fit each participant according to manufacturer's recommendations. This process is described in detail in Gardinier *et al.* [52]. Briefly, the device was attached to the patient's existing socket and pylon using standard attachments and then aligned. While wearing the device, we weighed each subject, and measured their height (the distance from the floor to the crown of head), and the length of both legs (the distance from the greater trochanter to the floor).

Next, the device was powered on and participants walked back and forth down a 20-meter-long hallway while the prosthetist increased the amount of net ankle work delivered by the device using a Bluetooth-enabled tablet computer. With each step, the tablet displays a dot corresponding to the net ankle work done by the prosthesis at the participant's chosen walking speed (Figure 2.1). This dot is green if it falls within the 95% confidence interval of normative data for net ankle work at that walking speed; net ankle work outside of this range is displayed as orange or red. It is important to note that while the prosthetist is able to see the net work performed by the device with each step, he or she does not have direct control over this parameter. Rather, the prosthetist tuning the device can change the power *setting* of the device, which ranges from 0% power to 100% power in increments of 1%. In our study, the prothetist adjusted the power setting until the net ankle work fell within the desired range and the walking looked smooth and felt comfortable to the participant. After the power setting, adjustments were made to other settings, such as the timing of the power onset. Participants were given 30 minutes to acclimate to the device once initial fitting was complete. During this time, the prosthetist fine-tuned the actuation settings to best fit the comfort of the user as necessary. The final power setting is subsequently referred to as prosthetist-chosen power setting. Participants' other settings (*e.g.,* timing) were held constant throughout the experiment. A complete list of all the BiOM settings for each participant is provided in Table 2.2.

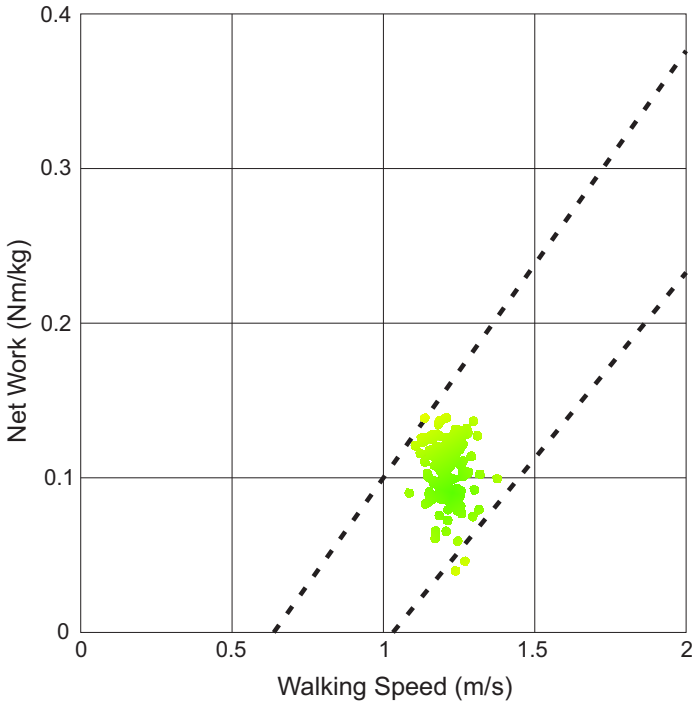

Figure 2.1: BiOM tuning interface. An example of the real-time display used to tune the actuation settings of the BiOM prosthesis (adapted from the BiOM user's manual [14]). As the subject walks, a dot appears on the screen that indicates the work done by the prosthesis at the user's chosen walking speed. The prosthetist modifies the power setting of the device until the dots fall into the 95% confidence interval for healthy biological net ankle work (dashed lines) and the patient is satisfied with the device performance.

Subject	Stiffness (%)	Power at Fast Cadence (%)	Power at Slow Cadence (%)	Power Sensitivity (%)	Timing at Fast Cadence (%)	Timing at Slow Cadence (%)	Cadence Range (%)
S1	12	57	40	0	44	20	47
S2	45	42	41	0	44	44	49
S3	65	45	24	0	44	16	42
S4	50	48	25	0	44	16	81
S5	63	44	33	0	44	20	57
S6	81	35	25	0	44	16	45
S8	75	38	28	0	66	30	66
S9	0	26	26	0	44	16	36
S10	75	39	20	0	65	20	28

Table 2.2: Prosthetist-chosen BiOM settings.

2.2.3 Experimental Protocol

After the initial acclimatization period, participants completed six different conditions in a randomized order while walking on a treadmill wearing the BiOM. Each of these conditions corresponded to a different ankle power setting. The conditions tested were 0 %, 25 %, 50 %, 75 %, 100 % power, and the prosthetist-chosen power setting (Figure 2.2).

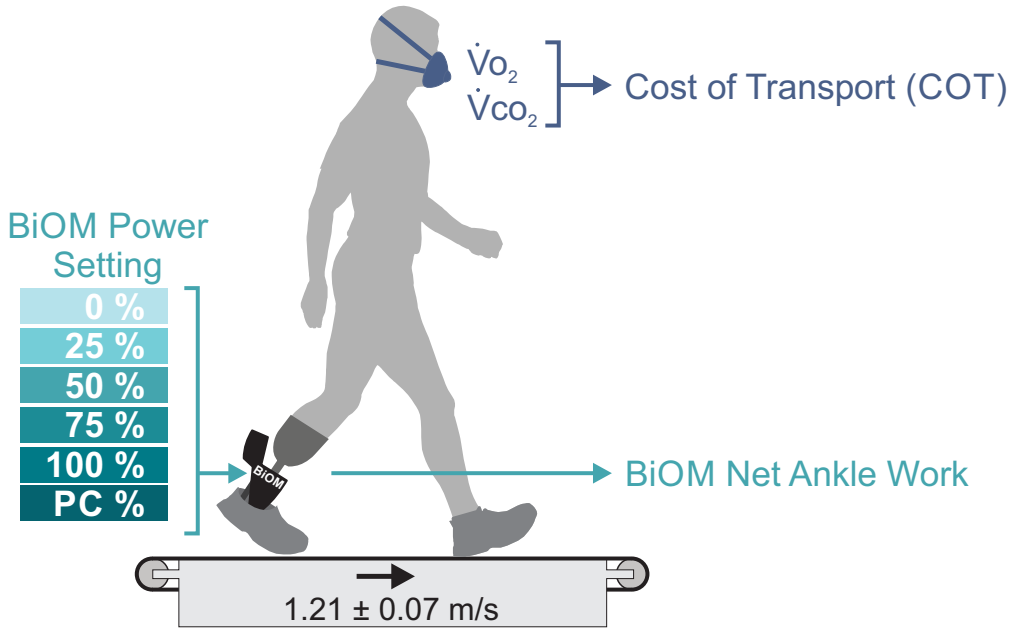

Figure 2.2: Experimental setup. In this experiment, 10 transtibial amputees walked on a treadmill wearing the BiOM powered ankle prosthesis with 6 different power settings: 0%, 25%, 50%, 75%, 100%, and the parameter setting chosen by the prosthetist during fitting (prosthetist-chosen, PC). We measured energy expenditure using indirect calorimetry and calculated the cost of transport (COT) from energy expenditure and treadmill speed. Treadmill belt speed was normalized to leg length, and determined using a Froude number of 0.16, which corresponds to the typical preferred walking speed for individuals with transtibial amputation [55]. In our experiment, subjects walked at a mean (SD) speed of 1.21 (0.07) m/s. We also collected net ankle work from the BiOM for each step during all conditions.

During all treadmill walking trials, participants wore a support harness (Likorail, Hill-Rom, Chicago, IL), which prevented them from falling in case of loss of balance, but did not provide any upward force or body weight support. Treadmill belt speeds were normalized according to a Froude number of 0.16 to scale belt speed to each participant's mean leg length [54]; the mean (SD) walking speed for all participants was 1.21 (0.07) m/s.

At the beginning of each session, subjects were fit with a lightweight portable metabolic system (K4b^2, Cosmed, Rome IT) consisting of a mask covering the nose and mouth and a portable unit attached to a harness. The metabolic system measured the rate of oxygen consumption ($\dot{V}O_2$) and carbon dioxide production ($\dot{V}CO_2$). Testing was generally performed first thing in the morning, and all participants were instructed to fast for at least two hours prior to arriving at the laboratory, as time of day and food consumption can affect one's basal metabolic rate [60,135]. Participants were initially asked to remain seated for a period of at least 10 minutes to establish a baseline metabolic

energy cost for each individual. We subsequently recorded participants' metabolic energy cost during all treadmill walking trials. Additionally, we recorded net ankle work and peak ankle power from each step of the prosthesis via the tablet computer for all walking conditions. The BiOM calculates ankle joint work from its on-board sensors as the integral of ankle torque-angle curve; it calculates peak ankle power by multiplying the measured ankle angular velocity with the ankle torque [63]. Ankle torque is calculated as the combined torque of the motor and the estimated torque of the carbon fiber foot [9].

Data were collected during one experimental session for most participants, but two participants (S1 and S5) required two sessions to complete all six conditions due to time constraints or physical fatigue. For these two subjects, we compared seated resting $\dot{V}O_2$ measurements from both experimental sessions against the between-day minimum detectable change threshold to assess the consistency of metabolic measurements across days. These between-day differences were less than the minimum detectable change threshold [34], and were therefore regarded as within the expected variation and not meaningful. As such, we combined the data collected across both days for analysis. Two trials were excluded due to BiOM battery failure in the middle of the walking trial, which cut the trials short (S2: 50 % condition, S8: 100 % condition). We were unable to re-collect these trials due to participant fatigue and time constraints. The BiOM battery also died during Subject 4's 75 % condition, but we were able to collect 5 steps at this power setting at the end of the session. During the data collection session for Subject 7, both BiOM batteries became defective and would not charge. As we were only able to collect complete data for 2 of 6 conditions, this participant's data was excluded from all analyses.

2.2.4 Data Analysis

Participants walked with each power setting for a minimum of five minutes in order to reach a steady-state oxygen consumption level, and then three minutes of steady-state data were collected and subsequently analyzed. We confirmed that the last three minutes of recorded breath measurements for each walking condition were indeed at a steady state by testing that they met

three criteria. These criteria were 1) average $\dot{V}O_2$ and 2) average $\dot{V}CO_2$ had less than 10 % variability between minutes [116, 136], and 3) the average respiratory quotient (RER) was between 0.7 and 1.0 [30]. We estimated metabolic power from the steady-state $\dot{V}O_2$ and $\dot{V}CO_2$ measurements using the Brockway equation [18].

To generalize energy expenditure across different walking velocities between participants, we calculated the cost of transport (COT) by dividing the metabolic power by the walking velocity and body weight while wearing the BiOM device. COT is a dimensionless quantity. We averaged the last 30 steps of net ankle work data and peak ankle power data from the BiOM to obtain the average net ankle work and average peak ankle power for each condition, and normalized these quantities by body mass (including the weight of the prosthesis). In the case of Subject 4 walking at the 75 % condition, they are the average value of the five collected strides.

2.2.5 Statistical Analysis

We use a generalized linear model to determine the effect of BiOM power setting (0 %, 25 %, 50 %, 75 %, 100 %) on COT. Power was a fixed factor, while subjects was a random factor. Post-hoc paired t-tests were used to explore significant differences between power settings. To determine a subject's energetically optimal power setting, we fit a third order polynomial to each subject's cost of transport data from the six power conditions (including the prosthetist-chosen). We then identified the minimum of each subject's third order polynomial, and chose the tested power setting (0 %, 25 %, 50 %, 75 %, 100 %, or prosthetist-chosen) closest to the minimum; this power setting is called the *best tested* power setting. This method of identifying the best tested power setting was chosen to minimize post-hoc selection bias, given the high level of breath-by-breath variability in the metabolic measurements and the sparse sampling of the parameter space. We used paired t-tests to evaluate the differences between the prosthetist-chosen and best tested power settings. Statistical comparisons were made using SPSS (IBM, Chicago, IL) with a level of significance of .05. Additionally, we calculated Pearson's correlation coefficients between COT, average net ankle work, and BiOM power setting using MATLAB (MathWorks, Natick, MA).

2.3 Results

Overall, there was a significant main effect of BiOM power setting on cost of transport (COT) ($p < .001$). On average, power settings higher than 50 % resulted in lower COT than 25 % or 0 % (Fig. 2.3). Post-hoc pairwise t-tests revealed significant differences between 0% power and 50% ($p = .02$), 75% ($p < .001$), and 100% ($p = .01$) power settings. There was also a significant difference between 25% power and 50% ($p < .001$), 75% ($p < .001$), and 100% ($p < .001$) power settings. There were no differences between the two lowest power settings ($p = .16$) and the three highest power settings ($p > .57$).

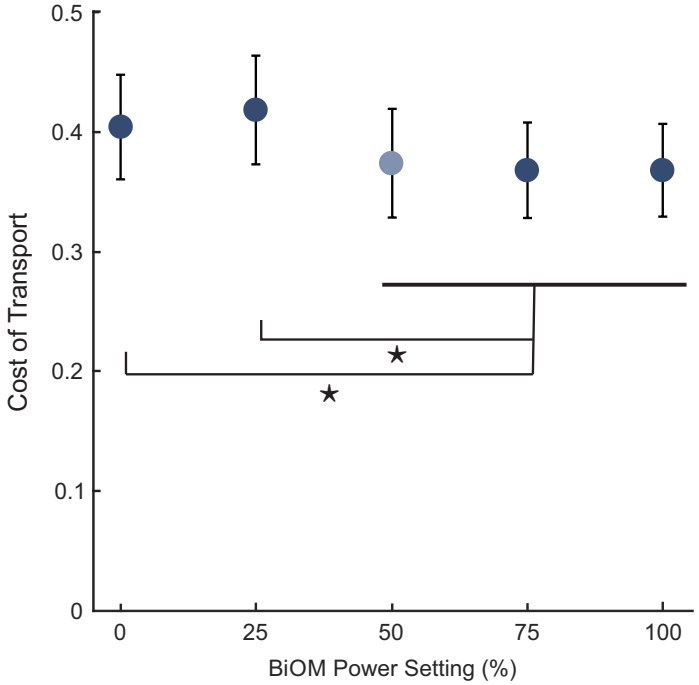

Figure 2.3: Average cost of transport at each BiOM power setting. On average, subjects exhibited a lower cost of transport (COT) as power setting increased. Data points represent the group mean COT (dark blue: n=9, light blue: n=8, error bars: ± 1 standard deviation (SD)). Stars ($\star$) indicate a statistically significant difference between conditions ($p < .05$). For example, 0 % had significantly higher COT than 50 %, 75 %, and 100 % power conditions.

Examining subject-specific responses, we observed considerable variability between subjects (Fig. 2.4). Yet, for all subjects, the best tested power setting was higher than the prosthetist-chosen power setting. Five (5) of 9 subjects had best tested power settings at 75 %, while the remaining 4 subjects were at 100 %. Additionally, some subjects showed an increase in COT when the device provided some power (25 %) compared to no power (0 %).

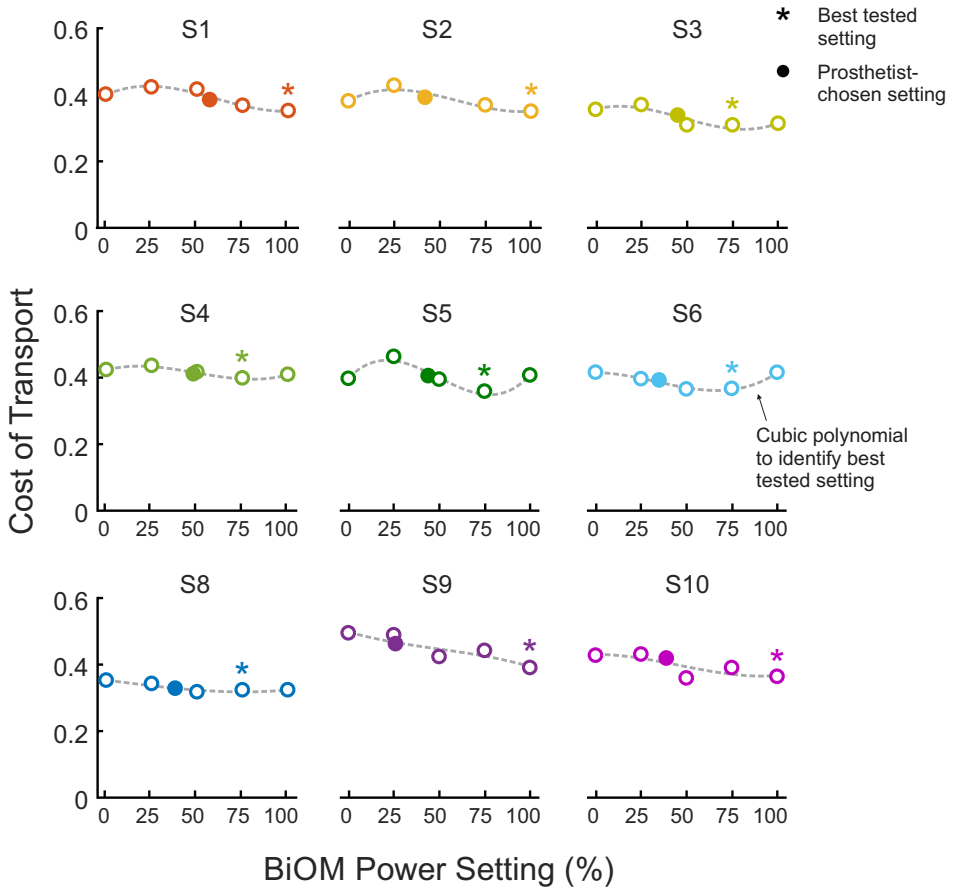

Figure 2.4: Subject-specific relationships between cost of transport and BiOM power setting. We observed large inter-subject variability in the relationship between power setting and the cost of transport (COT). Yet, for all subjects, the best tested power setting ($*$) was higher than the prosthetist-chosen power setting ($\bullet$). Each subject's best tested power setting was defined as the tested power setting closest to the minimum of the best fit cubic polynomial. This method was chosen to accommodate the noisy breath-by-breath measurements of metabolic cost and the sparse sampling of the parameter space. The third-order polynomials (dashed gray line) are presented for reference.

There was a significant difference in the magnitude of COT between the prosthetist-chosen setting (0.39 (0.04), mean (SD)) and best tested setting (0.36 (0.03)) (Fig. 2.5). On average, this difference in COT was 0.04 (0.02), which corresponds to an 8.8 (4.6) % reduction. The mean prosthetist-chosen power setting was 41.6 (8.7) % (corresponding to mean ankle work of 0.11 (0.06) J/kg) (Fig. 2.5). The mean best tested power setting was 86.1 (13.2) % (corresponding to mean ankle work of 0.24 (0.07) J/kg). The mean participant-specific difference between prosthetist-chosen and best tested power settings was 44.6 (16.2) % (corresponding to mean ankle work of 0.12 (0.09) J/kg).

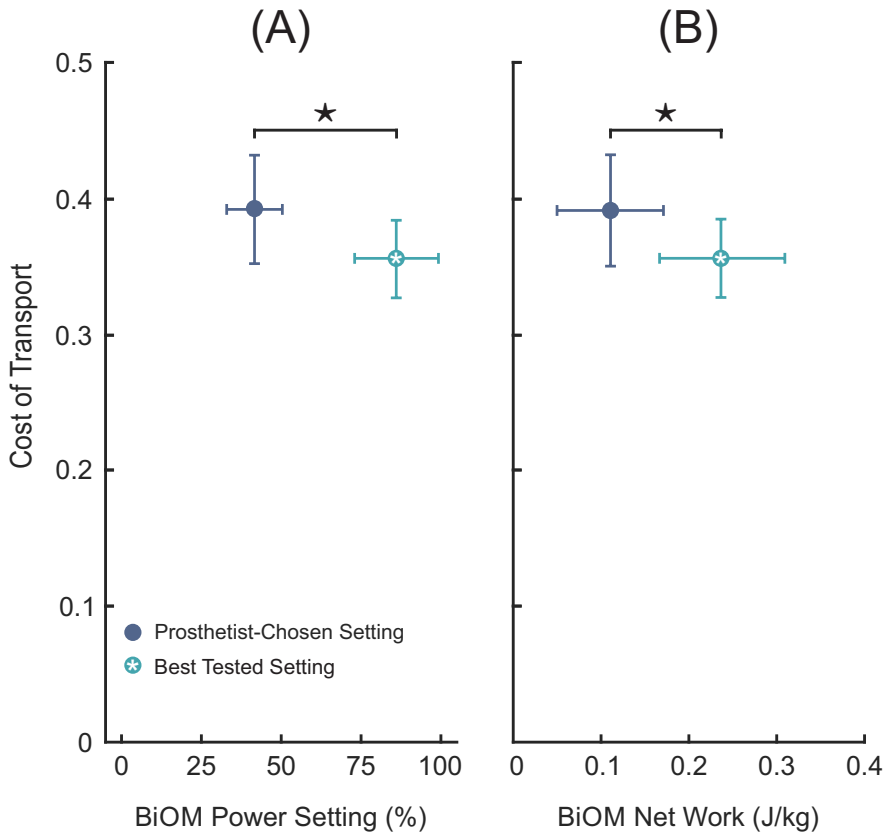

Figure 2.5: Comparison between cost of transport at prosthetist-chosen and best tested power settings. (A) On average, subjects walking with their best tested power setting had significantly lower cost of transport (COT) compared to walking with the prosthetist-chosen power setting ($p < .001$). The mean (SD) prosthetist-chosen power setting was 41.6 (8.7) %; the mean best tested setting was 86.1 (13.2) %. The star ($\star$) indicates a significant difference in the magnitude of COT between the prosthetist chosen and best tested power settings. Error bars represent ± 1 standard deviation in COT (vertical) and power setting (horizontal). **(B)** The corresponding mean (SD) net ankle work for the prosthetist-chosen and best tested conditions were 0.11 (0.06) J/kg and 0.24 (0.08) J/kg, respectively.

We also investigated how the power setting related to the net work performed by the BiOM. Overall, individual subjects showed a positive correlation ($r = .90$ (.07), mean (SD)) between power setting and net ankle work (Fig. 2.6). However, the relationship between power setting and ankle work was not strictly monotonic for all participants. Several subjects exhibited a plateau in ankle work as the power setting increased past 50 %, and some even elicited less ankle work at higher power settings. Additionally, we observed variable responses between no power (0 %) and a small amount of power (25 %) for different subjects. Five (5) of 9 subjects elicited considerably more ankle work at the 25 % condition than at the 0 % condition (mean (SD) differ-

ence: 0.07 (0.02) J/kg, n=5), while 4 other subjects elicited much more modest increases in ankle work between these conditions (mean (SD) difference: 0.01 (0.00) J/kg, n=4). Similar trends were observed in the relationship between power setting and average peak ankle power (Fig. 2.6).

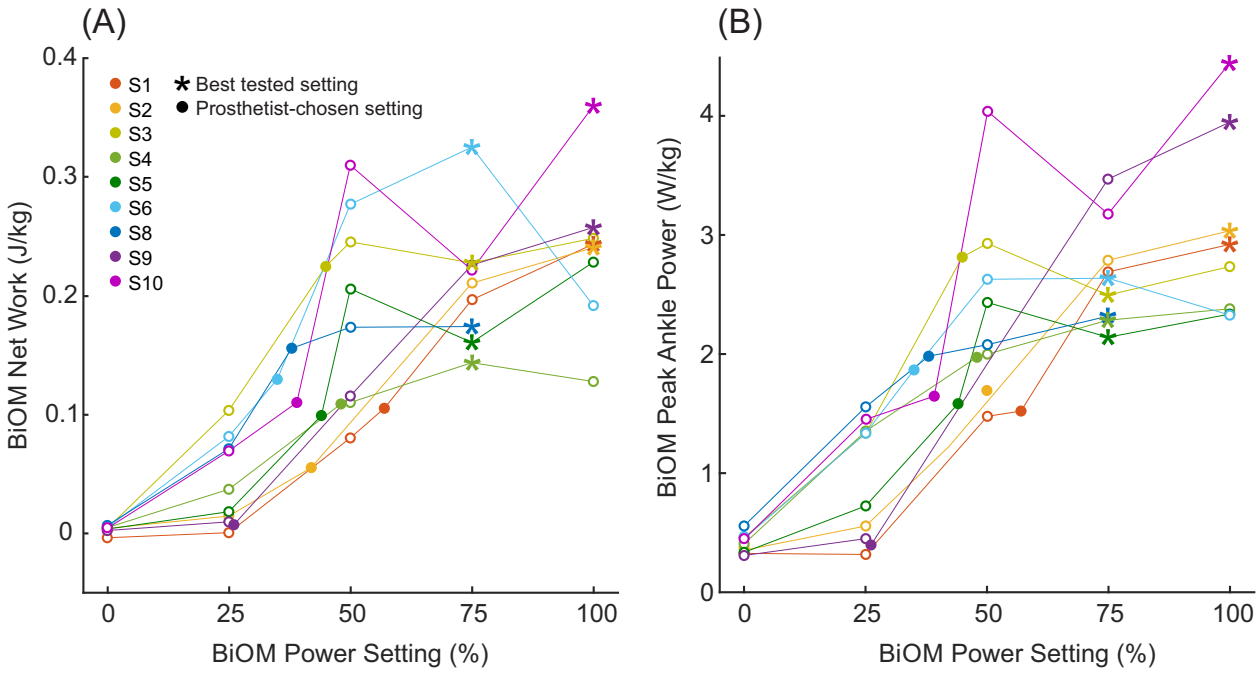

Figure 2.6: BiOM net work and peak ankle power as a function of power setting. (A) We found a linear correlation (Pearson's $r = .90$ (.07), mean (SD)) between power setting and net ankle work for individual subjects. **(B)** We found a linear correlation ($r = .90$ (.07)) between power setting and average peak ankle plantarflexion power for individual subjects. However, not all subjects exhibited a monotonically increasing relationship, and we observed a plateau in ankle work past the 50% power setting for some subjects. Filled circles (•) indicate the prosthetist-chosen power settings and corresponding net ankle work. Asterisks ($*$) indicate the subjects' best tested power settings. Net ankle work for each condition was calculated as the mean of the ankle work from the last 30 steps for all conditions except Subject 4's 75 % condition, which was the average of 5 steps.

Given that the relationship between BiOM power setting and net ankle work was not perfectly linear, we investigated the relationship between net ankle work and COT (Fig. 2.7). With all subjects pooled, a significant correlation between net ankle work and COT was identified ($r = -.55$, $p < .0001$). The best fit linear model resulted in a coefficient of determination (R^2) of .30. On an individual basis, subjects exhibited a stronger correlation between net ankle work and COT, with an mean (SD) correlation coefficient of $r = -.82$ (.15). Six (6) of 8 subjects (excluding S8) had best tested power settings that corresponded to their maximum delivered ankle work (*i.e.,* the

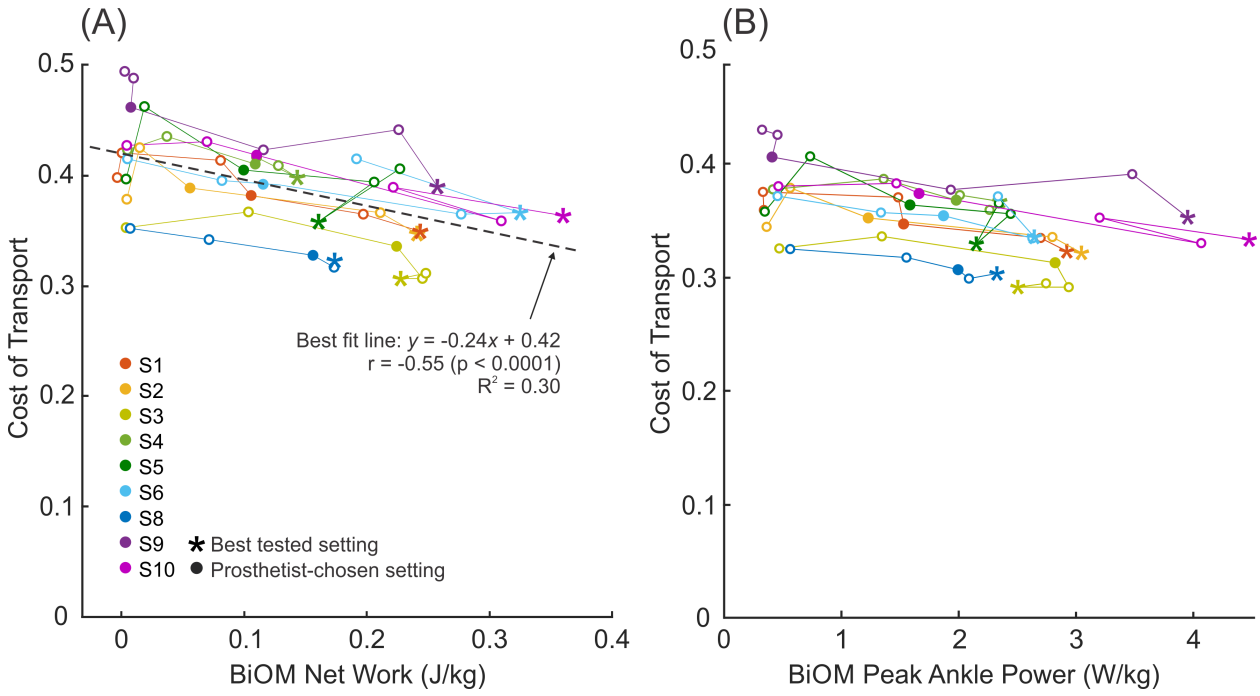

Figure 2.7: Cost of transport as a function of BiOM net work and peak ankle power. **(A)** With all subjects pooled, there was a moderate linear correlation between cost of transport (COT) and net ankle work (Pearson's $r = -.55$, $p < .0001$). The best fit linear model (dashed line) resulted in $R^2 = .30$. Individually, subjects exhibited a stronger linear relationship between COT and net ankle work ($r = -.82$ (.15), $R^2 = .69$ (.22), mean (SD)). The majority of subjects' best tested power settings corresponded to their maximum net ankle work. **(B)** Subjects individually exhibited a moderately strong linear correlation between cost of transport (COT) and average peak ankle power ($r = -.77$ (.15), mean (SD)), with large inter-subject variability. The majority of subjects' best tested power settings corresponded to their maximum ankle power. Starting from the left, colored lines connect increasing power settings for individual subjects. Filled circles (•) indicate the prosthetist-chosen power settings. Asterisks ($*$) indicate the best tested power settings. Net ankle work for each condition was calculated as the mean of the ankle work from the last 30 steps for all conditions except Subject 4's 75 % condition, which was the average of 5 steps.

asterisks in Fig. 2.7 lie farthest to the right). This did not coincide with the 100 % power setting for Subjects 4 and 6, whose best tested power settings were 75 %. The remaining two subjects analyzed (S3 and S5) had best tested power settings that did not correspond to their maximum ankle work. We do not have BiOM data from Subject 8 for the 100 % condition, so we can not evaluate this relationship for this subject. Similar trends were observed in the relationship between average peak ankle power and COT (Fig. 2.7).

2.4 Discussion

In this study, we evaluated the influence of different prosthetic ankle power settings on users' metabolic cost, using a commercially available powered prosthesis (BiOM). We hypothesized that 1) to minimize their energy cost, users would require a higher power setting than the power setting chosen by the prosthetist, which approximated the work of the biological ankle, and 2) the highest power setting (100 %) would not be the optimal power setting to minimize metabolic cost.

In support of our first hypothesis, we found that the best tested power setting was higher than the prosthetist-chosen power setting for all subjects. Moreover, we found that on average, subjects walking with their best tested power setting significantly reduced their cost of transport (COT) by 0.04 (0.02) compared to walking with the prosthetist-chosen power setting. This change is larger than the within-day minimum detectable change (MDC) for COT reported by Davidson *et al.* [34], which is 0.022. Therefore, we consider this difference a meaningful change and not likely due to measurement variation. On an individual basis, 6 of the 9 subjects analyzed had differences in COT greater than the MDC; the exceptions to this were Subject 1 (0.167), Subject 4 (0.012), and Subject 8 (0.005). The cost of transport calculated here is similar to values reported in other studies of people with transtibial amputation walking with the BiOM at similar walking speeds. In this study, the average COT for the prosthetist-chosen setting was 0.39 (0.04) , while the average was 0.40 (0.05) in Gardinier *et al.* [52] and 0.36 in Herr *et al.* [63].

Since the prosthetist-chosen power setting was chosen to approximate the work done by a biological ankle during walking, our results suggest that individuals with transtibial amputation may require ankle work in excess of biological norms in order to reduce their metabolic effort. The biological net ankle work for non-amputee subjects walking at 1.25 $^{m}/_{s}$ is approximately 0.1 $^{J}/_{kg}$ [63, 64]. On average, the best tested power setting in our study corresponded to net ankle work of 0.24 (0.07) $^{J}/_{kg}$, which is roughly double that of non-amputee subjects. As aforementioned, this finding is supported in studies conducted with able-bodied individuals wearing both exoskeletons [94, 119] and an ankle prosthesis emulator [23]. In contrast, a recent study by Quesada *et al.* [131] did not show an effect of prosthesis work on the metabolic work rate of amputee subjects. Differences

between our study and the study by Quesada *et al.* could have arisen from a variety of factors, including the functional level of the subject population and/or different prosthesis controllers. The subject cohort analyzed in our study comprised six K4 and three K3 individuals, while the cohort tested by Quesada *et al.* included six K3 and no K4 individuals. Gardinier *et al.* [52] found that individuals with K4 functional level were significantly more likely to receive an metabolic benefit from the BiOM than those with a K3 level, which may partially explain the different outcomes of these studies. Additionally, the various prosthesis emulator work conditions tested in by Quesada *et al.* were generated by modifying the torque-angle relationship, and as such, the prosthesis work remained mostly constant between steps. In contrast, the BiOM device is controlled using a reflexive controller, which utilizes a neuromuscular model to command a torque at each step, and varies depending on how the user loads the prosthesis [38, 63, 114]. The reflexive controller and powered plantarflexion on the BiOM device are hypothesized to help users maintain balance, especially on variable terrain [38, 53]. Similarly, Kim *et al.* [89] demonstrated that a stabilizing controller which modulates ankle push-off work each step reduced metabolic cost and step width variability in able bodied-users walking with an ankle prosthesis emulator. The reflexive nature of the BiOM's controller may have played an important role in the observed reduction of metabolic cost in the amputee subjects in our study, perhaps by reducing balance-related compensation efforts or allowing users to explore different assistance strategies by changing how they load the device step-to-step. Further studies will be required to determine the exact role that the reflex controller plays in reducing energy cost in prosthesis users.

The results of this study did not fully support our hypothesis that we would see an increase in COT at the highest power setting, indicative of too much ankle power. In fact, no significant group differences in COT were found between the 50 %, 75 %, and 100 % conditions (corresponding to mean net ankle work of 0.19 (0.08) J/kg, 0.22 (0.05) J/kg, and 0.24 (0.07) J/kg respectively). When examining the data on a subject-specific basis, we saw that 5 of the 9 subjects analyzed had the best tested power setting at 75 %. Yet, of these subjects, only Subjects 5 and 6 exhibited the anticipated trend of increased COT at 100 % (both S5 and S6 increased their COT by 0.049 from

75 % to 100 %; the minimum detectable change (MDC) is 0.022). The other three subjects (S3, S4, S8) had changes in COT smaller than the MDC between 75 % and 100 %. The remaining three participants had their best tested settings located at the maximum (100 %), which did not support our hypothesis. It is possible that these findings are also tied to the reflexive controller of the BiOM. Many subjects exhibited a plateau or a decrease in net ankle work past the 50 % power setting, which also corresponded to little change in COT past this power setting. Accordingly, it appears that some subjects down-regulated the amount of work they received from the device at the higher power settings, most likely by not loading the device with their full body weight and exploiting the reflexive nature of the controller. As our experimental setup did not include force plates, we were unable to experimentally confirm this in the current study.

When we examined the relationship between net ankle work and COT (with all subjects pooled), we found a moderate linear correlation ($r = -0.55$). Yet, the best fit linear model only resulted in $R^2 = 0.30$, which highlights the variability between subjects. On an individual basis, we found much stronger linear relationships between net ankle work and COT ($r = -0.82$ (0.15); R^2 =0.69 (0.22)). Accordingly, most of the subjects tested (6 of 8, excluding Subject 8) had best tested conditions that corresponded to their maximum ankle work (see Figure 2.7). For Subject 3, the difference in COT between the best tested condition (75 %) and the condition with the maximum net ankle work (50 %) was well below the minimum detectable change for COT (< 0.01) [34]; this difference was much larger in Subject 6 (0.04). We could not complete this analysis for Subject 8, because we were unable to collect BiOM ankle work data during the 100 % condition. Therefore, the power setting that maximized the ankle work delivered by the device was energetically optimal (or very close to energetically optimal) for nearly all participants in this study. Given this trend, however, it is also interesting to note that the net ankle work that corresponded to the best tested condition was quite variable between subjects, ranging from approximately 1.5 times (min: 0.14 J/kg) to 3.5 times (max: 0.36 J/kg) the work of a typical biological ankle.

The reflexive controller and corresponding variability in net ankle work that users produced between 0 % and 25 % power conditions could also explain why we saw an increase in COT for

some subjects between these conditions. Although there were no significant group differences between these conditions, we observed an increase in COT between the 0 % and 25 % power conditions greater than the MDC for Subjects 1, 2, and 5 (mean (SD) difference: 0.04 (0.02), n=3, see Figure 2.4). Qualitatively, we noticed that these same subjects were those who elicited only a modest increase in net ankle work between 0 % and 25 % power conditions (mean (SD) difference: 0.01 (0.01) J/kg, n=3, see Figure 2.6). Similarly, those five subjects (S3, S4, S6, S8, S10) who elicited substantially more ankle work at 25 % than at 0 % (mean (SD) difference: 0.07 (0.02) J/kg, n=5) exhibited a decrease or very small increase in COT between these conditions (mean (SD) difference: 0.00 (0.01)). The final subject (S9), exhibited very little change in ankle work or COT between the 0 % and 25 % power conditions. These results seem to classify our subject cohort into two groups when a small amount of power was provided: those who took advantage of the power, and those who appeared to modify their behavior to avoid receiving power from the device and correspondingly increased their COT. If the work the device performed at this power setting was disruptive to the natural walking dynamics or balance of the subject, it is possible that they were "fighting" the device by increasing muscular co-contraction or adopting atypical compensatory gait mechanisms, which may have driven up metabolic cost for those participants [48]. Given our small sample size and the high level of breath-by-breath variability in the metabolic measurements, further analysis of additional biomechanical quantities (*e.g.,* electromyography, inverse kinematics, spatiotemporal parameters) will be conducted to investigate these mechanisms in detail and to determine quantitative relationships, if any.

This study presents quantitative evidence regarding how users respond differently to various ankle power settings and exploit the BiOM's reflexive controller in order to reduce their metabolic cost. There are several hypothetical reasons that users might adapt their gait to reduce the plantarflexion power they receive from the device at various power settings. Users could be actively off-loading the device at higher power settings because they feel uncomfortable or unstable, and resultant compensatory gait strategies or "fighting" the device could lead to higher energy consumption. However, it is also possible that users are subconsciously or passively adapting their

interaction with the device in order to optimize a physiological objective function, such as minimal metabolic cost [36], minimal impact forces [42], or maximal stability [66], among others. Further detailed analyses of additional biomechanical measures (*e.g.,* electromyography, ground reaction forces) are required to elucidate the underlying causes for these observations. The results of this study provide some insight into additional elements of prosthetic control that may be necessary to reduce metabolic cost, beyond only the magnitude of ankle power delivered. The complex interactions between prosthetic control, metabolic cost, muscle activity, joint kinetics, stability, and patient satisfaction remain a hugely important topic for continued future research in order to inform powered prosthetic ankle prescription and improve patient outcomes.

Our study is not without its limitations. First, there was a limited sample of participants, and due to the walking stamina necessary to complete the experiment, we only tested active, healthy individuals. People with higher levels of ambulatory function (K3-K4) are capable of walking with variable cadence and performing advanced ambulation tasks, and may be able to better adapt their gait in order to take advantage of power from the device. Additional studies with a modified protocol will be necessary to determine how these results extend to individuals with lower levels of ambulatory function. Second, it is possible that the five-minute acclimation time was insufficient for some participants to adjust to each power setting. Compared to other studies in which the users had hours [63], multiple sessions [131], or even weeks [145] to acclimatize to a powered device and its conditions, the users in our study received less time to familiarize themselves with the power settings. Although the results from our study suggest that some participants were able to adapt to the power delivered from the device in this short amount of time, it is possible that some subjects may have required more time and/or specific training to fully adapt to each setting. Third, to prevent physical fatigue and respect time constraints, we tested participants in increments of 25 %, which is a very coarse sampling of the parameter space, and may have limited our ability to identify the true energetically optimal power setting for all users. Due to device limitations, we could not test users past 100 % power (0.24 (0.07) J/kg), so it is also possible that a more energetically favorable setting exists outside our tested range. As we do not know the exact physiological

relationship between power setting and COT, we can not currently extrapolate these results outside the tested range. Fourth, our experimental setup did not include an instrumented force treadmill so we were not able to experimentally validate the agreement between the BiOM's step-by-step calculations of net ankle work and average peak ankle power and those values obtained through inverse dynamics. Future studies with the BiOM prosthesis that include an instrumented treadmill will further improve the generalizability of this study's results. Finally, it is important to point out that our study was limited to evaluating the best power setting while participants walked on a level treadmill at a constant speed. It is likely that the optimal power setting would change when users walked at different speeds, at an incline or decline, or over variable terrain, so this study can not make universal claims about the optimal power setting for all tasks.

In conclusion, to minimize their metabolic energy consumption, subjects in this study required a higher power setting than the setting chosen by the prosthetist to approximate biological ankle kinetics. On average, the power setting setting which minimized energy cost corresponded to approximately double the net ankle work of the biological ankle. Furthermore, subjects walking at their best tested power setting exhibited a meaningful decrease in cost of transport compared to walking with their prosthetist-chosen power setting, which suggests that individuals may benefit metabolically from prescribed ankle power that exceeds biological norms. However, the varied responses between subjects also point to the need for subject-specific parameter tuning. As one solution, recent work has demonstrated the feasibility of automatically tuning assistive device parameters to minimize metabolic cost (*i.e.,* body-in-the-loop optimization) [35, 44, 92, 189], and continued research in this area has the potential to impact clinical device prescription. Finally, subjects' net ankle work was highly variable at different power settings, likely due to the reflexive controller of the BiOM and the user's adaptation to the device. As such, future work should focus on quantifying the mutual adaptation of the human user and the device to inform the design of optimal powered prosthesis controllers for individuals with transtibial amputation.

CHAPTER 3

Evaluating Physiological Signal Salience for Estimating Metabolic Energy Cost from Wearable Sensors

3.1 Background

Wearable assistive robotic devices, such as powered prostheses or exoskeletons, have the potential to aid and restore mobility in users who suffer from conditions such as stroke, spinal cord injury, or amputation. They can also augment performance in able-bodied individuals. Some of these devices are specifically developed to reduce the energy cost of ambulation tasks such as walking, running, or climbing stairs [7, 9, 43, 100, 120]. From a human-augmentation perspective, the less energy a person uses, the longer they can perform a particular activity. This reduction in energy cost also has clinical relevance, as individuals with movement disabilities often have a higher energetic cost of transport than able-bodied individuals, which further limits their activity level [96, 171].

The assistive devices employed for this purpose often have a large number of tunable controller settings that dictate their behavior and their interaction with the user. Such settings might include stiffness, actuation power, or actuation timing. In the clinic, these settings are tuned heuristically, often relying on visual inspection by a trained clinician and verbal feedback from the patient. This

This chapter has been previously published in full in the *Journal of Applied Physiology* [77].

approach is time consuming and subjective—two issues that are amplified for research prototypes that often have an even higher number of settings and tunable controller parameters. Identifying the best parameters for each individual is a challenge. As one proposed solution, "body-in-the-loop" optimization algorithms have been successfully employed as a method to quickly tune assistive device parameters on a subject-specific basis and in real time [35, 44, 90, 92, 189]. Body-in-the-loop optimization is the process of iteratively and automatically tuning parameters of assistive robotic devices to minimize a physiological cost function while an individual is using the device. While the aforementioned studies have differed in their optimization algorithms, assistive device hardware, and tuning parameters of interest, all of them used metabolic energy cost (*i.e.,* how much energy the body requires on a cellular level) as the physiological objective to minimize.

The automatic evaluation of energetic cost that is required for this method is currently one of the rate-limiting factors in the translation of these algorithms to real-world assistive devices. The state-of-the-art experimental measurement system for energy cost is indirect calorimetry, a method which estimates energetic cost using measurements of oxygen consumption and carbon dioxide production [18]. commercially available indirect calorimetry systems comprise a flowmeter embedded in a rubber mask that covers the nose and mouth of the user, and one or more processing units and battery packs typically secured to the user's body with a harness. While these systems are widely used, they are ill-suited for continuous use or long-term data collections because of the obtrustiveness of the mask and processing units. In addition to the bulky equipment, the data obtained from indirect calorimetry are extremely noisy, sparsely sampled, and dynamically delayed. As such, to obtain an accurate estimate of energy cost for a given activity, it is common practice to have individuals walk for 2–3 minutes to reach a steady-state oxygen consumption level [116, 154]. Once steady state has been reached, another 2–3 minutes of breath data are recorded and subsequently averaged into one estimate of energy cost for that activity. To mitigate some of these challenges, all body-in-the-loop optimization studies to date have leveraged the work of Selinger and Donelan [154], modeling and inverting the kinetics of the respiratory response as a first order linear system. While this approach enables researchers to use non steady-state breaths to estimate

energy cost, current body-in-the-loop optimization protocols still require 2–4 minutes to estimate the energy cost at a particular parameter setting [35, 44, 90, 92, 189]. These challenges prevent body-in-the-loop optimization algorithms from being used outside the lab during clinical fitting sessions and/or integrated into devices that individuals wear in their daily lives.

To realize the full potential of body-in-the-loop optimization algorithms, our estimates of energy cost need to be accurate, fast, comfortable (*i.e.,* no mask), and portable; current indirect calorimetry equipment cannot provide these properties. As such, the long-term goal of our work is to predict energy cost quickly and accurately using physiological signals from portable, wearable sensors, which can then be integrated into an online optimization of assistive device parameters. In this paper, we aim to quantify individual signal salience (*i.e.,* predictive capability) and discover which signals, or groups of signals, have the best predictive power for our application of body-in-the-loop optimization. In particular, we are interested in how well signals or groups of signals can predict *steady-state* energetic cost (*i.e.,* the amount of energy required to perform a particular activity), as well as generalize across different subjects and tasks.

This approach is enabled by recent developments in wearable sensor technology, that have resulted in devices that are capable of providing quick and accurate estimates of physiological quantities (*e.g.,* heart rate (HR), skin temperature) and information about movement (*e.g.,* acceleration). All these quantities relate to how much energy a person is expending. Several previous studies have utilized various types of sensors, signal pre-processing techniques (*i.e.,* filtering and feature extraction), and final predictive algorithms to estimate energy cost during a variety of activities. Commercial accelerometers placed on the chest, hip, feet, and/or wrists are historically the most common sensor used to predict energy cost, due to their simplicity and the inherent relationship between movement and energy expenditure [32, 33, 39, 61, 143, 163, 166, 167]. Heart rate is another common physiological signal used to estimate energy cost, and a linear relationship between heart rate and energy expenditure during sub-maximal physical activity has been established [86, 118, 137]. However, given that heart rate is tightly controlled by the nervous system, changes in heart rate can also be attributed to changes in emotional arousal and may require

individual calibration [2, 16, 86, 106]. Some of these factors have been mitigated by combining heart rate monitors with accelerometers [16, 17, 41], or biological parameters (*e.g.,* gender, age, weight) [86] in predictive algorithms. Some more recent studies have incorporated autonomic nervous system signals, such as skin temperature, humidity, and electrodermal activity into their estimates of energy cost [2, 3, 177]. Electrodermal activity (also called galvanic skin response), is a measure of skin conductance, which changes when an individual begins to sweat. While electrodermal activity has primarily been used as an indicator of emotional arousal or stress, changes in this signal have also been linked to changes in physical workload [126]. Skin temperature also has been widely used as a measurement tool for cognitive workload; decreasing skin temperature is associated with increased mental workload [126], but its relationship with metabolic work rate has not yet been well defined. Finally, muscle activity (*i.e.,* electromyography, EMG) and joint kinematics have been used to predict energy cost both in simulation [159], and experimentally [15, 67, 178]. Most often, EMG signals are processed in the time domain by calculating a linear envelope or mean activation to correlate with energetic cost [67], but frequency domain processing methods have also been utilized to resolve EMG signals into intensities to estimate energy cost during cycling [15, 178].

Many studies have concluded that predicting energy cost using a combination of sensors improves estimates of energy cost across participants and across physical activities [3,16,17,41,75,76, 143]. Recent work has thus utilized multi-sensor arrays embedded in textile clothing, such as EMG shorts [170] and the Hexoskin smart shirt (containing a waist-mounted accelerometer, heart rate monitor, and respiratory bands to measure breath frequency and minute ventilation) [13]. These systems enable high-fidelity measurement of physiological signals outside the laboratory setting. For example, the Hexoskin smart shirt uses two strain gauge bands positioned around the chest and abdomen to measure breathing frequency and minute ventilation; the physiological signals obtained using the Hexoskin smart shirt have been experimentally validated [175]. With these combinations of sensors, the most common predictive algorithm is multiple linear regression, which has been successful in predicting energy cost across multiple users and tasks. However, linear regres-

sion algorithms can produce large task-specific or subject-specific errors, so more sophisticated machine learning algorithms such as neural networks [12, 118, 143], and random forests [13, 39] have also been explored.

Most studies to date have investigated the predictive capability of only a small subset of sensors, which makes it challenging to compare which signals or groups of signals contain the most salient information for predicting energy cost across a variety of tasks and subjects. To design the best predictive algorithms for energetic cost, there is a need to understand the potential that individual signals and groups of signals hold for predicting energy cost. Altini *et al.* [3] compared the predictive power of a relatively large set of signals (two accelerometers, heart rate, electrodermal activity, skin humidity and respiration rate) across a range of physical activities, but only evaluated the performance of two *groups* of signals. Preliminary studies performed by our group investigated the predictive power of a large group of signals (including EMG, accelerometers, electrodermal activity, skin temperature, oxygen saturation, heart rate, and respiratory rate) during a small subset of activities [76], as well as the predictive power of different combinations of these signals [75]. The goal of the present work is to expand on previous methods and quantify individual signal salience for estimating energetic cost as well as characterize the ability of these signals to generalize to unknown subjects and unknown tasks. In this study, we focused on using simple signal processing techniques and prediction algorithms to stay as close as possible to the original sensor signals. This study is an important step toward the goal of improving the clinical viability and translation of body-in-the-loop optimization algorithms. Furthermore, the dataset made available in conjunction with this article [78] will enable researchers and clinicians across many fields to develop novel algorithms to predict energy cost from wearable sensors for a variety of applications.

3.2 Methods

3.2.1 Experimental Data Collection

Ten healthy subjects—8 male, 2 female, age: 27.4 (SD 4.5) years, height: 1.76 (SD 0.09) m, weight: 69.1 (SD 9.9 kg)—participated in this experiment, after giving informed consent to protocol approved by the University of Michigan Institutional Review Board. Each subject completed two experimental sessions in which they performed a variety of physical activities at various speeds and/or intensities (Table 3.1). The first experimental session consisted of sitting, standing, level walking, incline walking, and backwards walking on a treadmill (Bertec Corp., Columbus, OH); the second session consisted of sitting, standing, running on a treadmill, cycling on a stationary bike (Matrix Fitness, Cottage Grove, WI), and stair-climbing on a stairmill (Matrix Fitness, Cottage Grove, WI). For each activity, subjects stood quietly for 6 minutes, performed each speed/resistance condition for 6 minutes in a random order, and then sat quietly for 6 minutes. All changes between speeds or intensities were performed as instantaneous step changes. Subjects rested for approximately 10 minutes between activities.

During data collection, subjects wore a variety of sensors, detailed below and in Fig. 3.1. Oxygen consumption ($\dot{V}_{O_2}$), carbon dioxide production ($\dot{V}_{CO_2}$), breath frequency, and minute ventilation were recorded breath-by-breath using a portable respirometer (Oxycon Mobile, CareFusion, San Diego, CA). Breath frequency was defined as the number of breaths per minute (units: min^{-1}), and minute ventilation was defined as liters of expired air per minute (units: L/min). Heart rate (HR) was measured using a wireless heart rate monitor strapped around the chest (Polar Electro, Inc., Bethpage, NY). Blood oxygen saturation (SpO_2), defined as the percentage of oxygenated hemoglobin, was measured using a pulse oximeter (Oxycon Mobile, CareFusion, San Diego, CA) secured to the subject's left earlobe. The heart rate monitor and pulse oximeter were synchronized with the respirometer, so HR and SpO_2 were recorded breath-by-breath. Subjects wore 4 commercial tri-axial accelerometers (Opal, APDM Inc., Portland, OR) affixed to the left and right lateral ankle, the left hip, and the center of the chest using elastic velcro straps; these accelerometers mea-

Task	Speed	Intensity
Sitting	-	-
Standing	-	-
Level Walking	0.6 m/s	-
	0.9 m/s	-
	1.2 m/s	-
Incline Walking	0.6 m/s	4°
	1.2 m/s	4°
	0.6 m/s	9°
	1.2 m/s	9°
Backwards Walking	0.4 m/s	-
	0.7 m/s	-
Running	1.8 m/s	-
	2.2 m/s	-
	2.7 m/s	-
Cycling	70 rpm	Resistance 1
	70 rpm	Resistance 3
	70 rpm	Resistance 5
	100 rpm	Resistance 1
Stair Climbing	-	60 W
	-	75 W
	-	90 W

Table 3.1: Experimental activities.

sured accelerations in the x, y, and z axes at 128 Hz. Subjects also wore 2 bilateral wristbands (E4, Empatica, Milan, Italy) containing tri-axial accelerometers that recorded x, y, and z accelerations of the left and right wrist at 32 Hz. The Empatica wristbands also recorded electrodermal activity (EDA) and skin temperature at 4 Hz. Surface EMG electrodes (Biometrics Ltd., Ladysmith, VA) recorded muscle activity from 8 bilateral lower limb muscles: gluteus maximus (GMAX), biceps femoris (BF), semitendinosis (ST), rectus femoris (RF), vastus lateralis (VL), medial gastrocnemius (MGAS), soleus (SOL), and tibialis anterior (TA); EMG data were collected at 1kHz. All sensor signals were time-synchronized during collection.

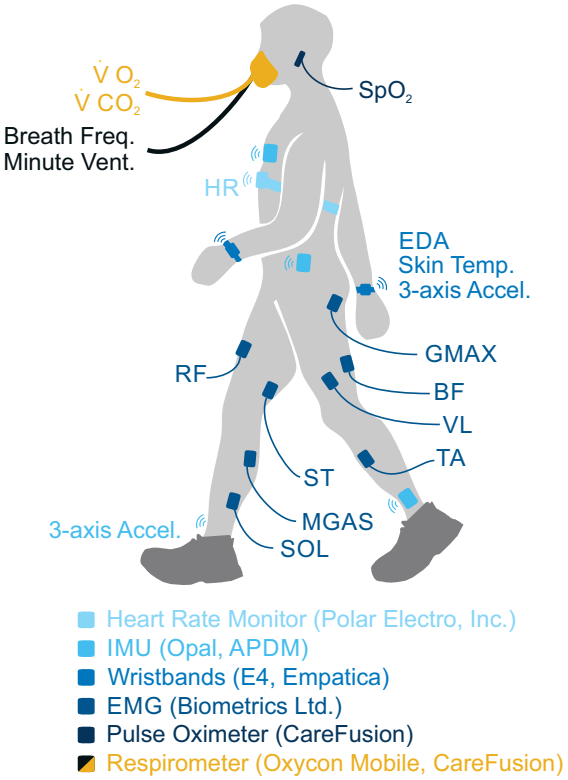

Figure 3.1: Physiological signals. Oxygen consumption ($\dot{V}_{O_2}$), carbon dioxide production ($\dot{V}_{CO_2}$), breath frequency and minute ventilation were measured breath-by-breath using a portable respirometer. Heart rate (HR) was measured using a wireless heart rate monitor strapped around the chest. Surface electromyography (EMG) electrodes recorded bilateral muscle activity from 8 lower limb muscles: gluteus maximus (GMAX), biceps femoris (BF), semitendinosis (ST), rectus femoris (RF), vastus lateralis (VL), medial gastrocnemius (MGAS), soleus (SOL), and tibialis anterior (TA). Skin temperature, electrodermal activity (EDA), and accelerations of the wrist were recorded using bilateral wrist sensors. An inertial measurement unit (IMU) placed on the chest, left hip, and ankles measured 3-axis limb acceleration. Blood oxygen saturation (SpO_2), was measured by a pulse oximeter secured to the subject's left earlobe.

Missing Data

Due to equipment malfunction, we were not able to collect complete data sets from all subjects. Subject 1 did not complete the stair-climbing activity. The right wristband worn by Subject 6 did not record data during Session 2 so their right wrist accelerometry, electrodermal activity, and skin temperature for running, cycling and stair climbing tasks were excluded from analysis. In addition, the following signals were excluded from analysis for the given task due to electrode disconnection: Subject 5's left gluteus maximus (walking), Subject 3's right gluteus maximus (cycling), and Subject 3's left semitendinosis (walking).

3.2.2 Ground Truth Energy Cost

We calculated whole-body metabolic energy cost from unfiltered $\dot{V}_{O_2}$ and $\dot{V}_{CO_2}$ using the Brockway equation [18], and normalized the data to subject body mass. We assumed that energetic cost was constant for a given 6-minute condition, and this value was obtained by averaging the final 3 minutes of steady-state energetic cost measurements for that condition. Henceforth, this value is referred to as the 'ground truth' energy cost for each condition. When the conditions were concatenated in time, the ground truth energy cost formed a stair-step function, with steps that are 6 minutes long. These data served as our target data (*i.e.,* the energy cost we aimed to predict). For each activity, the ground truth energy cost from the standing bout at the beginning of the trial was subtracted off to yield net energetic cost.

For this experiment, we used a portable metabolic system (Oxycon Mobile) to establish our ground truth energy cost, as opposed to a tethered system (*e.g.,* the Douglas bag method). The manufacturer's specifications for the Oxycon Mobile system report an accuracy of 0.05 L/min (3%) for measuring both $\dot{V}_{O_2}$ and $\dot{V}_{CO_2}$; the accuracy for measuring minute ventilation is reported as 0.05 L/min (2%). Some previous studies have also investigated the reliability of portable metabolic systems and found generally good measurement accuracy compared to other established methods (*e.g.,* [1, 141]).

3.2.3 Signal Processing

The total acceleration was calculated for each of the 6 accelerometers as the vector magnitude of the 3 axes ($\sqrt{x^2 + y^2 + z^2}$). We generated EMG linear envelopes by bandpass filtering the raw EMG signals between 30–350 Hz, full-wave rectifying, and low-pass filtering with a cutoff frequency of 5 Hz. Each subject's EMG linear envelopes were normalized to peak activation level of each muscle obtained across all activities performed during the same session. We calculated a composite sum of normalized linear envelope EMG signals (M_i) to provide an overall estimate of muscle activity for the left and right legs (8 muscles per leg):

$$M_{total(j)} = \sqrt{\sum_{i=1}^{8} M_i^2} \qquad j = \text{left}, \text{right}$$

Minute ventilation, breath frequency, $\dot{V}_{O_2}$, electrodermal activity, skin temperature, SpO_2, and heart rate were smoothed using a 1-minute Gaussian kernel with a cutoff frequency of 0.01 Hz (as in [13, 140]). Individual EMG linear envelopes, EMG composite sums, and vector magnitude accelerations were filtered using a 1-minute Gaussian kernel with a cutoff frequency of 0.1 Hz. EMG and acceleration signals were filtered more aggressively to account for the higher frequency content in these signals due to their fluctuation within the gait cycle.

3.2.4 Data Analysis

We employed two statistical methods to quantify the predictive capability of individual signals or groups of signals: 1) Pearson's correlation coefficients and 2) linear regression models. We evaluated the linear regression models using leave-one-*task*-out and leave-one-*subject*-out cross validation. We also found the best possible linear regression models from our data set (across all subjects) containing up to eight signals.

Pearson's Correlation Coefficients

The Pearson's correlation coefficient ($-1 < r < 1$) is a measure of the linear relationship between two variables, where 1 is perfect positive linear correlation, 0 is no linear correlation, and -1 is perfect negative linear correlation. We pooled data across all activities for one subject and one individual signal, and calculated the correlation coefficient between ground truth energy cost and the individual signal. We repeated this analysis for all 10 subjects, and calculated the mean and standard error across subjects. We then performed these calculations for all 16 individual signals both before and after filtering.

Although no formal cutoff exists to define 'good' correlation, in this study, we define a correlation greater than 0.7 as a strong linear correlation; this value agrees with other studies which

have investigated the relationship between physiological signals and energy cost (*e.g.,* [3, 167]). We also calculated the correlation of $\dot{V}O_2$ with ground truth energy cost to serve as a 'benchmark' for comparison; since $\dot{V}O_2$ is included in the calculation of ground truth energy cost, we expect this correlation to be very high. Yet, this correlation will not be perfect because breath-by-breath $\dot{V}O_2$ is a noisy signal with a dynamic delay, and only the final three minutes of $\dot{V}O_2$ measurements were included in the calculation of ground truth energy cost.

Linear Regression Models

We calculated single and multiple linear regression models of the following form,

$$\hat{\mathbf{y}} = b_0 + \sum_{i=1}^{k} b_i \mathbf{x}_i = \mathbf{X}\mathbf{b} \tag{3.1}$$

where k is the number of signals included in the regression model and m is the number of samples; $\hat{\mathbf{y}} \in \mathbb{R}^m$ is the vector of predicted energy cost; $\mathbf{X} = [\mathbf{1} \quad \mathbf{x}_1 \quad ... \quad \mathbf{x}_k] \in \mathbb{R}^{m \times (k+1)}$, where $\mathbf{x}_i \in \mathbb{R}^m$ are the signals included in the regression model, each interpolated and re-sampled at 1Hz, and $\mathbf{1} \in \mathbb{R}^m$ is a vector of ones; $\mathbf{b} = [b_0 \quad b_1 \quad ... \quad b_k]^T \in \mathbb{R}^{k+1}$ is the vector of signal weights, including an offset term.

We calculated 17 cross-validated single linear regression models with each of a smaller subset of signals, including 6 vector magnitude accelerations, 2 EMG composite sums, left/right electrodermal activity, left/right skin temperature, heart rate, oxygen saturation, breath frequency, minute ventilation, and $\dot{V}O_2$ (to serve as a benchmark for performance). In addition, we trained and evaluated 4 cross-validated multiple linear regression models (Table 3.2). For the multiple linear regression analysis, we divided the signals into two broad categories: the *global signals*, which give information about the state of the body as a whole (*e.g.,* heart rate, electrodermal activity, skin temperature), and the *local signals*, which give information about one particular limb segment (*e.g.,* accelerations, EMG). The first three multiple linear regression models contained 1) the local signals alone, 2) the global signals alone, and 3) the combination of global and local signals (see

Table 3.2 for a complete list of signals included in each model). For comparison to a commercially available portable sensor system (*i.e.,* the Hexoskin smart shirt [25]), the fourth multiple linear regression model contained the group of signals that can be obtained using the Hexoskin smart shirt (*i.e.,* breathing frequency, minute ventilation, heart rate, and waist accelerometry). This fourth group will henceforth be referred to as the 'Hexoskin signals.' All cross-validated single and multiple linear regression models were calculated using pre-processed and filtered signals.

	Vec. Mag. Accel.	Comp. Sum EMG	Electro-dermal Act.	Skin Temp.	Breath Freq.	Minute Vent.	Heart Rate	Oxygen Sat.
Local	×	×						
Global			×	×	×	×	×	×
Local + Global	×	×	×	×	×	×	×	×
Hexoskin signals	Hip only				×	×	×	

Table 3.2: Signal groups included in linear regression models.

Cross Validation

We utilized two evaluation methods to examine the effect of task- and subject-specificity on the predictive capability of each signal or group of signals: 1) leave-one-*task*-out cross validation, and 2) leave-one-*subject*-out cross validation. For each evaluation method, the data were partitioned into two sets, a training set (including training data, $\bar{\mathbf{X}}$, and training targets, $\bar{\mathbf{y}}$) and a testing set (including test data $\mathbf{X}'$, and test targets, $\mathbf{y}'$), according to Table 3.3. The cross validation was repeated until each *task* or *subject* had been left out once; therefore we conducted 6-fold leave-one-*task*-out cross validation and 10-fold leave-one-*subject*-out cross validation (Table 3.3).

For each fold (indexed by f), we used the training data ($\bar{\mathbf{X}}_f$) and training targets ($\bar{\mathbf{y}}_f$) to calculate the regression weights ($\mathbf{b}_f$) using the pseudoinverse,

$$\mathbf{b}_f = (\bar{\mathbf{X}}_f^T \bar{\mathbf{X}}_f)^{-1} \bar{\mathbf{X}}_f^T \bar{\mathbf{y}}_f \tag{3.2}$$

	Train	Test	Repeat	# of Folds (F)
Leave-one *task*-out	Subject $= \mathbf{S}$ Task $= \mathbf{T} \backslash \{t\}$	Subject $= \mathbf{S}$ Task $= t$	$t \in \mathbf{T}$	6
Leave-one *subject*-out	Subject $= \mathbf{S} \backslash \{s\}$ Task $= \mathbf{T}$	Subject $= s$ Task $= \mathbf{T}$	$s \in \mathbf{S}$	10

$\mathbf{S} = \{1, 2, ...10\}$ is the set of all subjects. If $s = \{1\}$, then $\mathbf{S} \backslash \{s\} = \{2, 3, ...10\}$
$\mathbf{T} = \{1, 2, ...6\}$ is the set of all tasks.

Table 3.3: Cross-validation methods for linear regression models.

We then used the identified signal weights ($\mathbf{b}_f$) to predict the energy cost ($\hat{\mathbf{y}}'_f$) from the test data ($\mathbf{X}'_f$),

$$\hat{\mathbf{y}}'_f = \mathbf{X}'_f \mathbf{b}_f \tag{3.3}$$

We compared the predicted energy cost ($\hat{\mathbf{y}}'_f \in \mathbb{R}^m$) to the ground truth energy cost (*i.e.,* the test targets, $\mathbf{y}'_f \in \mathbb{R}^m$) by calculating the root mean squared error ($\text{RMSE}_f \in \mathbb{R}$) between the two time-series signals,

$$\text{RMSE}_f = \sqrt{\frac{1}{m}(\mathbf{y}'_f - \hat{\mathbf{y}}'_f)^T(\mathbf{y}'_f - \hat{\mathbf{y}}'_f)} \tag{3.4}$$

We then computed the mean and standard deviation of the RMSE across all folds.

Best Possible Models

Finally, we determined the best possible combinations of signals from our data set. The goal of this analysis was to determine the best combinations of signals to use if an experimenter only chose to use a small number of sensors (here: fewer than eight). To achieve this, we pooled data across all subjects and all activities and trained every possible linear regression model using 1–8 of the 16 signals mentioned in the previous section. The number of tested models was calculated using the binomial coefficient ($_nC_k$, or "n choose k"), which calculates the number of ways to choose an unordered set of k elements from a fixed set of n elements,

$$ {}_nC_k = \frac{n!}{k!(n-k)!} \tag{3.5} $$

In this experiment, $n = 16$ and $k = [1,\ 2,\ ..,\ 8]$ so the number of tested models were,

$$ {}_{16}C_k = [16,\ 120,\ 560,\ 1820,\ 4368,\ 8008,\ 11440,\ 12870], \quad k = [1,\ 2,\ ..,\ 8] $$

For each k, we calculated the RMSE for all models, and then sorted the ${}_{16}C_k$ models based on their RMSE. We considered the model with the smallest RMSE the best model for a given k.

3.3 Results

To illustrate the general characteristics of the signals, time-series data and correlation coefficients for eight pre-processed and filtered signals are presented for one representative subject, along with ground truth energy cost (Fig. 3.2). Local signals (*i.e.,* acceleration and EMG) changed nearly instantaneously with a step change in energy cost, while global signals (*e.g.,* heart rate, electrodermal activity) demonstrated time-delayed responses to the same step change. Signals that were strongly correlated with ground truth energy cost (*e.g.,* heart rate and minute ventilation) were sensitive to small changes in energy cost. On the contrary, signals such as electrodermal activity, skin temperature, and oxygen saturation (SpO_2) had a coarser relationship with ground truth energy cost, and seemed to capture information about general increases or decreases in energy cost, while being unable to discern between individual steps.

The diverse set of tasks tested in this experiment resulted in a wide range of intensity levels between tasks. Across subjects, the activity that required the lowest net metabolic energy expenditure was level ground walking at 0.6 m/s; the mean ground truth energy cost for this activity was 3.33 (SD 0.48) W/kg. The activity that required the highest net metabolic energy expenditure was running at 2.7 m/s; the mean ground truth energy cost for this activity was 13.33 (SD 1.64) W/kg. Therefore, on average, the range of ground truth energy cost experienced by subjects across tasks was 10 W/kg.

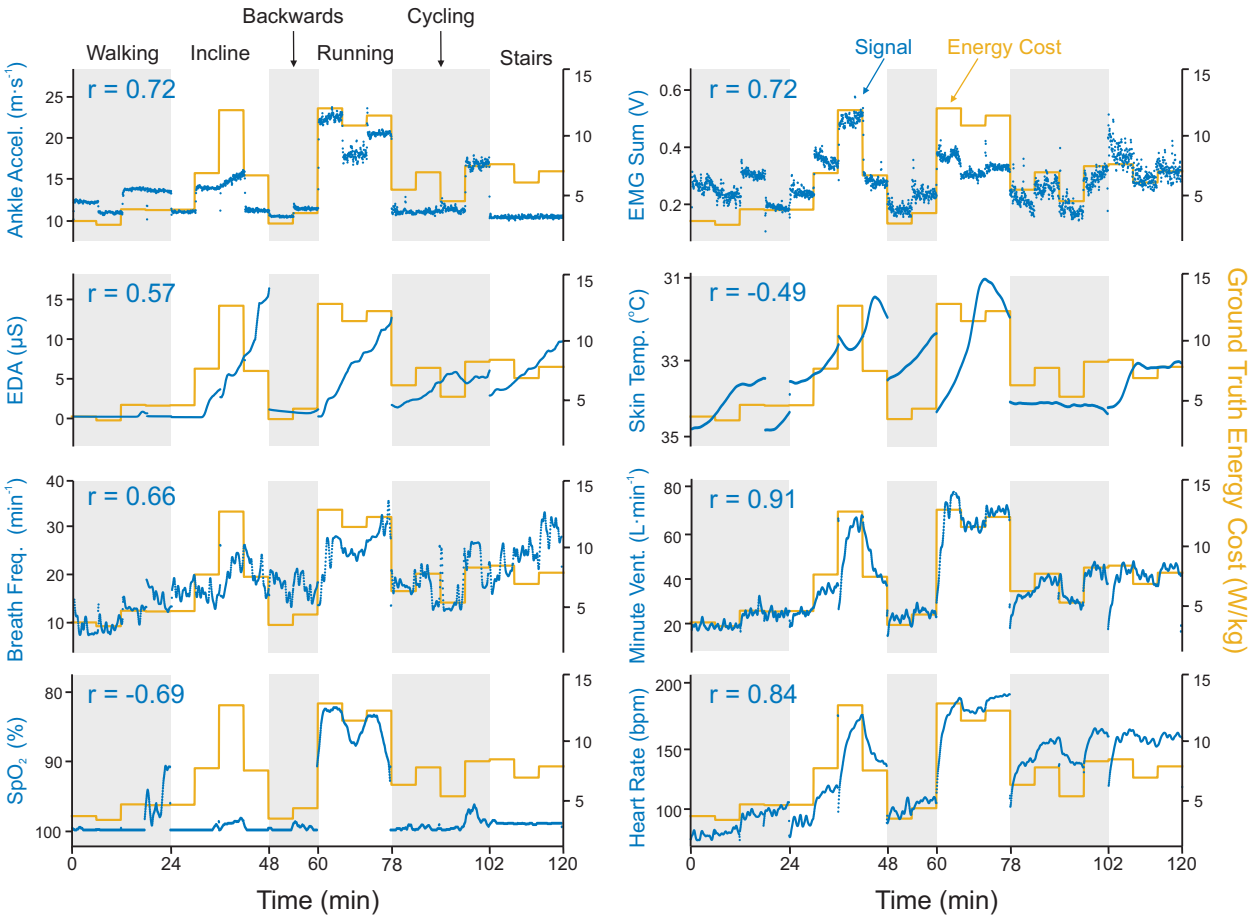

Figure 3.2: Time-series plots for filtered signals. Time-series plots for 8 filtered signals from one representative subject (Subject 4) for all 6 physical activities. Depicted ankle acceleration, EMG sum, electrodermal activity (EDA), and skin temperature are taken from the left side of the body. Left-hand axes measure the physiological signal (shown in blue); right-hand axes measure ground truth energy cost (shown in yellow). Pearson's correlation coefficient (r) between the signal and ground truth energy cost is written at the top left of each graph. For those signals with negative correlations (SpO_2 and skin temperature), the left-hand axis is reversed (*i.e.,* smaller values are at the top) to allow for visual overlay.

Effect of Filtering on Correlation with Ground Truth Energy Cost

Filtering dramatically improved the correlation of vector magnitude accelerations and linear envelope EMG signals, which are highly periodic with the gait cycle (Fig. 3.3). The filtered composite sum EMG signals had higher correlations than individual muscle signals, except for bilateral biceps femoris (Table 3.4). Filtering only slightly improved the average correlation of electrodermal activity, skin temperature, and heart rate, but improved correlation more substantially for breath frequency, minute ventilation, and oxygen saturation (Fig. 3.3). Unfiltered/filtered

$\dot{V}O_2$ had higher mean correlations than any of the other individual unfiltered/filtered signals tested. As aforementioned, we expected the correlation between ground truth energy cost and $\dot{V}O_2$ to be very high. Therefore, we use this correlation as a 'benchmark' to compare the correlation of other signals to.

	Unfiltered Mean (SEM)		*Filtered* Mean (SEM)	
	Left	Right	Left	Right
Vector Magnitude Accelerations				
Chest	.11 (.03)		.70 (.12)	
Waist	.15 (.04)		.68 (.13)	
Ankle	.30 (.07)	.30 (.07)	.69 (.14)	.68 (.14)
Wrist	.22 (.09)	.23 (.10)	.68 (.13)	.63 (.19)
Linear Envelope EMG				
Gluteus maximus	.20 (.03)	.10 (.04)	.47 (.07)	.23 (.09)
Rectus femoris	.12 (.03)	.12 (.03)	.33 (.09)	.29 (.06)
Vastus lateralis	.11 (.03)	.12 (.03)	.41 (.11)	.41 (.10)
Semitendinosis	.20 (.03)	.18 (.04)	.55 (.06)	.46 (.09)
Biceps femoris	.25 (.03)	.27 (.02)	.66 (.06)	.67 (.05)
Medial gastrocnemius	.09 (.03)	.12 (.02)	.33 (.09)	.44 (.07)
Soleus	.15 (.02)	.11 (.03)	.47 (.06)	.34 (.08)
Tibialis anterior	.08 (.04)	.06 (.03)	.19 (.10)	.17 (.08)
Composite sum	.27 (.03)	.25 (.03)	.61 (.07)	.51 (.07)
Respiratory and Physiological Signals				
$\dot{V}O_2$	.88 (.01)		.95 (.00)	
Breath Frequency	.52 (.04)		.68 (.04)	
Minute Ventilation	.86 (.01)		.91 (.01)	
SpO_2	-.36 (.10)		-.39 (.10)	
Heart Rate	.86 (.02)		.86 (.02)	
Electrodermal Activity	.42 (.04)	.42 (.06)	.42 (.04)	.43 (.06)
Skin Temperature	-.16 (.08)	-.20 (.08)	-.16 (.08)	-.20 (.08)

Mean (SEM) of Pearson's correlation coefficient was calculated across subjects (N=10). For signals collected bilaterally, results from left and right sides are presented in left and right columns, respectively.

Table 3.4: Effect of filtering on predictive capability of signals.

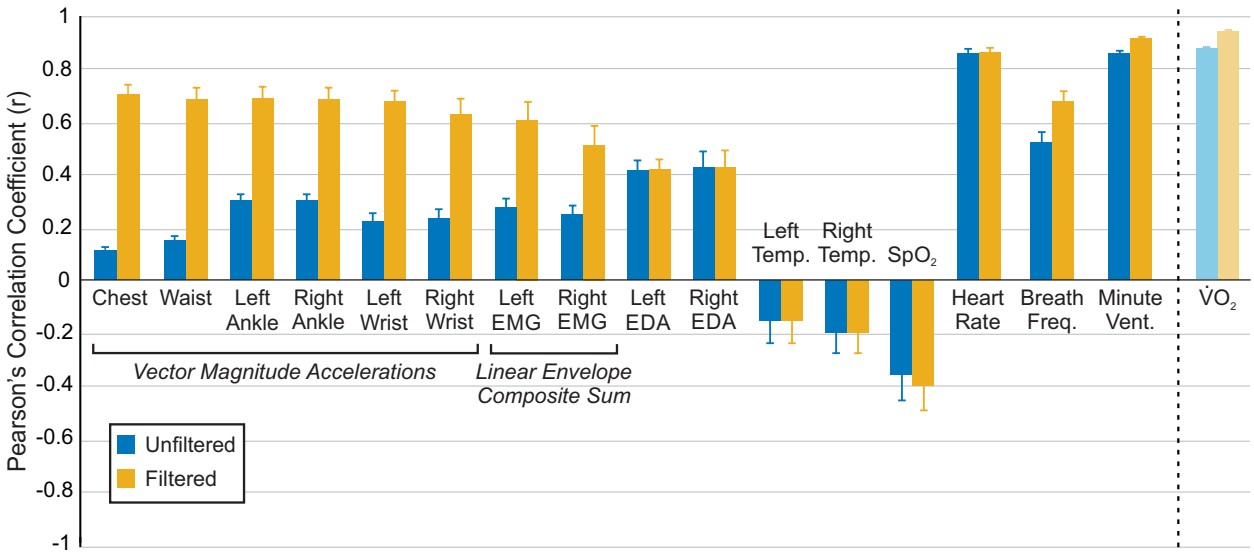

Figure 3.3: Pearson's correlation coefficients between unfiltered/filtered signals and ground truth energy cost. Correlations are calculated for individual subjects, with data pooled across all activities. Correlations are then averaged across subjects (N=10), and error bars represent ± 1 standard error of the mean (SEM).

Effect of Task and Subject Specificity on Prediction of Ground Truth Energy Cost

For the 17 cross-validated single linear regression models, evaluated using all six tasks, leave-one-*task*-out cross validation resulted in higher average root mean squared error (RMSE) than leave-one-*subject*-out cross validation, and we observed considerable variability between folds (Fig. 3.4A). For leave-one-*task*-out cross validation, the signal with the highest RMSE (*i.e.,* worst performance) was the chest accelerometer (4.56 (SD 5.76) $^{W}/_{kg}$); for leave-one-*subject*-out cross validation, the signal with the highest RMSE was left electrodermal activity (2.93 (SD 0.69) $^{W}/_{kg}$). minute ventilation had the lowest RMSE (*i.e.,* best performance) of any signals tested, for both leave-one-*task*-out (1.22 (SD 0.49) $^{W}/_{kg}$) and leave-one-*subject*-out (1.24 (SD 0.14) $^{W}/_{kg}$) evaluation methods (Fig. 3.4A). For comparison, the RMSE of filtered $\dot{V}_{O_2}$ was 0.88 (SD 0.42) $^{W}/_{kg}$ for *task*-left-out and 0.93 (SD 0.16) $^{W}/_{kg}$ for *subject*-left-out cross validation (Fig. 3.4A). We examined the results of each fold of leave-one-*task*-out cross validation individually; the mean RMSE (across all signals) was 1.5–3 times higher when running was left out of the cross validation than any other task (Table 3.5).

Left Out Task	Walking	Incline	Backwards	Running	Cycling	Stairs
RMSE, Mean (SD) (W/kg)	2.24 (0.81)	2.97 (0.75)	1.89 (0.53)	4.45 (1.40)	1.67 (0.94)	1.92 (0.60)

Table 3.5: Root mean squared error (RMSE) from individual folds of 6-fold leave-one-*task*-out cross validation.

Given these findings, we also performed a 5-fold leave-one-*task*-out cross validation, including level walking, incline walking, backwards walking, cycling, and stair climbing (*i.e.,* excluding running) (Fig. 3.4B). Across all signals, excluding running from the cross validation (Fig. 3.4B) resulted in a 31.1 (SD 9.6) % reduction in RMSE compared to the 6-fold leave-one-*task*-out cross validation (Fig. 3.4A). The signals with the largest differences were chest acceleration and left and right EMG composite sums; the signal with the smallest difference was waist acceleration.

Effect of Signal Groups on Prediction of Ground Truth Energy Cost

Regression groups that included the local signals (Local, Local+Global) performed worse when using leave-one-*task*-out cross validation than leave-one-*subject*-out cross validation (Fig. 3.5); regression groups which included mostly global signals (Global, Hexoskin) did not exhibit such a discrepancy in performance between evaluation methods. For leave-one-*task*-out cross validation, both the Global (1.23 (SD 0.54) W/kg) and Hexoskin (1.28 (SD 0.56) W/kg) models predicted energy cost almost as well as simply using filtered and scaled $\dot{V}O_2$ measurements (0.88 (SD 0.42) W/kg). For leave-one-*subject*-out cross validation, the Global (1.25 (SD 0.34) W/kg), Local+Global (1.28 (SD 0.34) W/kg), and Hexoskin (1.24 (SD 0.25) W/kg) groups were within 0.4 W/kg of filtered and scaled $\dot{V}O_2$ (0.93 (SD 0.16) W/kg).

Effect of Using Best Possible Models on Prediction of Ground Truth Energy Cost

As the number of predictors increased, so did the performance of the regression models, but with diminishing return. The RMSE and signals included in the best-performing regression models (trained and tested on data pooled across all subjects and tasks) with $k = [1,\ 2,\ ...\ 8]$ predictors are shown in Fig. 3.6. This sequence of best predictor combinations exhibited an incremental structure:

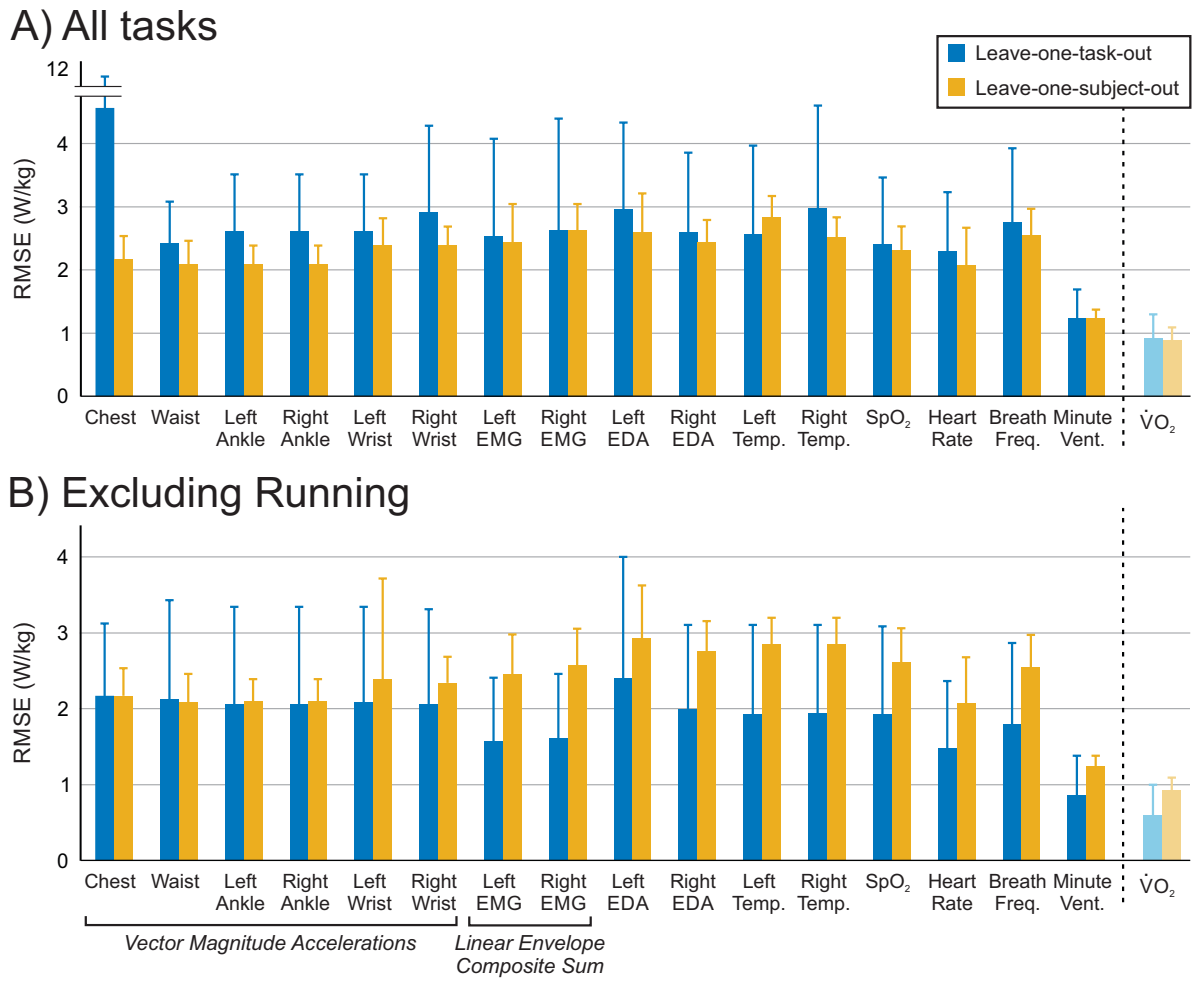

Figure 3.4: Average root mean squared error (RMSE) across folds for each signal using leave-one-task-out or leave-one-subject-out cross validation. (A) Leave-one-*task*-out cross validation is carried out including all tasks (6-fold). **(B)** Leave-one-*task*-out cross validation is carried out excluding running (5-fold). Leave-one-*subject*-out cross validation (10-fold) is the same in both A and B. Error bars represent ± 1 standard deviation.

the best combination of k predictors simply amended the best combination of $k-1$ predictors with an additional predictor. For example, for $k = 2$, we found minute ventilation and EMG composite sum (left) to be the best predictors, and for $k = 3$, minute ventilation, EMG composite sum (left), and waist acceleration. This structure was not enforced as we tested every possible combination of predictors independently for each k. In order, the signals found were minute ventilation, EMG composite sum (left), waist acceleration, electrodermal activity (right), breath frequency, heart rate, wrist acceleration (left), and ankle acceleration (right). With $k = 1$ predictor, the RMSE was 1.24 $^{\text{W}}/_{\text{kg}}$; With $k = 8$ predictors, the RMSE improved to 1.03 $^{\text{W}}/_{\text{kg}}$.

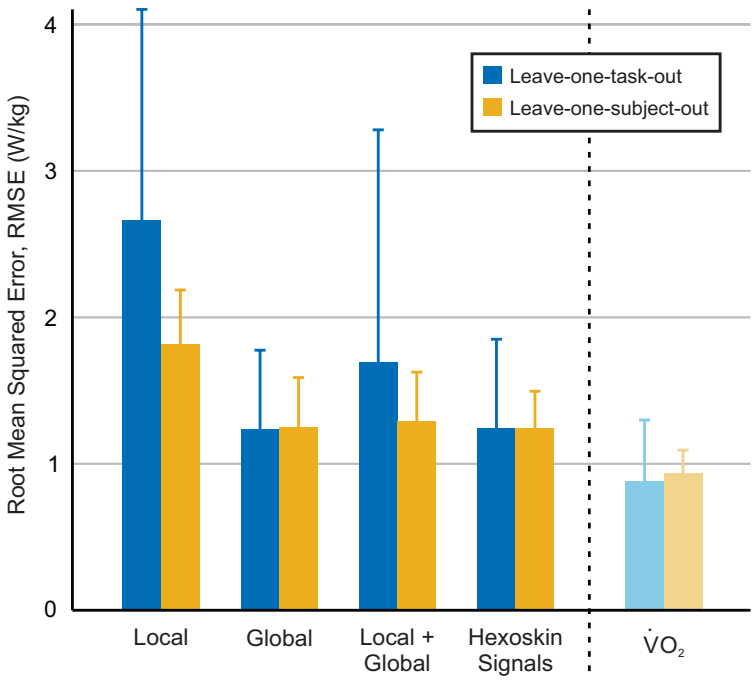

Figure 3.5: Average root mean squared error (RMSE) for each signal group using leave-one-task-out or leave-one-subject-out cross validation. Error bars represent ± 1 standard deviation. Signals included in each signal group are provided in Table 3.2.

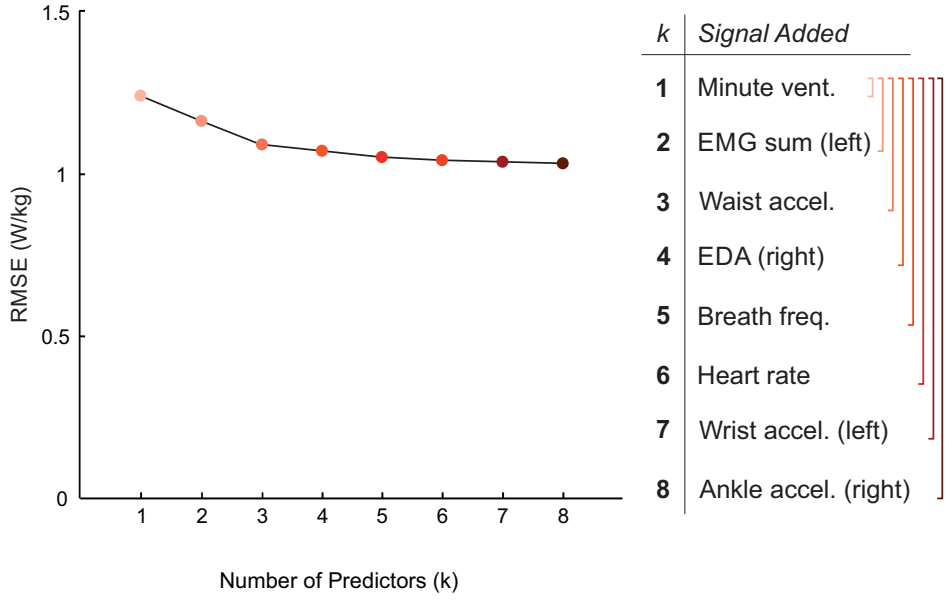

Figure 3.6: Root mean squared error (RMSE) for the best performing linear regression models with k predictors. The best model for each $k = [1, 2, ...8]$ was considered the model with the lowest RMSE out of all possible models containing k predictors. Adding additional signals improves predictive performance, but with diminishing return.

3.4 Discussion

In this study, we systematically evaluated the predictive capability of multiple physiological signals, alone and in combination with others, to predict metabolic energy cost. We focused on simple signal processing techniques and prediction algorithms to investigate the 'baseline' performance for each signal. The study is intended to inform the design and development of algorithms to estimate energy cost from physiological data acquired from wearable sensors in order to improve the speed, accuracy, and portability of body-in-the-loop optimization algorithms. These findings, as well as the large accompanying dataset [78], also have broad applications for any researchers or clinicians interested in predicting energy cost using portable sensors. These applications include, but are not limited to, the fields of exercise science, athletic training and performance, pediatric medicine, geriatric medicine, personal fitness, and rehabilitation. Using portable sensors to predict energy cost not only eliminates the need to use the bulky and cumbersome respiratory mask, but also enables researchers and clinicians to obtain estimates of energy cost outside the laboratory environment, for long or short periods of time, and from people across ages and ability levels.

We found that various combinations of physiological signals can predict steady-state energy cost quite well, even for unknown subjects and unknown activities. To illustrate this point, the overall predictive error obtained using the Global or Hexoskin signals groups represented only $\sim$1.25% of the *total range* of energy cost experienced across the tasks (which was approximately 10 $\mathrm{W/kg}$), and the error obtained using $\dot{V}O_2$ as a predictor was $\sim$0.9%. Therefore, the error obtained using the Global and Hexoskin signal groups comprised only a small fraction of the natural variation in energy cost across tasks, and was very close to the predictive power of $\dot{V}O_2$ itself. It is very difficult to compare predictive performance across studies due to the large number of confounding variables (*e.g.,* subjects, tasks, sensors, algorithms, units of metabolic power used), but our predictive performance falls into the range of previously reported performance using linear regression algorithms across a variety of tasks. For example, Altini *et al.* [3] reported an RMSE of 1.59 $\mathrm{kcal/min}$ ($\sim$1.5 $\mathrm{W/kg}$ for a 70-kg subject) across a set of tasks similar to those tested in our study. These findings demonstrated that without any advanced signal processing methods or machine learning

algorithms, it was possible to predict energy cost using portable, wearable sensors, without the use of the cumbersome respiratory mask.

In order to not bias our findings, we stayed as close as possible to the original sensor signals. Still, some basic data processing was required to improve the predictive capability of certain types of sensors. Filtering, for example, dramatically improved the correlations of ground truth energy cost with the accelerometers and the linear envelope EMG signals (Fig. 3.3). These filters effectively removed any periodic fluctuations associated with a gait cycle or other repetitive pattern, and generated a broader estimate of motion intensity. For signals that did not show such periodic fluctuations, including heart rate, electrodermal activity, or skin temperature, filtering did not improve the correlation. Yet, in a real-time application, it may still be beneficial to filter these noisy physiological signals to yield smoother estimates.

We also found a benefit to calculating the total EMG activity for one leg (*i.e.,* composite sum), as opposed to using individual muscle signals to predict energy cost (Table 3.4). Intuitively, different tasks require different muscle activation patterns, and therefore it is unlikely that one individual muscle would be able to capture the energetic requirements accurately across multiple tasks. Moreover, the composite sum EMG signals are more robust for real-time applications. EMG from individual muscles are highly sensitive to sensor placement, and can be affected by donning/doffing the sensors, sweat, and general sensor placement shifts. This effect was observed in this data set, considering the difference in mean correlations between filtered left and right gluteus maximus activity and ground truth energy cost (r=0.47 and r=0.23, respectively) (Table 3.4). Since the tasks tested in this experiment were all symmetrical tasks, these discrepancies were most likely due to differences in sensor placement between the left and right sides.

After filtering, the signals that correlated best with ground truth energy cost were minute ventilation, heart rate, and the accelerometers; the signal that was least correlated with ground truth energy cost was skin temperature. These results agree with a number of previous studies that have demonstrated good predictive ability with accelerations and heart rate (*e.g.,* [143]).

In terms of signal salience for body-in-the-loop optimization algorithms, these correlations can

only serve as a starting point. They were calculated on a subject-specific basis including data from all tasks; that is, they can explain how well a certain signal can predict the energy cost of a known subject doing a known task. Wearing adaptive assistive robotic devices, such as prostheses and exoskeletons, however, might very well constitute an unknown task and might lead to changes in an individual subject's physiology. This is particularly likely for body-in-the-loop optimization algorithms, which have the declared goal of continuously adapting and optimizing the behavior of assistive robotic devices. This adaptation leads to an ever changing 'task' for the user which will, in turn, influence the user's biomechanics (*e.g.,* gait kinematics, muscle activity) [20, 119]. Since the device is designed to assist with the motion, the changes in the task might be quite subtle. An exoskeleton, for example, might adapt to provide more torque during an activity without changing the kinematics of the motion. The user would adapt [91, 152], for example, by reducing muscle activation. Yet, from an outside perspective, it will appear as if the task remained the same while the physiological response of the subject has changed. Thus, for body-in-the-loop optimization algorithms to be successful, we need to incorporate signals that are robust to potentially unknown tasks and to unknown subjects. To identify these signals, we tested a broad range of different activities, and conducted leave-one-task/subject-out cross validations to explicitly test the predictive capability of the different signals for unknown tasks and subjects.

When all six activities were included in the analysis, every signal tested in this experiment had higher root mean squared error (RMSE) when evaluated with leave-one-*task*-out cross validation than with leave-one-*subject*-out cross validation (Fig. 3.4A). Broadly, this implied that all the signals evaluated in this study were more sensitive to unknown tasks than unknown subjects. The standard deviation between folds of the cross validation was very high, particularly for leave-one-*task*-out cross validation. Examining these results on a fold-by-fold-basis, we noticed that when the regression model was trained on walking, incline walking, backwards walking, cycling, and stair-climbing data and tested on running data, the resultant RMSE was 1.5–3 times higher than any other fold (Table 3.5). This indicated that the signal data collected during running were drastically different than the signal data collected during the other tasks. Since the running task

dominated the error of the leave-one-*task*-out evaluation, we believed it was masking some of the trends, particularly between different types of sensors.

As such, we performed a secondary analysis in which all running data were removed from the data set (Fig. 3.4B), and we performed a 5-fold leave-one-*task*-out cross validation with the remaining tasks. In this evaluation, we saw a dramatic reduction in RMSE compared to the 6-fold leave-one-*task*-out cross validation. Moreover, valuable insights emerged regarding how different types of signals were affected by unknown tasks and unknown subjects. As a broad signal class, the accelerometers were most affected by unknown tasks, with high variability between folds of the validation. We believe this is likely because the accelerometers provide information about the kinematics and the motion pattern of an activity, and therefore may be more sensitive to new patterns of activity. On the contrary, EMG and the global signals were more sensitive to an unknown subject than an unknown task. For EMG, this is potentially a result of differences in electrode placement between subjects. For the global sensors, this might be a result of the fact that they measure quantities related to the user's physiological responses, and may be affected by subject-specific factors like cardiovascular health [174] or individual differences in sweat and body temperature responses [139, 149]. To mitigate some of these factors, previous studies have proposed the personalization of energy expenditure estimates using normalization factors for physiological quantities, such as heart rate or electrodermal activity [2,4], which could be useful for our future implementations. For our study, we defined EMG activity as local signal, since it provides information about a specific muscle. However, shown in Fig. 3.4B, the composite sums of EMG signals are more robust to an unknown task than an unknown subject. Despite the fact that EMG activity is related to the kinematics of a task, individual subject differences may play a larger role in the generalizability of EMG signals for predicting energy cost. The difference in performance between the leave-one-*task*-out cross validations when the running task was included/excluded also highlights that a single algorithm may not be able to appropriately capture the energy cost requirements across *all* tasks. In this case, more sophisticated task-classification algorithms may be necessary to appropriately identify the task and apply the correct prediction algorithm based on

that task, a method which has been proposed by previous groups (*e.g.,* [4]).

In general, combining the signals into groups improved their predictive capability. As with the individual sensors, the groups that contained local signals (Local, Local+Global) performed worse when predicting an unknown task than an unknown subject. The Global and Hexoskin groups were the best performing groups for both leave-one-*subject*-out and leave-one-*task*-out cross validations, and both were able to predict energy cost almost as well as $\dot{V}O_2$ (within 0.4 W/kg). Although both groups performed similarly, the Hexoskin group (including the waist accelerometer, heart rate, breath frequency, and minute ventilation) may be a more robust predictor of energy expenditure when the user is wearing an assistive device because it combines information from both local and global sensors. The waist accelerometer has the potential to be a good predictor of energy expenditure across multiple tasks because it is located close to the body's center of mass (COM), and is therefore correlated with whole body energetic cost during ambulatory activities [4, 32, 33]. Other accelerometers on the ankles and wrists may be more affected by specific kinematic patterns, and less robust to unknown tasks, which could explain the larger error in the Local+Global group.

According to the results of this study, with only one signal (minute ventilation), we could achieve an RMSE of 1.24 W/kg. Adding additional predictors improved performance (up to RMSE = 1.03 W/kg for $k = 8$), but with diminishing return. These errors correspond to approximately 1–1.2% of the total range of energy cost measured in this experiment (10 W/kg); this error is quite small compared to the possible range of energy cost experienced during different activities. In particular, the signals able to be collected by the Hexoskin smart shirt were among the 6 best predictors of energy cost, which suggested that even with very simple signal processing and regression algorithms, a fully portable device could be used to accurately predict energy cost across a wide variety of tasks. Success using the Hexoskin smart shirt and a random forest regression algorithm to predict oxygen uptake *kinetics* has recently been demonstrated by Beltrame *et al.* [13] during activities of daily living and a pseudo-random walking sequence. In our study, the Hexoskin signals showed similar performance when predicting *steady-state* energy cost. Analysis of the best performing models demonstrated that adding more predictors improves performance up to a point, but with

the right choice of signals, good predictive performance can be achieved using fewer sensors. It would be up to an individual experimenter to decide the acceptable trade-off between regression performance and number and/or simplicity of sensors.

In this context, it is important to note that the differences between the top performing models that we identified were rather small. In particular, swapping out one signal for a related one (such as swapping left and right wrist accelerations or left and right EMG) resulted in similar performance. For example, using the right EMG instead of the left EMG increased the RSME only by $0.07\,\mathrm{W/kg}$ ($<1\,\%$ of the total range of metabolic cost) for $k = 2$. At the same time, it is interesting to note that none of the best performing models included bilateral signals (*i.e.,* both left EMG *and* right EMG). These analyses suggested that a more robust prediction can be made by combining information from multiple sensor modalities (*e.g.,* combine EMG and electrodermal activity), not simply adding additional examples of the same highly-correlated signals (*e.g.,* left EMG and right EMG). Along the same lines, we observed a mix of local and global sensors when we determined the best possible combinations of k signals. Furthermore, the order of signals added to the best combinations (with increasing k) was neither determined by the signals individual Pearson's correlation coefficients nor by their indiviudal RMSE values. This further highlights that it pays to have different types of sensor modalities when combining multiple signals, rather than just using the strongest individual sensors.

The goal of this study was to systematically evaluate the predictive power of physiological signals and groups of signals in order to draw conclusions about which types of signals are useful for our application of body-in-the-loop optimization. To keep our results general, we did so for multiple subjects and using a very broad range of activities. Yet, our choice of activities is not without drawbacks; we chose a broad set of activities with very different kinematics and dynamics, and some of which were novel for subjects (*e.g.,* backwards walking). Subjects' habituation to these activities may have changed their kinematics, kinetics, and energy consumption, and the current analysis did not account for these changes. Our current study is also limited by its small sample size and homogeneous, able-bodied subject population. One could argue that an algorithm

specifically designed to estimate the cost of using a particular assistive device should be more hand-tailored to that specific device and based on data that are specific to this device. That is, subsequent studies should investigate how well the predictive algorithms can work with data from individuals using assistive devices. However, such an approach might overly restrict the available data and consequently bias the estimation process. This would limit the applicability of body-in-the-loop optimization algorithms, which should ideally be unbiased in their search for the optimal device setting.

It is important to note that our study was not an attempt to find the best possible way to predict steady-state metabolic effort, and the models discussed in this study are not intended to be complete predictive models. In reality, many of the signals discussed could be cross-correlated, and more rigorous statistical methodology would be required to build an optimal model with the fewest number of predictors. For the current study, we prioritized simplicity over optimal algorithmic performance, and we used linear regression as a common algorithm to compare the predictive power across signals. In the future, other machine learning algorithms (such as random forest or neural networks) that may produce higher prediction accuracy and be more robust to unknown subjects and tasks should be explored. Additionally, more advanced feature extraction and signal processing techniques (*e.g.,* different filters or normalization) may improve predictive capability of various signals. To facilitate the collaborative development of such advanced techniques and algorithms, we decided to publish our complete dataset together with this paper [78]. Our hope is that it can be used to compare different approaches that predict energy cost from portable sensors for a wide variety of applications. Although the sensors we used in this study were not fully portable (*e.g.,* EMG, minute ventilation/breath frequency), fully portable versions of these sensors are commercially available, such as the BioStampRC® [115] and the Hexoskin® smart shirt [13,25, 175]. Future work should involve testing these algorithms in real time using fully portable sensors.

From our analyses of a broad set of physiological signals, we concluded that 1) filtering the accelerations and EMG linear envelopes to remove the periodicity from the gait cycle improved their ability to predict ground truth energy cost; 2) when the physical activities were diverse (*i.e.,*

including running), all the signals were more sensitive to unknown tasks than unknown subjects; 3) when the tasks were more similar (*i.e.,* excluding running), physiological signals (*e.g.,* heart rate or electrodermal activity) were more sensitive to unknown subjects than unknown tasks; 4) groups of signals could improve predictive capability over individual signals, but the errors were similarly affected by task and subject specificity; and 5) good predictive performance could be achieved using a small number of sensors. Given these conclusions, for body-in-the-loop optimization and other applications which require predicting energy cost from portable sensors, we recommend combining accelerometers (ideally located close to the center of mass) with heart rate, breath frequency, and/or minute ventilation to provide the most robust prediction of energy expenditure.

CHAPTER 4

Accelerating the Estimation of Metabolic Cost using Signal Derivatives: Implications for Optimization and Evaluation of Wearable Robots

4.1 Background

A chief goal for lower-limb wearable robots (*e.g.,* exoskeletons) is to augment the user's natural motion in a helpful, intuitive way. Accordingly, a common evaluation metric for such systems is the effect of the robotic assistance on the metabolic cost of the wearer. Recent hardware and control advancements have enabled researchers to achieve the challenging task of reducing the metabolic cost of walking below that of walking without an exoskeleton, using autonomous [119–121], tethered [111], and even passive [29] devices. While energy expenditure has long been used as an evaluation metric for device efficacy, it has also been proposed as a design specification (*i.e.,* the Augmentation Factor [121]), and, recently, used as a physiological cost function for the real-time control of wearable robotic devices—a strategy known as body-in-the-loop or human-in-the-loop optimization. These algorithms use real-time estimates of metabolic cost to iteratively tune the actuation profile of an assistive device to minimize the wearer's energetic cost [35, 44, 90, 92, 189].

The implementation of body-in-the-loop optimization algorithms, illustrated in Fig. 4.1, requires that the user wear a robotic assistive device (*e.g.,* a prosthesis or exoskeleton) which is

This chapter has been previously published in full in the *IEEE Robotics & Automation Magazine* [81].

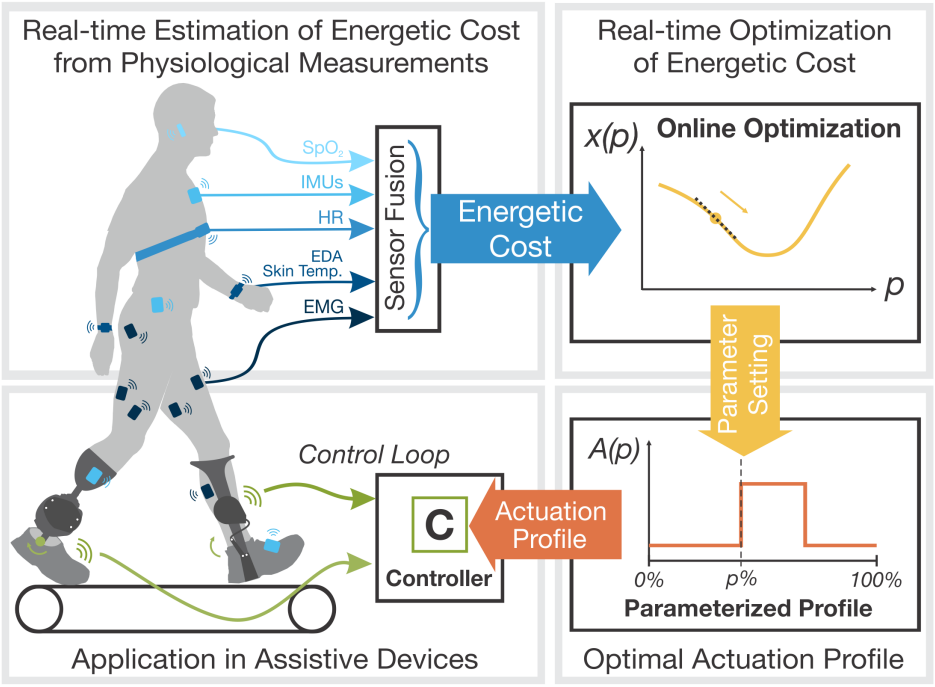

Figure 4.1: Schematic of body-in-the-loop optimization with estimation of metabolic cost from wearable sensors. Body-in-the-loop optimization is the process of iteratively and automatically tuning the actuation profile of wearable robotic devices (*e.g.,* prostheses and exoskeletons) to minimize the metabolic cost of the wearer. In the example depicted here, the individual wears robotic devices (bottom left), which are programmed to provide assistance. The parameterized actuation profile in this illustration ($A(p)$, bottom right) is a square pulse with an onset at $p\%$ of the gait cycle. This actuation profile is enforced by the closed-loop low-level device controller, with feedback from the robotic device (*Control Loop*, bottom left). While the user walks with the current actuation profile, their instantaneous energetic cost is measured or estimated (top left). The metabolic cost associated with the current parameter setting ($x(p)$, top right), is used in an online optimization algorithm, which chooses the next parameter setting to test. This loop continues until the optimizer converges to the optimal actuation profile to minimize the metabolic cost of the user.

programmed to provide assistance. The actuation profile of the device, $A(p)$, is shaped by one or more parameters p; $A(p)$ can be any parameterized assistance trajectory, but is commonly defined as joint torque or power as a function of time or gait phase (%). Once defined, the actuation profile is enforced by the device's low-level control loop, which often incorporates device feedback in a closed-loop controller. While the user walks with the current actuation profile, their metabolic cost is measured or estimated. Given the metabolic cost associated with the current parameter setting, $x(p)$, an online optimization algorithm (*e.g.,* gradient descent [93], CMA-ES [189], or Bayesian [35]) chooses the next parameter setting to test. This process continues, eventually converging on the optimal actuation profile that minimizes the cost of the wearer. Although body-in-

the-loop optimization experiments have shown promising results for real-time control of wearable robots, these algorithms have not yet gained adoption outside the laboratory environment.

One major factor delaying the widespread use of body-in-the-loop methodology is obtaining the required estimates of metabolic cost quickly, accurately, and with unobtrusive equipment. Consider the scenario in which an individual wears a robotic ankle exoskeleton that supplements 20 % of the biological ankle's work, and the device instantaneously changes to providing 40 % of biological ankle's work. This step change in device parameter setting results in a physiological step change in energetic demand—the body instantaneously requires a different amount of energy to accommodate the new conditions. The instantaneous energetic demand associated with a new set of experimental conditions is called the *instantaneous metabolic cost*. This quantity is utilized in body-in-the-loop optimization algorithms to automatically tune device parameters in real time to accommodate different users, different ambulation tasks, or changing environmental conditions. Measurements of instantaneous metabolic cost are also widely utilized across many fields for assessing outcomes of experiments, rehabilitation protocols, or athletic training, to name only a few.

Obtaining estimates of instantaneous metabolic cost is difficult because the measurements of metabolic cost collected from state-of-the-art systems are significantly delayed from the instantaneous metabolic demands of the body. To date, indirect calorimetry has been the method of choice for body-in-the-loop optimization and wearable robotics evaluations because portable measurement units (*e.g.,* Cosmed or Oxycon Mobile) allow the subjects to move freely about the laboratory environment while the system measures their respiratory gas exchange. Yet, the measurements obtained through indirect calorimetry exhibit high measurement noise, a low sampling rate, and require that the user wears a bulky rubber mask that covers his or her nose and mouth; these challenges make indirect calorimetry unsuitable for real-world applications. As an alternative, many researchers in the fields of physiology, exercise science, and clinical nutrition have focused on obtaining estimates of energy cost using portable wearable sensors. Most commonly, sensor signals are used as inputs to a black-box algorithm (*e.g.,* multiple linear regression, neural network, or random forest) which is then used to predict instantaneous metabolic cost [77], the metabolic

equivalent of a task (MET) (*e.g.,* [19, 32, 33]), or the dynamic response of energy expenditure or oxygen uptake (*e.g.,* [3, 13, 143]). While it is possible to obtain reliable estimates of metabolic cost using unobtrusive portable sensors and black-box machine learning algorithms, the *speed* of this prediction is still limited by the dynamics of the physiological input signals—a drawback that these signals share with indirect calorimetry.

The speed of estimating instantaneous metabolic cost is critical for the implementation of body-in-the-loop optimization outside the laboratory. During typical activities, it is rare that individuals perform long bouts of constant-intensity activity, and the device must quickly adapt actuation profiles accordingly, without waiting multiple minutes for each assessment. To understand the impact of signal dynamics on the speed of prediction, let us consider respiratory gas measurements, as well as measurements of minute ventilation ($\mathrm{L/min}$) and heart rate ($\mathrm{min^{-1}}$)[1]. Individually, the dynamics of each of these signals have been characterized as a linear first order system, with time constants of 40–42 s for respiratory gas measurements [154], 53–54 s for minute ventilation [99, 181], and 58 s for heart rate [73]. That means for a given step change in intensity level (*e.g.,* changing from walking at 1 m/s to 1.4 m/s) the individual must perform 2–3 minutes of constant-intensity activity before these signals reach steady state, and, depending on measurement noise, an additional 2–3 minutes are required to accurately estimate the steady state value. So while each of these signals is highly correlated with energy expenditure, their dynamics limit how quickly we can obtain estimates of instantaneous metabolic cost. For indirect calorimetry, model-based approaches [93, 154] have mitigated this issue, but in order to improve the real-time evaluation and control of wearable robots outside of the lab, we need to apply similar techniques to obtain faster predictions of instantaneous metabolic cost from portable, wearable sensors.

To improve the speed of instantaneous metabolic cost predictions using wearable sensors, we propose including the time derivatives of physiological signals in a black-box estimation process. We expect that this information is useful because a linear regression performed on a signal and its derivative has the same mathematical structure as an explicit first order dynamic model. Yet in

[1] In a previous study, we found that these physiological signals contained the most salient information for predicting energy cost [77], which agreed with many previous findings [3, 13]

contrast to such an explicit model, a data-driven approach takes into account the relative quality and information content of the signal and its derivative. In this context, one would expect that the various signals included in these estimators have different dynamic properties [77]. A signal such as limb acceleration provides *local* information about one segment and changes instantaneously with a step change in activity, yet it is a less direct measurement of total physiological energy expenditure. In contrast, signals such as heart rate or breath volume provide *global* information about the physiological state of the user—which is more closely related to physiological energy expenditure—but their response is dynamically delayed. For such dynamically delayed signals, we expect that the inclusion of derivative information can improve the overall prediction quality.

To test our theory, we evaluated the effect of including signals' derivatives on the predictive accuracy of multiple linear regressions trained to predict instantaneous metabolic cost. In this article, we show that including the derivatives of signals with known first order dynamics (*e.g.,* heart rate, minute ventilation) in the regressions improves the predictive accuracy during the transient phase of an activity (*i.e.,* before the signals have reached steady state) without deteriorating the prediction during the steady state phase. For certain signal groups, including derivatives in the regressions resulted in accurate estimates of instantaneous metabolic cost (here defined as error less than 1 $\mathrm{W/kg}$) more than *three times faster* than regressions without derivatives. The demonstrated error reductions and increased prediction speeds during the transient phase of an activity highlight the potential for this approach to improve the estimation of metabolic effort, and in particular, to advance the translation of body-in-the-loop optimization algorithms into real-world applications.

4.2 Methods

4.2.1 Derivation

By identifying explicit dynamic models of physiological responses, it has been shown that the transient behavior of many physiological signals, including energy cost measurements [154], minute ventilation [99, 181], and heart rate [73] can be described by a first order linear differential

equation. These linear input-output models take the form:

$$\dot{y}(t) = \frac{1}{\tau}\Big(x(t) - y(t)\Big) \tag{4.1}$$

where $x(t)$ is the signal value that will eventually be reached in *steady state* when the activity performed at time t is continued at the current level, and where $y(t)$ is the *transient* signal measurement at time t. The subject-specific time constant τ represents the speed with which a signal can respond to a step change in intensity. This dynamic representation might prove similarly useful when trying to approximate the transient behavior of other physiological signals or groups of signals.

Employing (4.1) to predict the steady state values $\hat{x}(t)$ of a physiological signal from its current transient measurements $y(t)$ just requires a simple rearrangement of terms:

$$\hat{x}(t) = y(t) + \tau\dot{y}(t). \tag{4.2}$$

Unfortunately, using this equation suffers from some severe practical limitations. Subject-specific time constants τ need to be determined, and the measurements $y(t)$ are typically quite noisy, which is further amplified in the process of obtaining the derivatives $\dot{y}(t)$. To minimize the error in the prediction of $\hat{x}(t)$, one has to filter the signal while carefully trading noise for signal attenuation and phase delay, especially if the activity and corresponding target value $x(t)$ are changing rapidly.

Here, we instead focus on the linear nature of (4.2) and restate the output as weighted sum of the measurement and its derivative $\hat{x}(t) = y(t) + \tau\dot{y}(t) = \begin{bmatrix} y(t) & \dot{y}(t) \end{bmatrix} \begin{bmatrix} 1 \\ \tau \end{bmatrix}$, in which the vector $\begin{bmatrix} 1 & \tau \end{bmatrix}^T$ can be thought of as a coefficient vector applied to two individual signals $y(t)$ and $\dot{y}(t)$. When using experimental data of n sampled pairs of ground truth steady state $x(t_i)$ and measurements $y(t_i)$, such a coefficient vector (β) can be obtained via linear regression[2], minimizing the predictive

[2]Please note that in standard regression notation, the dependent variable would be labeled y and the independent variable x. Here we have switched this convention to be consistent with the input-output notation of the related work.

error of the model:

$$\hat{x}(t) = \begin{bmatrix} y(t) & \dot{y}(t) \end{bmatrix} \beta, \tag{4.3}$$

with the coefficients:

$$\beta = \left(\mathbf{Y}^T\mathbf{Y}\right)^{-1}\mathbf{Y}^T\mathbf{x}, \text{ where } \mathbf{Y} = \begin{bmatrix} y(t_1) & \dot{y}(t_1) \\ \vdots & \vdots \\ y(t_i) & \dot{y}(t_i) \\ \vdots & \vdots \\ y(t_n) & \dot{y}(t_n) \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x(t_1) \\ \vdots \\ x(t_i) \\ \vdots \\ x(t_n) \end{bmatrix}. \tag{4.4}$$

In the presence of measurement noise, the identified coefficients of β will not match the coefficients $\begin{bmatrix} 1 & \tau \end{bmatrix}^T$ of the theoretical model. Depending on the signal-to-noise ratio (SNR) of the individual signals $y(t)$ and $\dot{y}(t)$, the regression will be biased towards one or the other. Typically the SNR of $\dot{y}(t)$ is much lower than that of $y(t)$, resulting in a $\dot{y}$ coefficient in β that is considerably smaller than the time constant τ. A more detailed derivation of this phenomenon can be found in the appendix.

With (4.3) and (4.4), we can predict the steady state values $x(t)$ of different physiological sensor signals without waiting until steady state is actually reached. This ability is particularly useful if we try to estimate the energetic cost $E(t)$ of a time-varying activity or try to rapidly optimize the parameters of an assistive device. In practice, one might want to fuse the steady state estimate of multiple sensor signals $\hat{x}_j$, $j \in [1 \cdots m]$ into an aggregated estimate of $\hat{E}(t) = f(t, \hat{x}_1(t), \cdots, \hat{x}_j(t), \cdots, \hat{x}_m(t))$; expressed here as the function f. A very basic version of this approach is to let f be a linear function of the individual estimates $\hat{x}_j$. In this case, one can actually skip the intermediate step of predicting an individual signal's steady state response with (4.3) and (4.4), and instead directly perform a linear regression using all the raw signals y_j and their derivatives $\dot{y}_j$ to predict steady state energy cost, $E(t)$. While this approach is likely not the most effective estimator, we chose it deliberately as a simple means to experimentally evaluate the concept of including derivative information in a data-driven approach.

4.2.2 Experiment

The details of this experimental protocol have been previously published in Ingraham *et al.* [77] and are described in Chapter 3. For brevity, we summarize the most relevant information here. Additional information, including the equipment used, sensor placement, signal filtering, and missing data, can be found in Section 3.2 of this dissertation.

Subjects

Ten healthy participants (8 male, 2 female, mean (SD) age: 27.4 (4.5) years, height: 1.78 (0.08) m, weight: 69.0 (9.9) kg), participated in our experiment, after providing informed consent to a protocol approved by the University of Michigan Institutional Review Board.

Protocol

Each subject performed six physical activities at a variety of speeds and/or intensities: level walking (0.6, 0.9, 1.2 m/s), incline walking (0.6, 1.2 m/s at 4° incline; 0.6, 1.2 m/s at 9° incline), backwards walking (0.4, 0.7 m/s), and running (1.8, 2.2, 2.7 m/s) on a treadmill, cycling on a stationary bike (Resistance 1, 3, 5 at 70 rpm; Resistance 1 at 100 rpm), and stair climbing on a stairmill (60, 75, 90 W). In total, subjects completed 20 conditions divided between two sessions (subjects performed level walking at 1.2 m/s during both session 1 and session 2). For each physical activity (*i.e.,* level walking), subjects stood quietly for 6 minutes, performed each speed/resistance condition for 6 minutes in a random order, and then sat quietly for 6 minutes. The changes between the six minute intervals thus constituted instantaneous step changes, from rest to activity, from one speed/resistance condition to another, and from activity to rest. Between different physical activities, subjects rested for at least 10 minutes.

Wearable Sensor Recordings

During data collection, we instrumented subjects with a suite of wearable sensors, listed in Table 4.1. We recorded oxygen consumption ($\dot{V}_{O_2}$), carbon dioxide production ($\dot{V}_{CO_2}$), breath

frequency, and minute ventilation breath-by-breath using a portable respirometer (Oxycon Mobile, CareFusion, San Diego, CA). Heart rate was measured using a wireless heart rate monitor strapped around the chest (Polar Electro, Inc., Bethpage, NY). Blood oxygen saturation (SpO_2) was measured using a pulse oximeter (Oxycon Mobile, CareFusion, San Diego, CA) affixed to the subject's left earlobe. Since the recording of heart rate and oxygen saturation was synchronized with the respirometer, these signals were also recorded breath-by-breath. Subjects wore 4 tri-axial accelerometers (Opal, APDM Inc., Portland, OR) at the center of the chest, on the left hip, and on the left and right lateral ankle; these accelerometers measured accelerations in the sensor x, y, and z axes at 128 Hz. Subjects also wore bilateral wristbands (E4, Empatica, Milan, Italy) containing tri-axial accelerometers that recorded x, y, and z accelerations of the left and right wrist at 32 Hz. The Empatica wristbands also contained sensors to measure electrodermal activity (EDA) and skin temperature at 4 Hz. Surface EMG electrodes (Biometrics Ltd., Ladysmith, VA) recorded muscle activity at 1kHz from 8 bilateral leg muscles: gluteus maximus, biceps femoris, semitendinosis, rectus femoris, vastus lateralis, medial gastrocnemius, soleus, and tibialis anterior. All sensor signals were time-synchronized during collection.

Signal Group	Signal(s)	Sampling Freq. (Hz)	Filter Cutoff Freq. (Hz)	Sensor	Sensor Location[§]
Local	Total acceleration (x4)	128	0.1	3-axis accelerometer (APDM)	Chest, hip, L/R ankle
	Total acceleration (x2)	32	0.1	3-axis accelerometer (Empatica)	L/R wrist
	Total EMG activity (x2)	1000	0.1	Surface EMG (Biometrics)	L/R legs
Global	Electrodermal activity (x2)	4	0.01	E4 wristband (Empatica)	L/R wrist
	Skin temperature (x2)	4	0.01	E4 wristband (Empatica)	L/R wrist
	Heart rate	$*$	0.01	Heart rate monitor (Polar)	Chest
	Breath frequency	$*$	0.01	Respirometer (Oxycon Mobile)	Mask
	Minute ventilation	$*$	0.01	Respirometer (Oxycon Mobile)	Mask
	Oxygen saturation, SpO_2	$*$	0.01	Respirometer (Oxycon Mobile)	L earlobe
Ground Truth	O_2 consumption, $\dot{V}_{O_2}$	$*$	-	Respirometer (Oxycon Mobile)	Mask
	CO_2 production, $\dot{V}_{CO_2}$	$*$	-	Respirometer (Oxycon Mobile)	Mask

*These signals were sampled one measurement per breath. Each signal was linearly interpolated to 10Hz before filtering.
[§]L = left, R = right

Table 4.1: Physiological signal collection and processing.

Signal Processing

For each accelerometer, we calculated the total acceleration by calculating the vector magnitude ($\sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2}$). We derived the linear envelope for all individual EMG signals, then calculated total EMG activity for each leg as the norm of all 8 individual muscle signals ($\sqrt{M_1(t)^2 + ... + M_8(t)^2}$). Each of the 16 signals was filtered using a 1-minute Gaussian filter kernel with appropriate cutoff frequencies derived from the literature (Table 4.1). Breath frequency, minute ventilation, heart rate, oxygen saturation, left/right electrodermal activity and left/right skin temperature were filtered using a cutoff frequency of 0.01 Hz; the 6 limb accelerations and left/right muscle activity were filtered using a cutoff frequency of 0.1 Hz. To compute each filtered signal derivative, the raw signal was convolved with the derivative of each Gaussian filter kernel. All signals were filtered and re-sampled to 1 Hz.

We divided the 16 filtered signals into two categories: the *local signals* and the *global signals* [77]. The 8 local signals (6 vector magnitude accelerations and the left/right total EMG) provide information about an individual muscle or limb segment and vary with the periodicity of the activity. In contrast, the 8 global signals (heart rate, left/right electrodermal activity, left/right skin temperature, breathing frequency, oxygen saturation, and minute ventilation) provide information about the state of the whole body.

Ground Truth Instantaneous Metabolic Cost

We used the Brockway equation [18] to calculate metabolic rate from unfiltered oxygen consumption ($\dot{V}_{O_2}$) and carbon dioxide production ($\dot{V}_{CO_2}$); we normalized this value to the subject's body mass. We assumed that the instantaneous metabolic cost was constant for a given condition, and we obtained this value by averaging the final 3 minutes of breath measurements for that condition (assuming that the measurements have reached steady state). The instantaneous metabolic cost forms a stair-step function that changes every time the activity level changes; this quantity represents the instantaneous metabolic demands of the body that correspond to a step-change in activity, and will henceforth be referred to as the 'ground truth' energy cost for a given condition.

In the regressions, ground truth energy cost served as our target data (the energy cost we aimed to predict). For each activity, the average energy cost from the standing bout at the beginning of the trial was subtracted off to yield net metabolic rate.

4.2.3 Linear Regression Analysis

In total, we trained 6 cross-validated multiple linear regression models, containing: 1) the 8 local signals, 2) the 8 global signals, 3) the 16 combined signals (*i.e.,* local + global), 4) the 8 local signals plus their 8 derivatives, 5) the 8 global signals plus their 8 derivatives, and 6) the 16 combined signals plus their 16 derivatives. Regression models 1–3 provided a baseline for comparison of model performance without derivatives (as done in [77]). Regression models 4–6 were designed to evaluate how the addition of derivatives can improve the prediction of instantaneous metabolic cost.

All 6 linear regression models have the form:

$$\hat{E}(t) = \mathbf{y}(t)\mathbf{b}, \tag{4.5}$$

where $\hat{E}(t)$ is the estimate of the steady state energy cost that will eventually be reached if the activity performed at time t were to be continued at the current intensity (*i.e.,* the instantaneous metabolic cost). The vector $\mathbf{y}(t)$ contains a constant offset term and either m sensor signals, $\mathbf{y}(t) = [1, y_1(t), \ldots y_m(t)] \in \mathbb{R}^{m+1}$ (regressions 1–3) or a constant offset term, m sensor signals, and m derivatives, $\mathbf{y}(t) = [1, y_1(t), \ldots y_m(t), \dot{y}_1(t), \ldots \dot{y}_m(t)] \in \mathbb{R}^{2m+1}$ (regressions 4–6). The vector $\mathbf{b}$ contains the $m + 1$ or $2m + 1$ regression coefficients.

For this analysis, we considered each condition an individual trial; although subjects performed each condition for 6 minutes, we excluded the last minute and only included the first 5 minutes in our regressions to eliminate any inconsistencies from switching between tasks. In total, the data set analyzed in this study comprised 187[3] 5-minute trials. We performed a 5-fold cross validation in which we randomly divided the 187 trials into a training set (80%, 149 trials), and a testing set

[3]Note that this number is smaller than 200, as not all 10 subjects successfully performed all 20 conditions. Please see [77] for details.

(20%, 38 trials)—this process was repeated 5 times until each trial had been left out of training once. For each cross validation fold, we used the training data to identify the model coefficients $\mathbf{b}$ from n pairs of sensor signals $\bar{\mathbf{y}}(t_i)$ and associated ground truth energy cost values $E(t_i)$, with all training trials concatenated:

$$\mathbf{b} = \left(\bar{\mathbf{Y}}^T\bar{\mathbf{Y}}\right)^{-1}\bar{\mathbf{Y}}^T\mathbf{E}, \tag{4.6}$$

where the matrix $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}^T(t_1), \ldots, \bar{\mathbf{y}}^T(t_n)]^T$ contains the signal vectors of all n data points in the training set, and the vector $\mathbf{E} = [E(t_1), \ldots E(t_n)]^T$ is the associated ground truth energy cost.

Using the testing data, $\mathbf{y}'(t)$, we used the identified model coefficients from Eq. 4.6 and calculated the predicted energy cost, $\hat{E}(t)$, associated with the testing conditions using Eq. 4.5. We repeated this process 5 times, once for each fold of the cross validation, until we obtained $\hat{E}(t)$ for each condition.

To compare the predicted energy cost $\hat{E}(t)$ to the ground truth energy cost $E(t)$, we considered each 5-minute condition as an individual trial. For each trial (with index k), we computed the absolute error $e_k(t) = \left|\hat{E}_k(t) - E_k(t)\right|$ as a function of trial time. This trial time was measured from the onset of the step change at the beginning of the trial and thus ran from 0 to 5 minutes. The mean absolute error (MAE) $\bar{e}(t)$ was obtained as the average of all 187 individual trials. We further averaged this error metric over the entire trial (*i.e.,* minutes 0–5), and over individual minutes (*i.e.,* minutes 0–1, 1–2, ..., 4–5). We then calculated the reduction in MAE between models trained with and without derivatives for the whole trial and per minute. For example, we calculated the change in MAE between regression model 1 (local signals only), and regression model 4 (local signals and their derivatives) for each of the 5 individual minutes and the entire trial. We repeated this analysis for the global signals (regressions 2 and 5), and the combined signals (regressions 3 and 6).

Finally, we computed four additional metrics for each regression: the initial error $\bar{e}(0)$, the steady state error (the average MAE over the last two minutes), the error decay time (the time at which $\bar{e}(t)$ decrekgased to 36.8% of the range from initial value to steady state value), and the time to 1 W/kg (the time at which $\bar{e}(t)$ decreased below 1 W/kg). We chose the threshold of 1 W/kg because it is 10 % of the total range of metabolic cost experienced by all subjects across all activities [77].

4.2.4 Statistical Analysis

We statistically analyzed the average MAE of the entire 5-minute trial (*i.e,* the average of minutes 0–5). We performed three paired t-tests in MATLAB (one for each signal group, $\alpha = .05$) comparing the average MAE of all 187 conditions between regressions with and without derivatives. Using this paired t-test, we generated 95% confidence intervals around the difference between regressions trained with and without derivatives, for a given signal group. We did not perform statistical tests on the average MAE from individual minutes.

4.3 Results

Including derivatives in regressions that contained the global signals (*i.e.,* global, combined) visibly improved their performance, especially during the first 2 minutes (Fig. 4.2). Over the entire 5-minute trial, the average mean absolute error (MAE) was significantly reduced for all signal groups with the inclusion of derivatives (Fig. 4.3A). The difference in average MAE between regressions with and without derivatives was 0.004 W/kg for local signals (95% confidence interval: $\text{CI}_{95\%} = [0.001, 0.006]$ W/kg, p-value: $p = .004$), 0.088 W/kg for the global signals ($\text{CI}_{95\%} = [0.052, 0.123]$ W/kg, $p = 1.8 \times 10^{-6}$), and 0.081 W/kg for the combined signals ($\text{CI}_{95\%} = [0.052, 0.109]$ W/kg, $p = 1.7 \times 10^{-7}$) (Fig. 4.3B).

On average, in the first minute (minute 0–1), including derivatives in the regression with global signals resulted in a 0.48 W/kg (32.5 %) reduction in prediction error compared to the regression containing global signals only (Figs. 4.4, 4.5). The performance of the combined signals improved by 0.31 W/kg (25.4 %) in the first minute with the inclusion of derivatives, while the local signals exhibited less than a 0.02 W/kg (1 %) difference. In minutes 2–5, the difference between regressions with and without derivatives was less than 0.06 W/kg for the global signals, less than 0.05 W/kg for the combined signals, and less than 0.001 W/kg for the local signals.

For regressions without derivatives, the initial error $\bar{e}(0)$ was between 2–2.3 W/kg for all signal groups (Table 4.2). With the addition of derivatives, the initial error of the local, global and

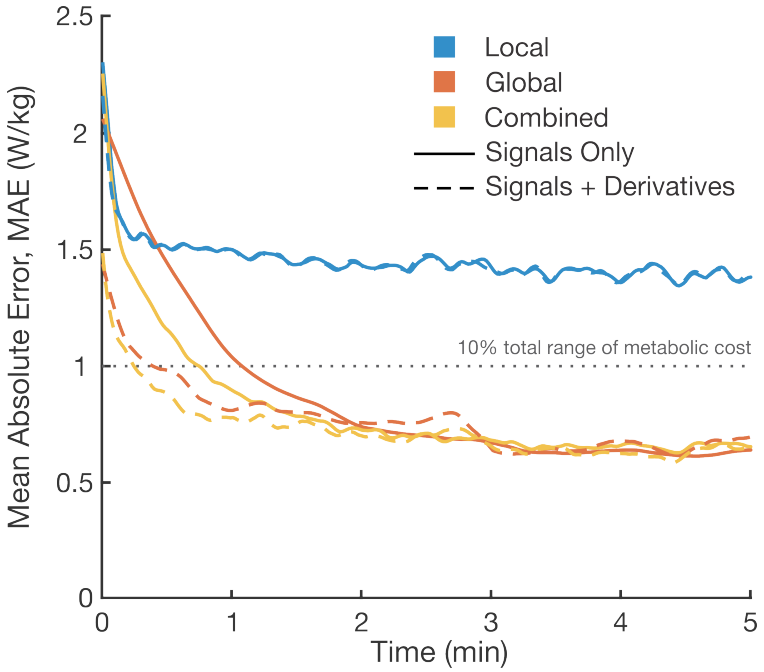

Figure 4.2: Mean absolute error (MAE) as a function of trial time for all linear regression models. The performance of the global and combined signal groups visibly benefited from the inclusion of derivatives. Mean absolute error (MAE) was calculated between predicted energy cost and ground truth energy cost as a function of trial time, and averaged over 187 5-minute trials. The solid lines depict performance of the regression models trained with the local (blue), global (red), and combined (yellow) signals only; the dashed lines of the same color depict performance of the models trained with the signals and signal derivatives. The gray dotted line indicates the threshold of 1 W/kg, which corresponds to 10% of the total range of metabolic cost over all conditions (10 W/kg).

combined regressions dropped by 0.14 W/kg, 0.63 W/kg and 0.77 W/kg, respectively. The difference in steady state error with the inclusion of derivatives was less than 0.03 W/kg for all signal groups (Table 4.2). The mean absolute error (MAE) $\bar{e}(t)$ (Fig. 4.2) of the local signals (with and without derivatives) decayed faster than any of the regressions trained with global or combined signals (Table 4.2). Adding derivatives to the regressions shortened the decay time by 12 s for the global signals, and by 3 s for the local signals; the decay time for the local signals did not change with the addition of derivatives.

When derivative information was included, the MAE from the global and combined regressions reached 1 W/kg in 23 s and 15 s, respectively (Table 4.2); in contrast, without derivatives, the MAE from the global and combined regressions required 63 s and 44 s, respectively. Neither regression trained with just local signals (with or without derivatives) crossed the 1 W/kg threshold.

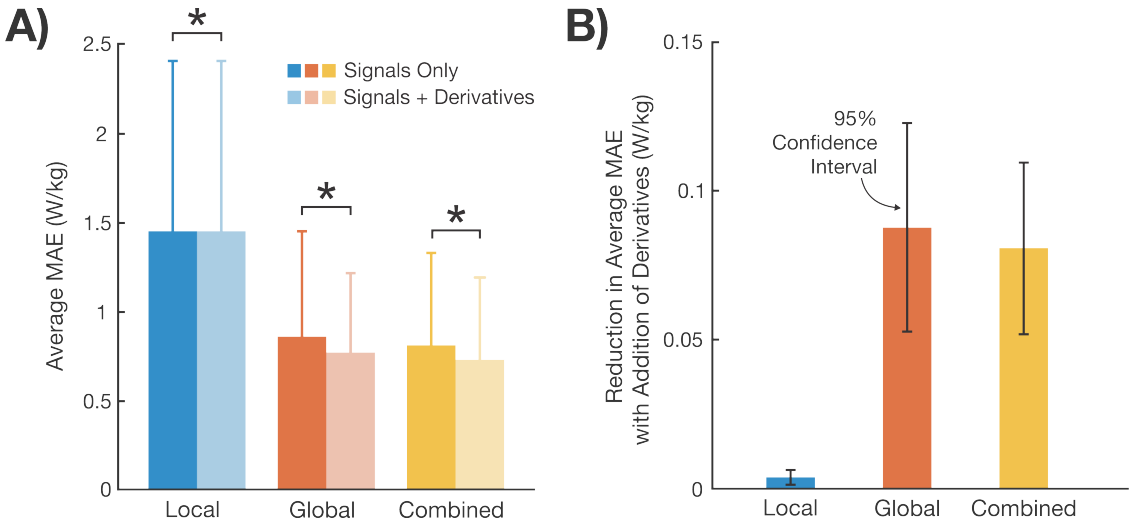

Figure 4.3: Comparison of mean absolute error (MAE) between models with and without derivatives. **(A)** Paired t-tests revealed significant differences in average mean absolute error (MAE) between regressions trained with and without derivatives for all 3 signal groups ($*$ indicates $p < .05$). The average MAE was calculated as the mean across all 5 minutes of MAE for each trial (as depicted in Fig. 4.2). Colored error bars indicate the standard deviation across 187 trials. **(B)** The reduction in average MAE with the addition of derivatives was much smaller for the local signals (0.004 W/kg) than for the global signals (0.088 W/kg) or combined signals (0.081 W/kg). These reductions were calculated as the difference between dark and light bars in Fig. 4.3A; black error bars indicate the 95% confidence interval around the difference, generated from the paired t-test.

	Initial Error* (W/kg)		Steady State Error§ (W/kg)		Decay Time† (s)		Time to 1 W/kg‡ (s)	
	Signals	*+Derivatives*	*Signals*	*+Derivatives*	*Signals*	*+Derivatives*	*Signals*	*+Derivatives*
Local	2.30	2.16	1.39	1.39	7	7	-	-
Global	2.06	1.43	0.63	0.65	49	37	63	23
Combined	2.25	1.48	0.65	0.63	24	21	44	15

*Initial error, $\bar{e}(0)$ (Fig. 4.2).
§Steady state error defined as the average of the error, $\bar{e}(t)$ (Fig. 4.2) for the last 2 minutes (minutes 4 and 5).
†Decay time defined as the time at which $\bar{e}(t)$ decreased to 36.8% of the range from its initial value to its steady state value.
‡Time to 1 W/kg defined as the time at which $\bar{e}(t)$ decreased below 1 W/kg.

Table 4.2: Error metrics.

4.4 Discussion

In this study, we investigated how including signals' derivatives affected the predictive accuracy of multiple linear regression models trained to predict instantaneous metabolic cost. As we expected, the groups containing *global* signals (such as heart rate and minute ventilation) significantly benefited from the inclusion of derivatives (Fig. 4.3). Adding derivatives to the regression trained with *local* signals alone also significantly reduced its predictive error, although the mag-

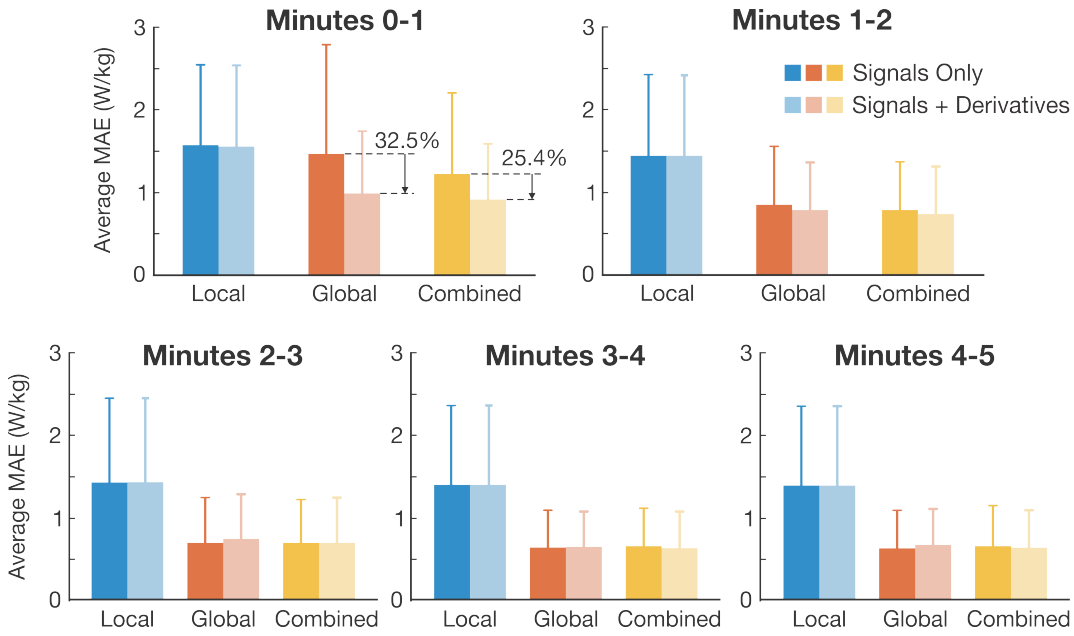

Figure 4.4: Minute-by-minute comparisons of mean absolute error (MAE) between models with and without derivatives. In the first minute, including derivatives resulted in a 32.5 % reduction in error for the global signals and a 25.4 % reduction for the combined signals. Mean absolute error (MAE) from Fig. 4.2 was averaged over individual minutes; error bars indicate the standard deviation over 187 trials. Dark colors indicate regressions trained with local (blue), global (red) and combined (yellow) signals only; light colors indicate regressions trained with signals and their derivatives. We did not perform statistical analyses for each minute separately.

nitude of this reduction was very small (0.004 W/kg). Examining these data minute-by-minute, we observed that the difference in performance between regressions with and without derivatives was most evident during the first minute of the trial (Figs. 4.4, 4.5). In the first minute, regressions trained with the global and combined signal groups demonstrated reductions in mean absolute error (MAE) of 0.48 W/kg (32.5 %) and 0.31 W/kg (25.4 %), respectively, with the addition of derivatives. The initial error for the global and combined regressions also decreased dramatically (by 0.63 W/kg and 0.77 W/kg, respectively) with the inclusion of derivatives. For context, the inter-subject mean metabolic difference between walking at 0.6 m/s (cost: 3.33 W/kg) and walking at 0.9 m/s (cost: 3.74 W/kg) was only 0.41 W/kg. These error reductions are thus quite substantial, especially considering the need to discriminate small changes between similar activities. The results of this analysis suggest that with dynamically delayed signals, the *magnitude* of the error associated with predicting instantaneous energetic cost during the transient phase of an activity can be meaningfully reduced with the inclusion of derivatives.

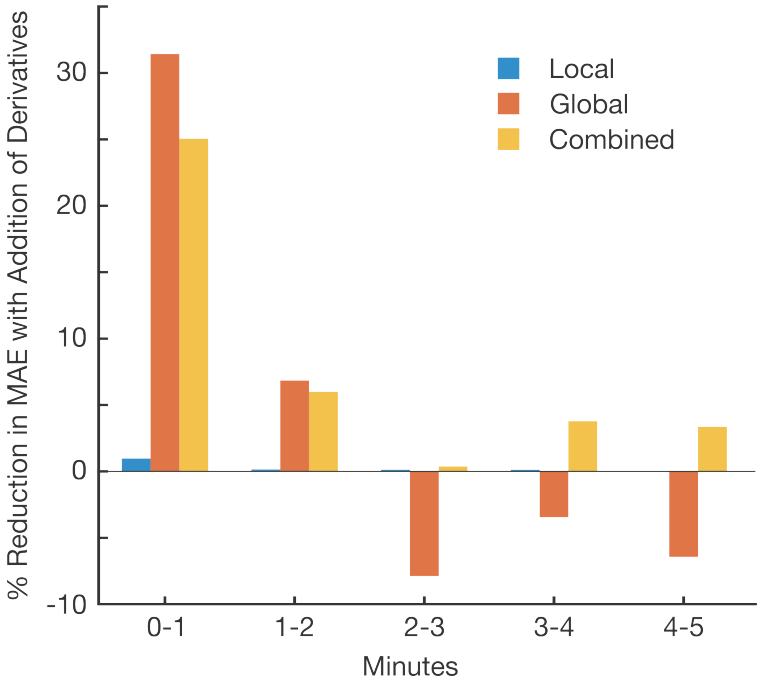

Figure 4.5: Percent reduction in mean absolute error (MAE) with the addition of derivatives. Adding derivative information to the regressions trained with global (red) and combined (yellow) signal groups substantially improved their predictive performance in the first minute of the trial (by 0.48 W/kg and 0.31 W/kg, respectively), compared to regressions trained without derivatives. These reductions were calculated for each signal group and each minute as the difference between dark and light bars in Fig. 4.4. We did not perform statistical analyses for each minute separately.

The reduction in error magnitude obtained from the inclusion of derivatives can also be interpreted as an increase in the *speed* of prediction. This speed is crucially important for many wearable robotics applications, in particular for body-in-the-loop optimization. For this application, the faster we can obtain an accurate estimate of the instantaneous energetic cost for a given activity or set of device parameters, the more conditions we can test in a fixed amount of time, both increasing the efficiency and speed of the optimization routine. Furthermore, during typical activities of daily living, the tasks are highly dynamic and non-steady-state, which means minimizing the time required to obtain an accurate estimate of instantaneous metabolic cost for the current activity is critical for the translation of these algorithms. For other clinical or research evaluations, decreasing the speed associated with obtaining an accurate estimate of instantaneous energetic cost can reduce the experimental time and physical burden on participants.

In this study, we quantified prediction speed using both decay time and the time for the MAE to decrease below 1 W/kg. Both the global and combined signal groups obtained slightly faster decay times with the addition of derivatives (12 s faster for the global signals, and 3 s faster for the combined signals) (Table 4.2). However, it is important to note here that the initial error $\bar{e}(0)$ is already much smaller when derivatives are included. So even though the decay time is not much faster, the absolute magnitude of the error diminishes much more quickly. This effect becomes most evident when we look at a constant error threshold (that is, a metric of an 'acceptable' prediction accuracy) to quantify the speed of the prediction. The MAE from the regression trained with global signals and their derivatives reached 1 W/kg in only 23 s, compared to the 63 s required when no derivatives were included. Similarly, the error from the regression with combined signals and their derivatives reached 1 W/kg in 15 s, compared to 44 s with no derivatives. Therefore, simply including derivatives in the multiple linear regressions can be responsible for large time savings (roughly 30–40 s) when trying to achieve some acceptable error threshold. In this analysis, the error threshold of 1 W/kg was chosen because it represents 10 % of the total range of metabolic cost experienced by all subjects across all tasks [77], and was a common threshold to compare all regressions against. In practice, the acceptable error threshold for a particular application would be determined by individual experimenters.

Importantly, the observed performance improvement in the transient phase did not diminish the performance of the steady state prediction. For all signal groups, the steady state error did not appreciably change between regressions trained with or without derivatives (Table 4.2), and the differences in the MAE in minutes 3, 4, and 5 were smaller than 0.04 W/kg for all signal groups (Figs. 4.4, 4.5). When a signal has reached steady state, its derivative is essentially zero and contains mostly noise, not salient information. The regressions in our study were trained on five minute conditions that included both transient and steady state phases, so the algorithm automatically balanced the relative weights of the signal and derivative terms to use derivative information during the transient phases and reject noise during the steady state phases. In practical implementations, to optimize the trade-off between transient and steady state predictive accuracy, the training data

should be dynamically representative of the data that the estimator will be used on.

Although the motivation for including derivatives in the regressions came from models of linear first order systems, the black-box nature of the linear regression algorithm does not require that the signals fit this form; in fact, this approach extends to other signals whose dynamics have not been explicitly quantified. The other physiological signals included in the global signals group—breath frequency, oxygen saturation, skin temperature, and electrodermal activity—do not have explicitly defined first order dynamics, but are still dynamically delayed in response to a step change in metabolic effort, as documented in [77]. Regardless of the explicit dynamics of an individual signal, the derivative encodes rate of change information, which can be thought of as an additional feature to improve the predictive accuracy during the transient phase if the signal is changing significantly. The regression coefficients will be scaled optimally to exploit this information. The mathematical framework we developed (outlined in Appendix A), allows us to compare the coefficients of a regression model with those identified from an explicit model. However, at this time, it is very difficult to make a one-to-one comparison between the predictive performance of a regression model and an explicit model. Signal filtering, system identification methods, and signals with higher-order or undefined dynamics are all likely to strongly affect this comparison. Although it is outside the scope of the current study, expanding the mathematical framework to systematically compare the predictive performance of explicit dynamic models and regression models is a worthwhile and valuable avenue for future work.

It is important to state that this study was not an attempt to build the best algorithm to predict instantaneous metabolic cost and minimize steady state error. Rather, the purpose of this study was to use multiple linear regression as a common platform to demonstrate the potential that derivatives have to reduce both the *magnitude* and *delay* of predicting instantaneous metabolic cost. This goal motivated several practical choices in our study, which each have limitations. To facilitate comparisons between models with and without derivatives, we sought to keep the signals as close to their raw form as possible. In reality, many of the signals are highly cross-correlated with one another and have standard feature sets that we did not explore in this work. All the regressions we

tested in this study were trained using signal *groups*, and we did not evaluate the relative importance of individual signals and their derivatives within these groups, which would be interesting to explore in future studies. The cross validation technique we used in this study (*i.e.,* training on 80 % and testing on 20 % of the conditions) allowed us to evaluate the performance of our algorithms when predicting novel conditions, but does not allow us to make inferences about how these methods perform when we specifically test them on unknown subjects or unknown tasks. A detailed analysis of how the signal groups we studied generalize to unknown subjects or tasks can be found in [77]. Finally, we chose to use the multiple linear regression algorithm deliberately, because it could be formulated as the corollary to the first order linear model. However, we believe this theory can be applied to many different predictive algorithms in order to further increase predictive capability. For example, Beltrame *et al.* included a feature they called ΔHR (*i.e.,* the difference between the current heart rate value and the previous value by a 1 s lag operator) in their random forest algorithm trained to predict oxygen uptake dynamics [13]. Although they did not evaluate the algorithm with and without the derivative information, using ΔHR in combination with heart rate, minute ventilation, breathing frequency, and hip acceleration resulted in good predictive performance of oxygen uptake kinetics, which highlights that derivative information can be useful in other machine learning algorithms. In future work, it will be important to tune the predictive algorithm and improve steady state error by exploring different feature sets, signal processing techniques, and regression algorithms. We hope to enable other researchers to explore these future avenues by accessing our full open source dataset online [77, 78].

The results of this study indicate that simply including derivative information in the training data has the potential to improve predictive performance, with all other factors held constant. We believe these methods are useful for wearable robotics and body-in-the-loop optimization applications because they may reduce the time needed to evaluate the metabolic cost associated with a robotic device or a particular set of conditions. The dataset we used in this study comprised a diverse set of tasks and intensities that provided substantial metabolic cost variability over a range of 10 $\mathrm{W/kg}$. However, this range is likely greater than that experienced by subjects during a typical

body-in-the-loop optimization routine, so we do not yet know the exact timing or error improvements that may be realized for this particular application. In the future, it would be beneficial to apply these methods to a more specific dataset in order to fully answer this question. These methods can also be applied across a broad spectrum of fields (*e.g.,* rehabilitation medicine, exercise science), and may be useful for any researcher or clinician who relies on measurements of metabolic cost for outcomes. Regardless of the application, to maximize predictive performance, we recommend using training data that is not only representative of the activities one aims to predict, but also captures the desired ratio between transient and steady state behavior.

Incorporating physiological information from the human body into control algorithms for wearable robots can help advance the design and evaluation of these devices for a variety of applications. One such physiological quantity is the user's metabolic energy expenditure, which can provide valuable real-time information about the state of the user. As such, quick, reliable, and portable estimates of instantaneous metabolic cost are required for translating our wearable robots beyond the laboratory environment. Inspired by model-based estimation techniques, in this article we showed that including signal derivatives in a black-box estimation process improves both the magnitude and speed of predicting instantaneous metabolic cost from wearable sensors when the input signals are dynamically delayed. These performance improvements were observed during the transient phase of an activity, while steady state performance remained unchanged. To some extent, signal derivative information can be included in prediction models 'for free,' since it requires no additional sensor hardware and minimal signal processing, yet it has the potential to enhance the predictive capability of the model during transient phases of an activity. This article provides a practical foundation for improving the speed of predicting metabolic energy cost for wearable robotics applications.

CHAPTER 5

User Preference as a Holistic Objective for the Control of Robotic Exoskeletons

5.1 Background

The newest generation of lower-body exoskeletons are high-performance, untethered, and lightweight [88, 107, 120, 151], but have not yet achieved widespread adoption as devices to augment able-bodied ambulation. One major barrier to translation is that designing effective control systems for such devices is challenging due to the added complexity of interfacing directly with the human body. For lower-limb exoskeletons to achieve their full potential as augmentative devices, the robotic assistance must provide a benefit to the user while also synergistically integrating with the user's intact neuromotor system. Fundamental to the realization of this vision is the definition of appropriate efficacy criteria, or objectives, for exoskeleton assistance (*i.e.,* how should robotic assistance impact the wearer?). This is currently a rich field of study, and a universally agreed-upon objective for optimal able-bodied augmentation has not yet been identified. To accelerate the translation of robotic exoskeleton systems, it is imperative that control system designers analytically investigate metrics that capture a broad spectrum of relevant outcomes associated with exoskeleton use. Furthermore, to enable synergistic human-robot interaction, we must expand our definition of optimality and consider not only how robotic assistance affects biomechanical elements of gait, but also how individuals perceive, interact with, and learn from the assistance.

As of this writing, this chapter has been submitted as a manuscript for publication and is currently under review.

In many state-of-the-art exoskeleton control systems, the desired objective is maximized through the tuning–or optimization–of the controller architecture and associated parameters. Commonly, controllers for lower-limb exoskeletons command parameterized assistance (*e.g.,* joint torque or power) as a function of time [120], presumed location in the gait cycle [50, 82, 95, 111, 132], or muscle activity [94], among others. The choice of the controller parameterization and subsequent tuning of these parameters not only govern the behavior of the robot but also dramatically impact the wearer's walking mechanics and energetics [50, 82, 111, 132] and experience wearing the device [186]. While various objectives for lower-limb exoskeleton assistance have been explored (*e.g.,* [83]), by far the most common physiological objective for able-bodied augmentation has been to reduce the metabolic cost of the user [151]. A promising technique to minimize the user's energy cost is known as human-in-the-loop optimization [35, 44, 90, 189]. In this paradigm, the user performs a continuous activity, such as walking on a treadmill, while an online optimization algorithm iteratively modifies characteristics of the exoskeleton's assistance profile to arrive at the optimal settings. A real-time estimate of the user's metabolic cost serves as the cost function that the algorithm seeks to minimize [154]. Recent experiments have demonstrated that human-in-the-loop optimization has the potential to yield large energetic savings while wearing an exoskeleton with optimized control settings compared to unassisted walking [35, 189]. However, current human-in-the-loop optimization methods are fundamentally limited due to their reliance on one, predefined, measurable physiological objective as a metric of device efficacy.

In reality, it is likely that users prioritize many different metrics simultaneously while wearing a robotic exoskeleton, such as comfort, stability, pain, symmetry, or perceived effort. Thus, current controller optimization techniques based solely on minimizing energy expenditure omit valuable information. Yet, it is challenging to design a control system based upon subjective quantities. Even if it were possible to robustly measure such metrics, we would not know how to properly assign weights to the discrete metrics to meet an individual user's needs. Alternatively, asking users about their *preference* may inherently encode relevant information from multiple sources related to the user's experience. With user preference as a holistic objective function for exoskeleton control

optimization, the potential emerges for individuals to customize the device assistance according to their own criteria for optimality. In this paradigm, the user performs their own individual optimization internally, which eliminates the burden on researchers to identify, measure, and assign individual weights to the myriad unknown elements that comprise optimal exoskeleton assistance.

In addition to its role as a target for control parameter optimization, user preference may also be used to promote synergistic human-robot interaction. While current exoskeleton controllers with the human 'in the loop' measure physiological signals (*e.g.,* muscle activity, metabolic cost) to inform the assistance generation, this type of shared-control paradigm is unidirectional–the wearer has no conscious control over the assistance provided by the device. Researchers across many fields of assistive and rehabilitation robotics have demonstrated that bidirectional interaction (*i.e.,* control sharing) is generally preferred by users, and that high levels of user satisfaction are required for the acceptance of autonomous robotic assistance [10, 40, 56, 84, 87]. Yet, control sharing is a relatively nascent field when applied to robotic exoskeleton technology, especially in systems designed for able-bodied augmentation. It is not yet known which types of control sharing paradigms may be appropriate for lower-limb exoskeleton control systems. To address this challenge, it is important that we understand the characteristics and temporal features of an individual's preference, like how quickly and reliably users can identify their preferences or how their preference changes over time. Understanding these facets of user preference may provide practical benchmarks for implementing future shared-control frameworks.

To utilize preference as an optimization criterion for lower-limb exoskeletons, it must be robustly measured, and we need to understand how it changes based on adaptation, learning, and interaction with the device. Recent studies have investigated users' preferences in characteristics of the assistance provided by lower-limb prostheses [22, 28, 155, 157, 158, 168] and exoskeletons [80, 105, 172, 173, 186]. To measure preference, some studies utilized a two-alternative forced choice paradigm, in which subjects were presented with pairwise comparisons (*i.e.,* A-B testing) [105, 168, 172, 173], or asked to compare a condition to their own internalized preference [157, 158]. Data obtained from forced choice methods can be used to identify or learn a

user's preference, but it may require extended experimental time to obtain a sufficient number of comparisons. To mitigate this challenge and accelerate the estimation of a user's preference, one study implemented a coactive learning paradigm in which users not only chose between paired conditions but also directly suggested improvements for the assistance each trial (*e.g.,* longer or shorter step length) [173]. As an alternative to forced choice methods, some studies employed a method of adjustment, in which individual users (or experimenters directed by users) adjusted control parameters until they achieved their preferred assistance [22, 28, 80, 155]. Such 'self-tuning' methods are advantageous because they can quickly yield an individual's preference and are intuitive to the subject–it is easy to imagine someone using a smart phone or smart watch to adjust their settings during activities of daily living. Previous studies have demonstrated success using self-tuning to identify individual preferences in one dimension [28, 155], yet it is an open question how to generalize these methods to multiple dimensions. One option is to perform one-dimensional sweeps through each parameter simultaneously (as in [22]), but this method is time consuming and does not allow subjects to assess how interactions between parameters affects their preference. Prior studies have laid the foundation for the robust measurement of user preference, yet to our knowledge, no studies have investigated how user preference changes as individuals adapt to and learn from robotic assistance.

Incorporating user preference into exoskeleton control systems has the potential to increase user satisfaction, promote acceptance, and add a rich source of holistic information that cannot be easily measured using standard biometric sensors. To inform future preference-based control frameworks, the goal of this study was to measure and analyze users' preferences of the applied torque characteristics of bilateral robotic ankle exoskeletons. Furthermore, we sought to characterize how preference varies with walking speed, how preference differs between individuals with different technical backgrounds, and how preference changes as individuals learn and adapt to robotic assistance. To this end, we developed a two-dimensional grid interface that allowed users to explore in two dimensions simultaneously to identify their preferred assistance [80]. Importantly, individuals adjusted their assistance using a blank grid, and they were not given any

information about the control system parameters—the only instruction they were given was to 'find your preference'. Within this paradigm, users had to identify the assistance that satisfied their internal criteria by relying on their perception of the robotic assistance to guide their tuning.

In this study, 24 individuals wearing bilateral robotic ankle exoskeletons self-tuned two parameters of the device's assistive torque profile to identify their *preferred settings* (Fig. 5.1A). Participants were presented with a blank two-dimensional grid displayed on a touch screen tablet (Fig. 5.1B), and touching the grid instantaneously changed the torque profile the user experienced by commanding the magnitude and the timing of peak torque delivery (Fig. 5.1C). Subjects were instructed to explore the grid before deciding on their preference (Fig. 5.1D). Of the 24 participants, twelve were *naive* users (no prior exoskeleton experience) and twelve were *knowledgeable* users (researchers in the field of wearable robotics with prior exoskeleton knowledge or experience). The naive cohort completed three blocks of eight preference-identification trials, where the first trial block comprised trials 1–8, the second comprised trials 9–16, and the third comprised trials 17–24. The participant's walking speed (1.0, 1.2, or 1.4 m/s) varied between trial blocks and we block randomized the presentation of speeds within the subject pool. The knowledgeable cohort performed a shortened version of the protocol and performed eight preference-identification trials while walking at 1.2 m/s. To characterize user preferences, we statistically analyzed eight outcome measures related to subjects' preferences, precision, and exploration strategy. These outcomes were subjects' preferred settings (magnitude and timing), the standard deviation of subjects' preferences across repeated trials (magnitude and timing), the total number of settings explored per trial, the total exploration time per trial, the average number of strides taken per setting, and the average time spent per setting. For the naive cohort, we used linear mixed effects models to evaluate the effects of walking speed and trial block (*i.e.,* time spent interacting with the device) on each outcome. For the knowledgeable cohort, we used two-tailed *t*-tests to compare the preference and precision outcomes to the naive cohort during the speed-matched condition.

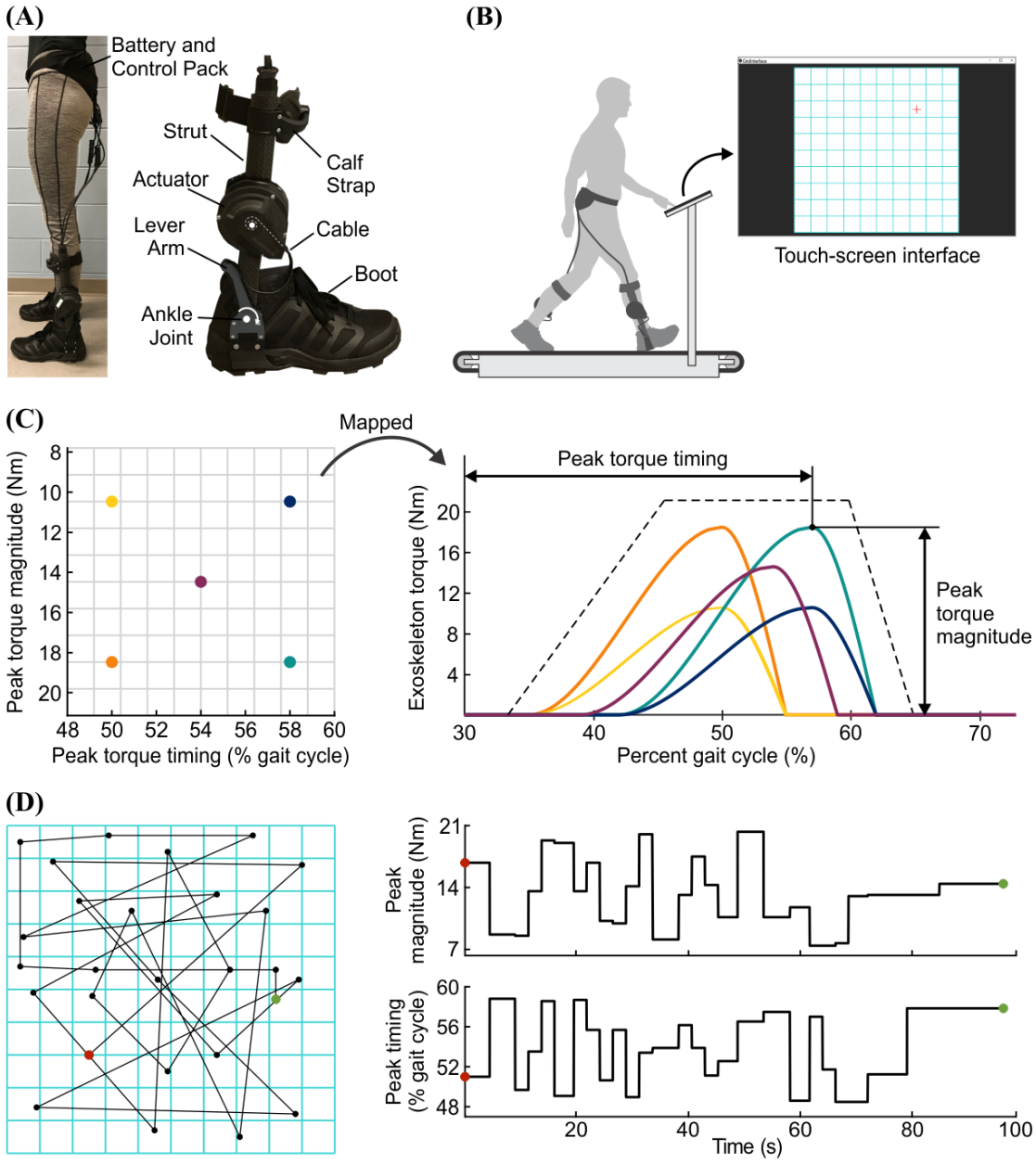

Figure 5.1: Bilateral ankle exoskeletons and preference-identification protocol. **(A)** Bilateral ankle exoskeletons (Dephy Inc., Maynard, MA) **(B)** Participants self-tuned their exoskeleton control system using a blank 2-dimensional grid displayed on a touch screen tablet mounted to the treadmill. **(C)** Touching the grid instantaneously changed the torque profile the user experienced by commanding the peak torque magnitude and timing. The axes of the grid were hidden from the subject but are depicted here to illustrate the mapping between the parameter coordinates and their corresponding torque profiles. The black dashed line encompasses the approximate bounds of torque profiles the user could experience within the limits of the grid. **(D)** A representative preference-identification trial. Subjects explored a blank 2-dimensional grid, beginning at a random initial condition (red circle), and ending at their preferred settings (green circle). Black circles indicate the settings the participant sampled. Black lines connect settings in the order they were explored; these lines were added for visualization and did not appear on the subject's screen. The corresponding time-series representation of the participant's exploration is shown to the right.

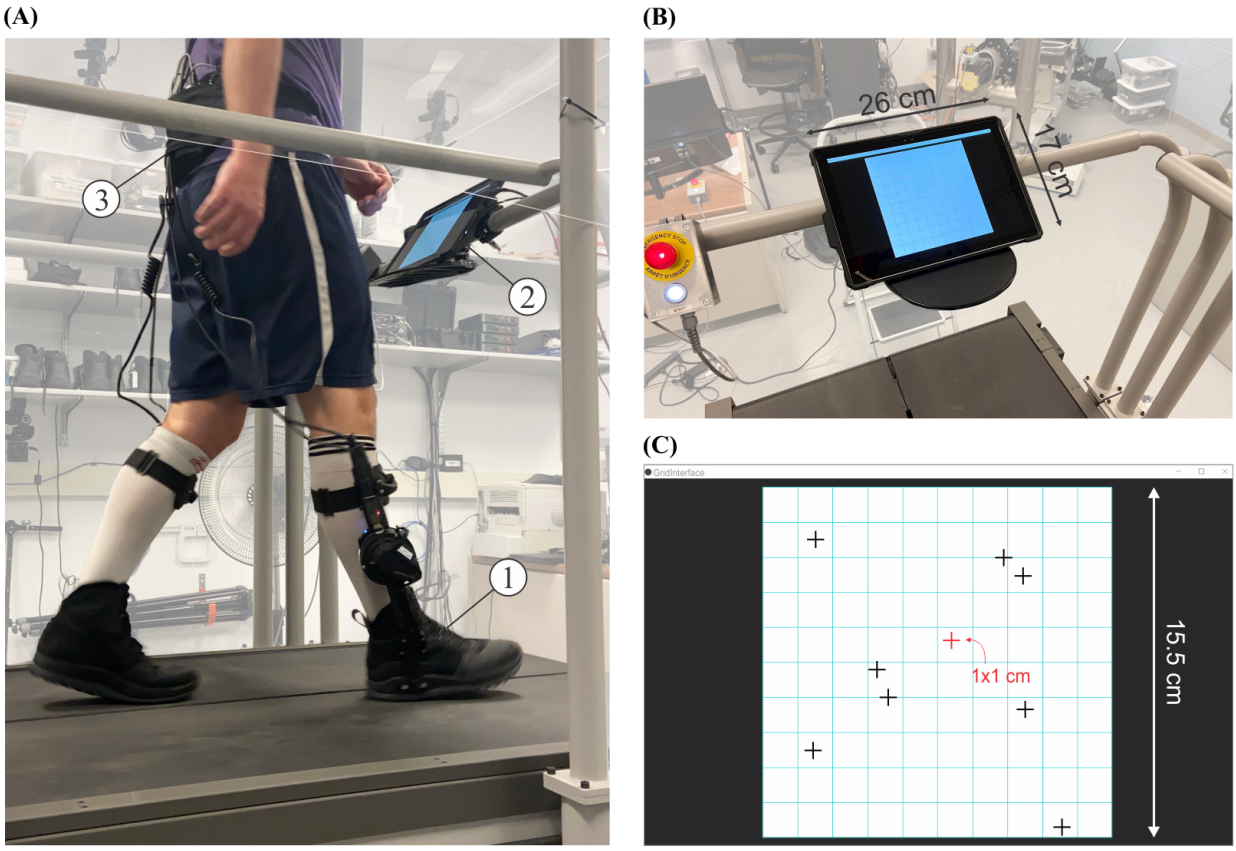

Figure 5.2: Experimental setup and touch screen interface. **(A)** (1) Subjects wore bilateral ankle exoskeletons (Dephy Inc.) as they walked on a treadmill (Bertec). (2) The two lithium polymer batteries and the control microprocessor (Raspberry Pi) were carried in a hip sack. (3) Subjects interacted with a touch screen tablet (Microsoft Surface) mounted to the treadmill in front of them to self-tune their exoskeleton control system. The subject's selections from the touch screen were transmitted wirelessly to the Raspberry Pi. An emergency stop button was located to the left of the touch screen. **(B)** The touch screen tablet (Microsoft Surface Pro 4) was mounted to the treadmill. The screen measured 26×17 cm. **(C)** The tablet displayed a 2-dimensional grid. The displayed grid was 10×10 squares and measured 15.5×15.5 cm. When the subject touched the screen, a small red crosshair (1×1 cm) appeared to indicate the current setting; previous settings remained on the screen as black crosshairs.

5.2 Methods

5.2.1 Study Design

The objective of this study was to quantify individual users' preferences in their bilateral ankle exoskeleton assistance settings. This study utilized a crossover design, in which twelve naive exoskeleton users self-tuned the magnitude and timing of the assistive torque burst using a touch

screen tablet (Fig. 5.1). We conducted repeated preference-identification trials at each of three walking speeds, presented in a block-randomized order. With this design, we sought to quantify how individual preferences varied with both walking speed and time spent interacting with the device (*i.e.,* trial block). Within this paradigm, we characterized how individuals searched for and identified their preferences and learned to use the exoskeleton assistance and tuning interface. Finally, we investigated how prior knowledge of exoskeleton control systems affected user preference. We repeated a shortened version of the self-tuning protocol at one speed with a group of twelve knowledgeable subjects (researchers in the field of wearable robotics) and compared their preferences to those of the naive subjects.

5.2.2 Participants

We recruited a naive cohort of twelve able-bodied subjects (6 male, 6 female; mean (SD) age: 32.1 (8.3) years, height: 1.77 (0.08) m, weight: 78.6 (12.5) kg) who had no prior experience walking using a robotic exoskeleton and were not researchers in the field of wearable robotics; this information was self-reported by subjects on a post-experiment questionnaire. We also recruited a knowledgeable cohort of twelve able-bodied subjects (9 male, 3 female; mean (SD) age: 27.3 (3.5) years, height: 1.77 (0.11) m, weight: 69.0 (10.3) kg) who self-identified on the questionnaire as researchers in the field of wearable robotics and reported from 0–100+ hours of walking using a robotic exoskeleton. Prior to data collection, all participants provided informed consent to a protocol approved by the University of Michigan Institutional Review Board. Individuals recruited for this study had no history of serious lower limb injury, no neurological diseases affecting their movement or balance, and were not pregnant. Before beginning the experiment, subjects self-reported their gender, height, and age, and we measured their weight (Table 5.1).

Naive cohort						Knowledgeable cohort					
Subject	Gender	Age (years)	Height† (m)	Weight* (kg)	Prior time walking with exo† (hours)	Subject	Gender	Age (years)	Height† (m))	Weight* (kg)	Prior time walking with exo† (hours)
S01	F	25	1.80	100.0	0	S13	F	30	1.63	70.4	5
S02	F	30	1.63	67.1	0	S14	M	26	1.70	86.3	0
S03	M	33	1.85	86.4	0	S15	M	22	1.83	62.0	1
S04	M	40	1.71	67.9	0	S16	M	27	1.83	61.0	0
S05	M	35	1.85	92.8	0	S17	M	22	1.80	85.7	>100
S06	M	37	1.83	86.0	0	S18	M	32	1.73	73.5	0
S07	F	31	1.78	85.9	0	S19	M	30	1.75	72.5	7
S08	F	19	1.63	60.6	0	S20	F	26	1.70	58.7	8
S09	M	24	1.80	76.0	0	S21	M	26	1.68	57.7	>100
S10	M	51	1.83	76.3	0	S22	M	26	2.06	64.6	0
S11	F	32	1.68	61.8	0	S23	F	27	1.69	58.3	1
S12	F	28	1.80	82.9	0	S24	M	33	1.79	77.3	25
Mean (SD)	-	32.1 (8.3)	1.77 (0.08)	78.6 (12.5)	0 (0)	Mean (SD)	-	27.3 (3.5)	1.77 (0.11)	69.0 (10.3)	20.6 (37.8)
(Min, Max)	-	(19, 51)	(1.63, 1.85)	(60.6, 100.0)	(0, 0)	(Min, Max)	-	(22, 33)	(1.63, 2.06)	(57.7, 86.3)	(0, 100)

*Subjects were weighed while wearing the exoskeleton. The weights presented here have had the weight of the exoskeleton (5 kg) subtracted.
†Self-reported.

Table 5.1: Subject characteristics of naive and knowledgeable cohorts.

5.2.3 Ankle Exoskeleton Hardware

Subjects wore bilateral ankle exoskeletons (Dephy Inc., Maynard, MA) (Figs. 5.1, 5.2). Each exoskeleton consists of an actuator, a boot, and a strut. The actuator (T-motor U8-KV100, Nanchang, Jiangxi, China) is a brushless motor that provides plantarflexion torque about the ankle joint by spooling an inelastic cable that is rigidly attached to a short lever arm on the boot. The lever arm is rigidly connected to a footplate embedded in the sole of the boot. The actuator is mounted to a strut, which is connected to the boot via the ankle hinge joint and affixed to the wearer's shank with an adjustable calf strap. Each exoskeleton is equipped with onboard sensors: an absolute encoder at the ankle joint that measures ankle joint angle and velocity, an incremental encoder on the motor that measures motor angle and velocity, a current sensor that measures motor current, and an inertial measurement unit (IMU) that measures acceleration and angular velocity of the shank.

The exoskeleton geometry creates a nonlinear transmission ratio, which varies as a function of ankle angle (as measured by the absolute encoder). Using a benchtop test, we empirically characterized the exoskeleton's transmission ratio curve (*i.e.,* transmission ratio as a function of ankle angle) and represented this function as a fourth-order polynomial. Over the ankle angle's normative range of motion (40 degrees plantarflexion to 20 degrees dorsiflexion) the transmission ratio varies from 17:1 to 7:1, with a mean of 15:1. A detailed description of the characterization procedure is provided in Section 5.2.5.

Subjects wore a hip sack that contained the exoskeleton lithium polymer batteries and the control microprocessor (Raspberry Pi 4, Cambridge, UK). In total, the bilateral device weighed 5 kg (boots: 1.1 kg; bilateral exoskeletons: 2.7 kg; battery and control pack: 1.2 kg). Before data collection, we fit each subject with the exoskeletons by selecting the properly sized boot (US men's sizes 7, 9, 10, or 12), securing the calf straps and hip sack, and adjusting the posterior wires so they did not hinder the subject's movement.

5.2.4 Ankle Exoskeleton Control

The exoskeleton controller employed in this study prescribed exoskeleton torque as a function of stride percentage. The torque profiles exerted by the device were characterized by two shaping parameters: the magnitude of peak torque (in Nm) and the timing of peak torque (in percent gait cycle) (Fig. 5.1C). These torque profiles were produced using the methodology established by Zhang *et al.* [189], which generates polynomial coefficients that define the rising and falling portions of the exoskeleton torque burst as a function of stride percentage. A finite state machine governed the exoskeleton torque generation during each stride. The state machine consisted of five states, and transitions between states occurred at set percentages of the stride (Fig. 5.3). The total stride time used to calculate the current stride percentage was the average of the past three stride times, where stride time was measured from left heel contact to left heel contact. We identified heel contact using shank angular velocity measurements from the on-board gyroscope. During states when the exoskeleton produced plantarflexion torque, we controlled the device using current control. At each time step, we calculated the current stride percentage and used the torque profile's polynomial coefficients to calculate the desired exoskeleton torque [189]. Desired exoskeleton torque was then converted to desired motor torque using the instantaneous transmission ratio (obtained using the transmission ratio curve and instantaneous measured ankle angle). Desired motor torque (T_m) was translated into desired current (i) using $i = T_m/k_t$, where k_t is the q-axis torque constant; for this motor, the torque constant was empirically determined to be $0.14\ ^{\text{Nm}}/_{\text{A}}$ [103]. The commanded current was enforced on the motor through a closed-loop current controller.

During the remaining states, which included early stance and swing phases, the desired exoskeleton torque was zero. In these states, the exoskeleton behaved 'transparently' and exerted zero torque on the user, while keeping the cable nearly taught so the actuator was prepared to produce torque with minimal lag time. This was accomplished using position control to enforce the desired motor position that maintained enough tension in the cable, while also exerting no torque on the user. Further details about the implementation of the transparent controller can be found in Section 5.2.5.

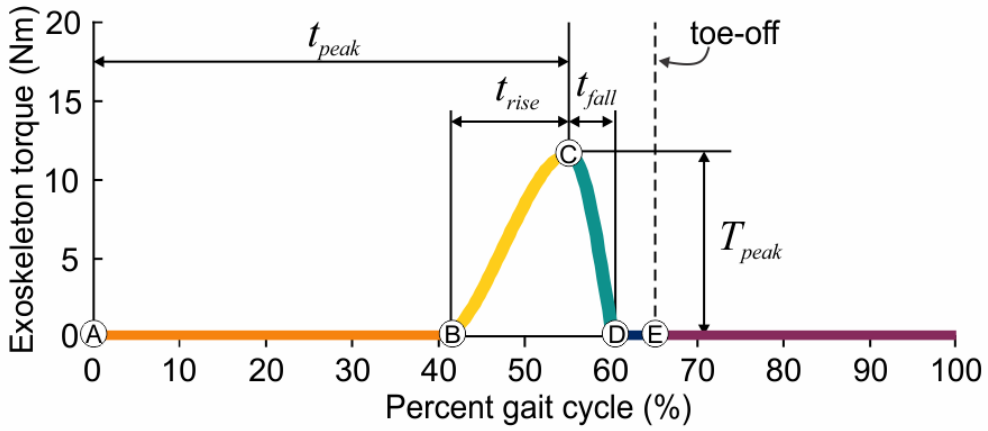

Figure 5.3: Parameterized torque profile as defined by the finite state machine. The exoskeleton produced torque as a function of percent gait cycle. The torque profile was parameterized by four shaping parameters, as in [189]: the magnitude of peak torque (T_{peak}, Nm), the timing of peak torque (t_{peak}, % gait cycle), the rise time of the torque (t_{rise}, % gait cycle), and the fall time of the torque (t_{fall}, % gait cycle). The magnitude and timing of peak torque were commanded by the subject in real time using the touch screen, while the rise time (15 % gait cycle) and fall time (5 % gait cycle) were held constant. We commanded exoskeleton torque using a finite state machine with five states. Early stance (orange) was defined from (A) heel contact (0 % gait cycle) to (B) $t_{peak} - t_{rise}$, or the time when the exoskeleton began to produce torque. Ascending push-off (yellow) was defined from (B) $t_{peak} - t_{rise}$ to (C) t_{peak}. Descending push-off (teal) was defined from (C) t_{peak} to (D) $t_{peak} + t_{fall}$. Late stance (blue) was defined from (D) $t_{peak} + t_{fall}$ to (E) toe-off (dashed black line, fixed at 65 % gait cycle). Swing (purple) was defined from (E) toe-off to (A) the subsequent heel strike. During early stance, late stance, and swing phases, the desired exoskeleton torque was 0 Nm. During ascending and descending push-off, desired exoskeleton torque was defined as a cubic spline; polynomial coefficients for each ascending and descending portion of the curve were generated according to the methodology established in [189].

5.2.5 Ankle Exoskeleton Benchtop Characterization

We characterized the mechanics of the exoskeleton using a benchtop test. With the boot fixed to the table, we held the shank of the exoskeleton vertically. We provided the motor with a very low amount of current (∼0.5 A), enough to spool the cable into tension but not enough to provide appreciable torque. We slowly moved the exoskeleton shank through the ankle's full range of motion (from 60 degrees plantarflexion to 40 degrees dorsiflexion) and repeated this sweep 4 times over 20 seconds. With data recorded from the absolute encoder at the ankle joint and the incremental encoder on the motor, we determined the relationship between motor angle and ankle angle, with the cable taught but exerting no torque. We represented this function as a fourth-order polynomial, and it is referred to as the *motor-ankle curve* (Fig. 5.4A). We then numerically differentiated the motor-ankle curve to produce the transmission ratio as a function of ankle angle and represented this using another fourth-order polynomial. This function is referred to as the *transmission ratio curve* (Fig. 5.4B).

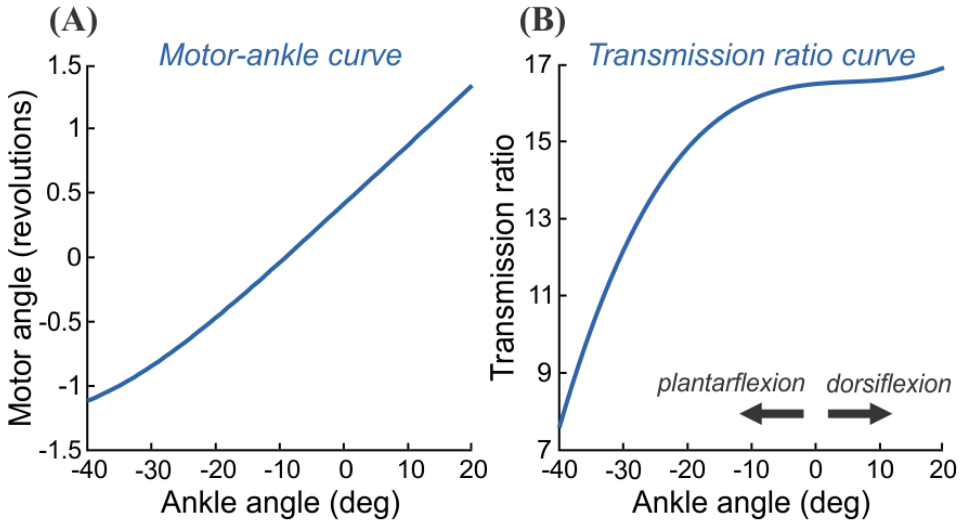

Figure 5.4: Benchtop characterization of exoskeleton mechanics. **(A)** Motor-ankle curve used to determine the required motor position to keep the cable taught but exert no torque on the user at a given ankle angle. **(B)** Transmission ratio curve used to determine the instantaneous transmission ratio at a given ankle angle. A positive ankle angle indicates dorsiflexion and a negative ankle angle indicates plantarflexion. The motor-ankle curve and transmission ratio curve are shown over the normative range of ankle angle observed during walking, and are both characterized as fourth-order polynomials.

In our control system, during states when the exoskeleton produced plantarflexion torque, we required the instantaneous transmission ratio to convert desired ankle torque into desired motor torque at each time step. We calculated the instantaneous transmission ratio using the transmission ratio curve as a lookup function, given the instantaneous ankle angle measured by the absolute encoder. During states in which the exoskeleton behaved transparently (*i.e.,* the desired ankle torque was zero), we utilized the motor-ankle curve. We used this function to calculate the motor position required to keep the cable taught but exert no torque on the user, given the instantaneous ankle angle measured by the absolute encoder. We added approximately 40 degrees (0.1 revolutions) to the calculated taught motor position to introduce enough slack in the cable for the user to comfortably move their ankle through its range of motion during walking. This method ensured that the exoskeleton exerted no torque on the user, yet the cable was nearly taught so the device could generate plantarflexion torque with minimal lag time.

5.2.6 Touch Screen Interface

To enable subjects to manually tune their exoskeleton assistance settings in real time, we utilized a touch screen tablet (Microsoft Surface Pro 4, Microsoft Corporation, Redmond, WA) mounted to the treadmill (Fig. 5.1B). The touch screen displayed a blank 2-dimensional grid (Fig. 5.2). When the subject touched the screen, a small red crosshair appeared to indicate the current setting; previous settings remained on the screen as black crosshairs. Touching the screen changed the assistance settings that the user experienced, beginning at the following left heel contact. Behind the scenes, the unlabeled Cartesian coordinates of the 2-dimensional grid were mapped to the parameter coordinates (*i.e.,* magnitude and timing of peak torque) (Fig. 5.1C). To prevent the subjects from returning to the same (*x, y*) location each trial, we created eight different maps, in which the *x*- and *y*-axes either corresponded to peak torque magnitude or peak torque timing. Then, each axis was either oriented smallest-to-largest or largest-to-smallest, with respect to the origin. All combinations of axis parameters and orientations resulted in eight unique maps, which were hidden from the subjects and randomized between trials.

For the peak torque magnitude parameter, subjects could explore from approximately 7.8–20.7 Nm. For the lower limit of torque magnitude, we chose 7.8 Nm instead of 0 Nm to maximize the resolution of the grid. Based on pilot testing [80], we determined that this was the approximate inflection point at which subjects could begin to feel the exoskeleton assistance. Therefore, by setting the lower limit to 7.8 Nm, we ensured that users could feel the exoskeleton assisting them at all locations of the grid. Because the peak commanded exoskeleton torque was not always the same as the peak measured exoskeleton torque, subjects did not always experience the exact same boundaries of the grid. Across all subjects and strides, the mean error between the peak commanded torque and the peak measured torque was 3.9 %. Therefore, we expect that the lower bound of subjects' exploration was within 7.5–8.1 Nm, and the upper bound was within 19.9–21.5 Nm. In this study, we report each subject's preferred magnitude in *measured peak torque* to capture what users experienced when they chose their preferred settings.

For peak torque timing, the full range of assistance settings available to the subjects to explore was 48–60 % gait cycle. The bounds for the timing parameter were selected from empirical testing and informed by the literature. For our exoskeleton and control system, timings earlier than 48 % gait cycle were very uncomfortable for users and decreased their ability to walk normally. Yet, 48 % gait cycle was low enough to capture the metabolically-optimal exoskeleton push-off timing reported in previous studies [119, 189]. For both parameters, the grid was linearly interpolated on both axes between the minimum and maximum values over 74 points, resulting in a resolution of 0.174 Nm for torque magnitude and 0.162 % gait cycle for torque timing.

5.2.7 Preference-Identification Protocol

In this experiment, subjects identified their preferred settings by self-tuning the magnitude and timing of their peak exoskeleton assistance in real time using the touch screen interface. During a single preference-identification trial, participants began with a random assignment of one of the eight maps described in Section 5.2.6, and a randomized initial condition between 50–57 % gait cycle for torque timing and 10.3–19.4 Nm for torque magnitude. All subjects were read an identical

short script at the beginning of each experiment to communicate the goals of the study. Subjects were told, "The goal of this device is to assist you while you walk. This device can help you move your ankle by pushing your toes down. Try to let it assist you. The setting on the screen changes the way it assists you and may change how it feels. Your goal is to find the setting you prefer. We are going to do this many times under different conditions, but your goal is always the same—find the setting you prefer. Each new condition is different, so the same (*x, y*) location on the grid may not feel the same between conditions. You should explore the grid to find the setting you like the best. Try to walk for at least 3 steps before changing the setting." Subjects explored the grid to find their preferred settings (as depicted in Fig. 5.1D). When they found their preferred settings, they informed the experimenter, who stopped the trial. We verbally reminded subjects to explore the space until they were confident in their preference but did not enforce any time limit on subject exploration.

Participants in the naive cohort performed three blocks of eight preference-identification trials, for a total of 24 trials. During each trial block, subjects walked on a treadmill (Bertec, Columbus, OH) at 1.0 m/s (slow), 1.2 m/s(nominal), or 1.4 m/s (fast). The order of speeds presented was block-randomized within the subject pool to ensure even distribution, and two subjects performed each of the six possible orders of speeds. Participants in the knowledgeable cohort performed one block of eight preference-identification trials at the nominal speed (1.2 m/s).

Before beginning the preference-identification trials, all subjects performed a practice trial, during which they were allowed to explore the grid for as long as they wished to get used to the exoskeleton assistance. We encouraged subjects to explore out to the edges of the grid to familiarize themselves with possible assistance settings they might experience during the experimental trials. On average, naive subjects practiced walking with the exoskeletons for 4.6 (2.1) minutes (mean (SD) across ten naive subjects; we were unable to save the practice trials from Subjects 2 or 3). Knowledgeable subjects practiced walking with the exoskeletons for 3.5 (1.4) minutes (mean (SD) across twelve knowledgeable subjects).

5.2.8 Data Analysis

During each preference-identification trial, we recorded time series of all assistance settings and (*x, y*) coordinates that participants experienced (e.g., Fig. 5.1D). We simultaneously recorded data from the exoskeleton, including motor angle, velocity, voltage, and current, ankle angle and velocity, and shank angular velocity (Figs. 5.5, 5.6). At the termination of each trial, we noted the final settings and coordinates corresponding to the subject's preference. We assessed eight outcomes in this study, divided into three categories: 1) preference outcomes, 2) precision outcomes, and 3) exploration strategy outcomes. The two preference outcomes were preferred peak torque magnitude (Nm) and preferred peak torque timing (% gait cycle), which corresponded to the final settings the participant chose. The two precision outcomes were designed to assess subjects' repeatability in identifying their preferred settings across trials. They were the standard deviation of preferred peak torque magnitude (Nm) and the standard deviation of preferred peak torque timing (% gait cycle). The four exploration strategy outcomes described how the participants performed the experiment and interacted with the touch screen. They included the number of settings an individual sampled before confirming their preference, the total exploration time before confirming their preference (seconds), the average number of strides (left heel contact to left heel contact) taken per setting before changing to a new setting, and the average time spent per setting before changing to a new setting (seconds).

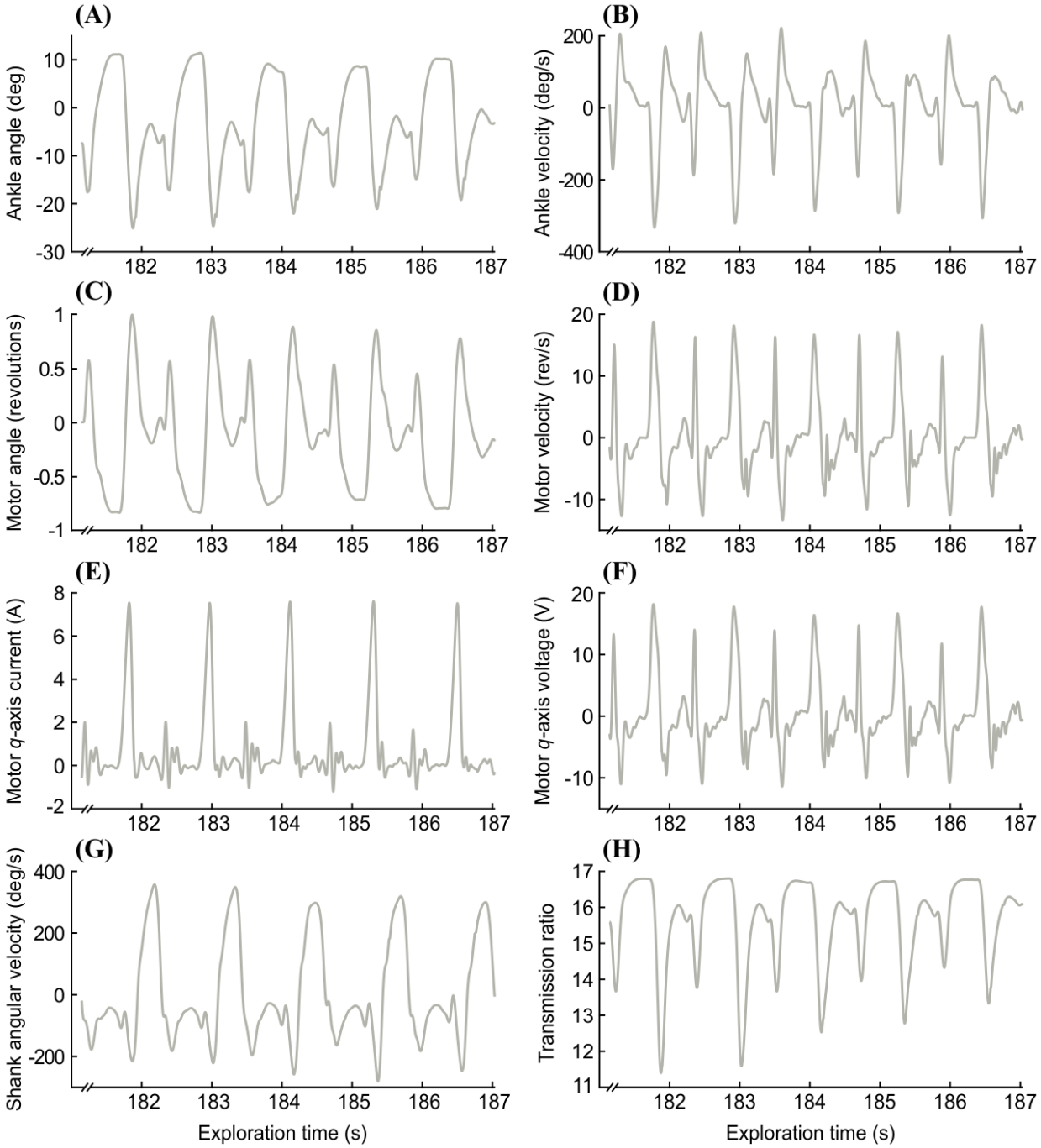

Figure 5.5: Signals collected from the exoskeleton. The signals depicted here were collected from the last five steps at the preferred setting from a representative trial (Subject 9, nominal speed). The traces begin and end at heel strike. **(A)** Ankle joint angle (dorsiflexion is positive). **(B)** Ankle joint velocity. **(C)** Motor angle. **(D)** Motor velocity. **(E)** Motor q-axis current. **(F)** Motor q-axis voltage. **(G)** Shank angular velocity. **(H)** Transmission ratio. For this figure, ankle velocity, motor angle, motor current, shank angular velocity, and transmission ratio trajectories were smoothed using a zero-lag low-pass filter with a cutoff frequency of 25 Hz. Motor voltage and motor velocity trajectories were smoothed using a zero-lag low-pass filter with a cutoff frequency of 50 Hz. The sampling frequency was approximately 500 Hz.

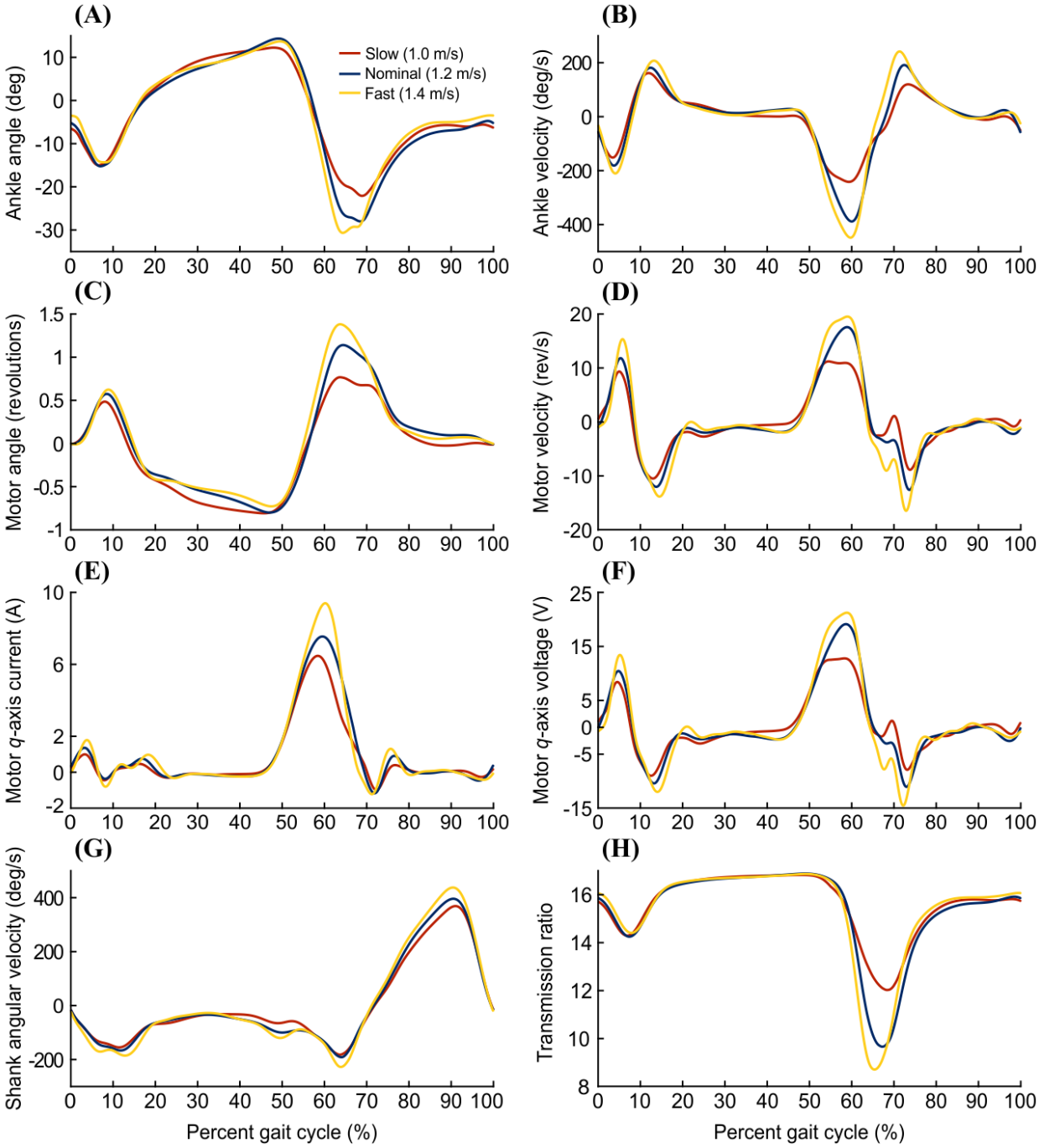

Figure 5.6: Mean exoskeleton signals at each walking speed. Mean signal traces were generated for Subject 5 as the average of all the steps taken at the preferred setting, concatenated across each trial performed at a particular speed. The slow speed is shown in red ($N = 71$ steps), the nominal speed is shown in blue ($N = 68$ steps), and the fast speed is shown in yellow ($N = 65$ steps). Heel strike corresponds to 0 % gait cycle. **(A)** Mean ankle joint angle (dorsiflexion is positive). **(B)** Mean ankle joint velocity. **(C)** Mean motor angle. **(D)** Mean motor velocity. **(E)** Mean motor q-axis current. **(F)** Mean motor q-axis voltage. **(G)** Mean shank angular velocity. **(H)** Mean transmission ratio. Signals were filtered before averaging, as in Fig. 5.5.

5.2.9 Statistics

All statistical analyses were performed using MATLAB (MathWorks, Natick, MA). To compare preference and precision outcomes between the knowledgeable cohort and the naive cohort (during the speed-matched condition of 1.2 m/s), we utilized two-tailed *t*-tests ($\alpha = .05$).

Using data collected from the naive cohort, we built eight linear mixed effects model (LMEM), one for each outcome. Each LMEM built in this study had the following form:

$$y = \mathbf{X}\beta + \mathbf{Z}b + \epsilon \tag{5.1}$$

where $y \in \mathbb{R}^{N\times 1}$ was the response vector of outcome measurements, and N was the number of total observations. $\mathbf{X} \in \mathbb{R}^{N\times m}$ was the fixed-effects design matrix and $\beta \in \mathbb{R}^{m\times 1}$ was the fixed-effects vector, where m was the number of fixed effects plus an intercept. $\mathbf{Z} \in \mathbb{R}^{N\times qK}$ was the random-effects design matrix, and $b \in \mathbb{R}^{qK\times 1}$ was the random-effects vector, where q was the number of random effects and K was the number of levels. $\epsilon \in \mathbb{R}^{N\times 1}$ was the error vector.

For all outcomes, the base model included speed and trial block as fixed effects ($m = 3$) and included a random intercept per subject ($q = 1$, $K = 12$). Starting at this base model, we optimized each model using a manual forward selection process based on the Akaike information criterion (AIC). We iteratively added random effects of subject-specific slopes for speed and trial block. When the difference in AIC was not significant between models with additional terms (as determined using MATLAB's `compare()` function), we chose the model with the fewer number of predictors. Through this process, we determined which combinations of fixed and random effects produced the model with the lowest AIC for each outcome. For the two precision outcomes, the best model was the base model. For the two preference outcomes and the four exploration strategy outcomes we selected the base model plus an additional subject-specific slope for trial block ($q = 2$, $K = 12$). For the number of strides per setting, adding an additional subject-specific slope for speed to the chosen model slightly improved the AIC from 762 to 758 ($p = .01$), but we opted for the simpler model to be consistent with the other exploration strategy outcomes.

The linear mixed effects models were constructed using data collected from twelve naive subjects who performed eight preference-identification trials at each of three speeds (24 trials per subject). For each preference outcome, we analyzed $N = 285$ observations. Please note this number is less than 288 because we were only able to analyze seven trials for Subjects 4 and 8 during the slow condition and Subject 6 during the nominal condition. For the exploration strategy outcomes, we analyzed the same $N = 285$ observations as with the preference outcomes. For each precision outcome, we analyzed $N = 36$ observations—the standard deviation across all trials at each of three speeds for twelve participants.

Each LMEM with two fixed effects produced a 2-D surface that best fit the outcome data, with walking speed on the *x*-axis and trial block on the *y*-axis (as depicted in Fig. 5.8). The surface was defined by the model coefficients for each fixed effect (*i.e.,* slopes), each with an associated *p*-value. If the LMEM *p*-value for a fixed effect coefficient was less than the significance threshold of .05, we considered that slope to be significantly different from zero, and thus concluded there was a significant effect of that fixed effect on the given outcome.

5.3 Results

Individuals Reliably Identified Unique Preferences in Characteristics of their Robotic Exoskeleton Assistance

Across all trials, naive subjects' preferred peak torque magnitude ranged from 7.94 Nm (95% confidence interval, $\text{CI}_{95\%} = [7.68, 8.19]$ Nm) to 19.42 Nm ($\text{CI}_{95\%} = [18.91, 19.92]$ Nm) (Fig. 5.7). As a metric of subjects' precision across trials, the mean within-subject standard deviation of preferred magnitude was 1.72 Nm, with a minimum of 0.60 Nm and a maximum of 2.92 Nm. As a metric of the variability between subjects' preferences, the between-subject standard deviation of preferred magnitude was 3.66 Nm (2.1 times the within-subject standard deviation). There was no correlation between preferred magnitude and body mass (Pearson's correlation coefficient, $r = .13$) or height ($r = .31$).

Naive subjects' minimum preferred timing was 54.08 % gait cycle ($CI_{95\%}$ = [53.32, 54.85] % gait cycle), and the maximum was 59.24 % gait cycle ($CI_{95\%}$ = [59.0, 59.47] % gait cycle) (Fig. 5.7). The mean within-subject standard deviation of preferred timing was 1.53 % gait cycle, with a minimum of 0.50 % gait cycle and a maximum of 2.82 % gait cycle. The between-subject standard deviation of preferred timing was 1.47 % gait cycle (0.96 times the within-subject standard deviation). There was no correlation between preferred timing and body mass (r = -0.21) or height (r = -0.08).

Naive participants sampled 21.5 (between-subject SD: 6.0) settings and explored for 104.9 (34.2) seconds before confirming their preference. On average, subjects took 4.3 (1.0) strides per setting and spent 4.8 (1.3) seconds per setting before changing to a new setting.

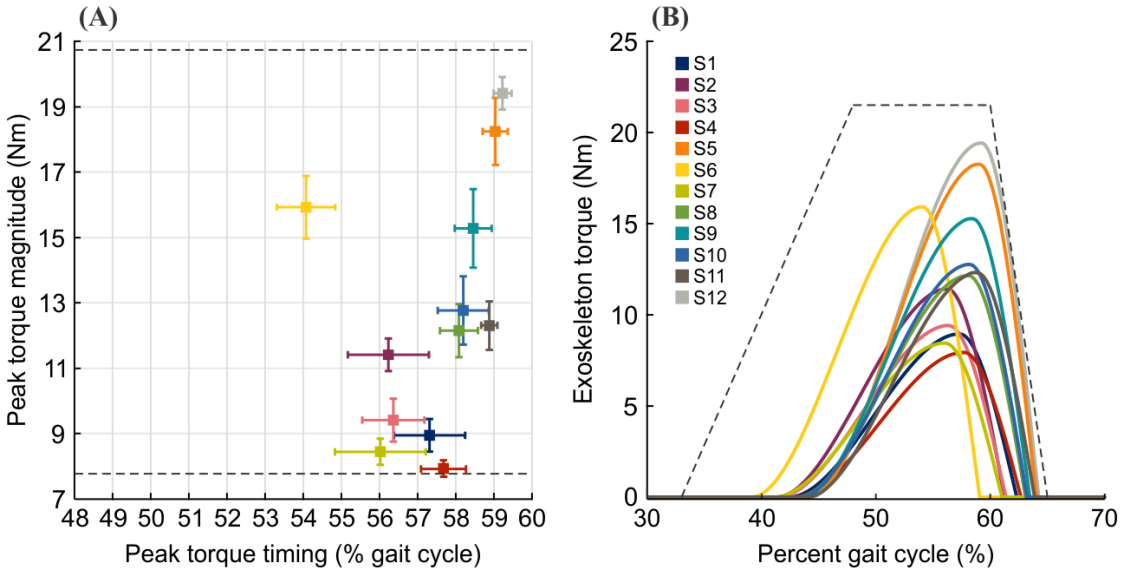

Figure 5.7: Naive subjects' preferred exoskeleton assistance. **(A)** Subjects' mean preferred magnitude and timing settings across all 24 preference-identification trials. Colored squares and error bars depict the mean and 95% confidence intervals for each parameter. The black dashed lines indicate the approximate bounds of torque magnitude individuals could explore using the touch screen. **(B)** Subjects' preferred torque profiles corresponding to their mean preferred magnitude and timing parameters. The black dashed lines encompass the approximate area that individuals could explore using the touch screen.

Individuals' Preferred Torque Magnitude Increased as the Experiment Progressed

There was a significant effect of trial block on preferred magnitude (LMEM, $p = .01$) (Fig. 5.8B). According to the slope of the model, we would expect a subject's preferred magnitude to increase by 1.54 Nm ($CI_{95\%} = [0.41, 2.66]$ Nm) from the first to the last trial block (Table 5.2). We additionally observed a mild positive relationship between preferred magnitude and walking speed, but it was not statistically significant (LMEM, $p = .06$) (Fig. 5.8A). Subjects' preferred timing did not vary with speed (LMEM, $p = .44$) (Fig. 5.9A), or trial block (LMEM, $p = .56$) (Fig. 5.9B).

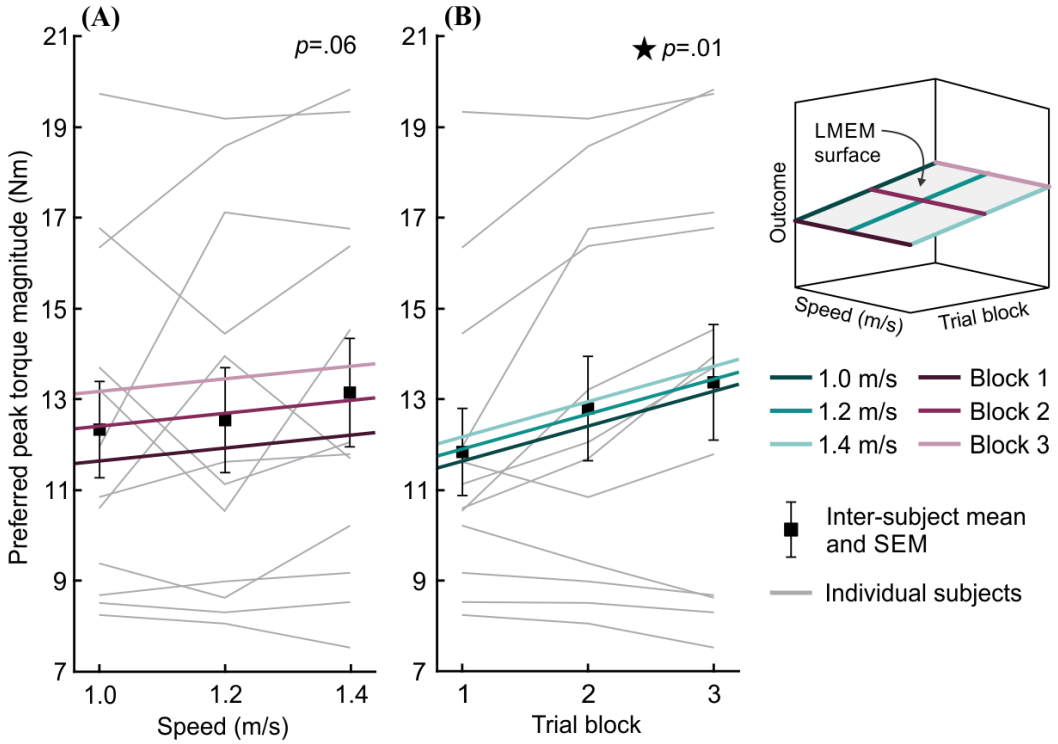

Figure 5.8: Preferred peak torque magnitude as a function of speed and trial block. **(A)** Preferred magnitude as a function of walking speed. **(B)** Preferred magnitude as a function of trial block. Black squares represent the inter-subject mean, and black error bars denote 1 standard error of the mean (SEM). Gray lines connect individual subjects' means. Each linear mixed effects model (LMEM) generates a surface that characterizes the relationship between the outcome and the two fixed effects (speed and trial block); a cartoon example of an LMEM surface is illustrated on the far right. Left and right panels are 2-dimensional views of the same LMEM surface. Purple lines (left) show the slope of the model with respect to walking speed and represent slices of the LMEM surface at trial block 1 (dark), trial block 2 (medium), and trial block 3 (light). Teal lines (right) show the slope of the model with respect to trial block and represent slices of the LMEM surface at 1.0 m/s (dark), 1.2 m/s (medium), and 1.4 m/s (light). The LMEM p-value for each fixed effect's slope is depicted at the top right of each panel; a significant effect is denoted with a black star.

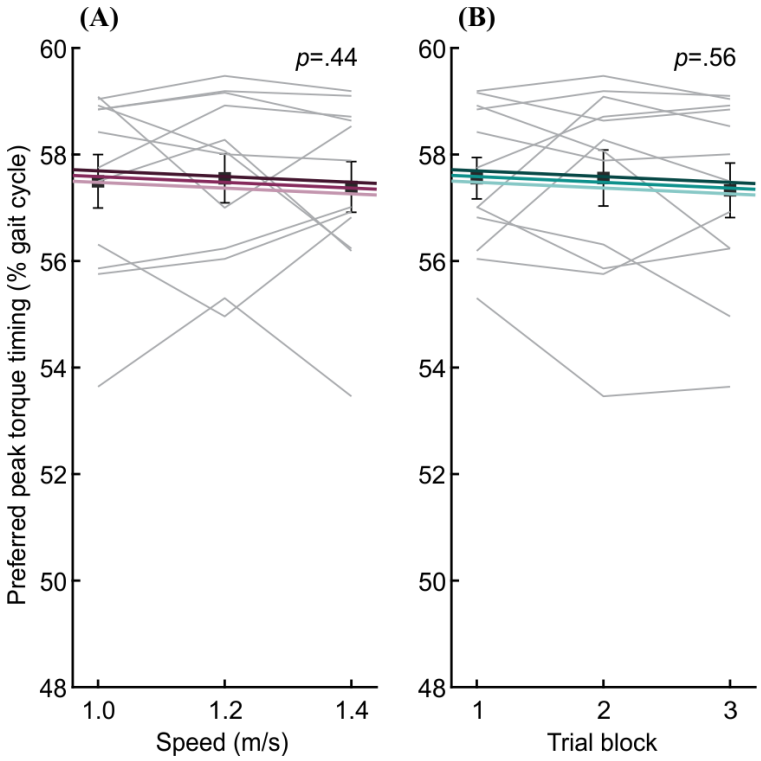

Figure 5.9: Preferred peak torque timing as a function of speed and trial block. **(A)** Preferred timing as a function of walking speed. **(B)** Preferred timing as a function of trial block. For a complete figure legend please refer to Fig. 5.8.

Individuals Were More Precise Identifying Their Preferred Torque Magnitude at Faster Walking Speeds

The standard deviation of subjects' preferred magnitude significantly decreased (*i.e.,* subjects became more precise) as walking speed increased (LMEM, $p = .01$) (Fig. 5.10A). According to the slope of the model, we would expect a subject's standard deviation of preferred magnitude to decrease by 0.50 Nm ($CI_{95\%} = [0.15, 0.86]$ Nm) as walking speed increased from 1.0 to 1.4 m/s (Table 5.2). With respect to trial block, the standard deviation of subjects' preferred magnitude exhibited a moderate decreasing trend, but the slope did not surpass the threshold for significance (LMEM, $p = .05$) (Fig. 5.10B). The standard deviation of preferred timing did not significantly vary with speed, but we observed a trend toward increased precision with increasing speed (LMEM, $p = .07$) (Fig. 5.10C). The standard deviation of preferred timing did not vary with trial block (LMEM, $p = .98$) (Fig. 5.10D).

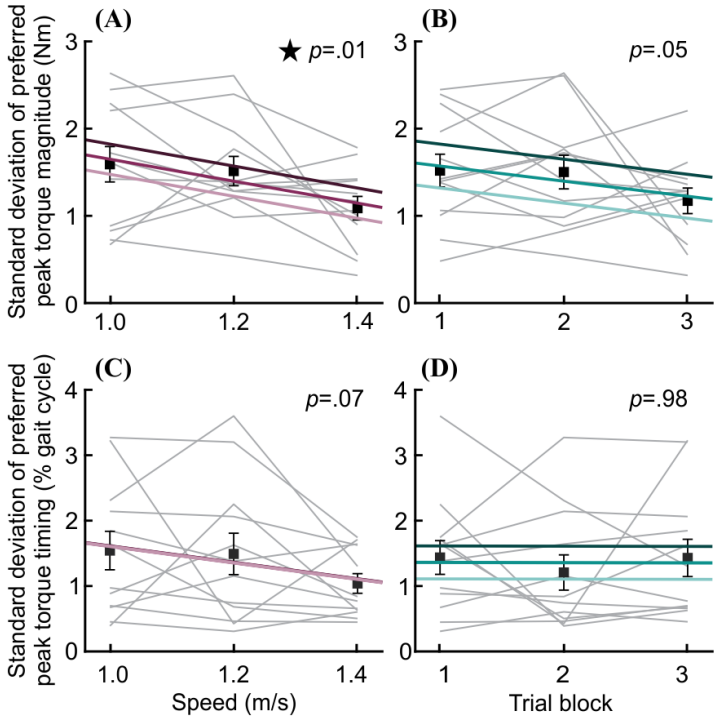

Figure 5.10: Precision outcomes as a function of speed and trial block. **(A)** Standard deviation of preferred magnitude as a function of walking speed. **(B)** Standard deviation of preferred magnitude as a function of trial block. **(C)** Standard deviation of preferred timing as a function of walking speed. **(D)** Standard deviation of preferred timing as a function of trial block. For a complete figure legend please refer to Fig. 5.8.

Individuals Became More Efficient Identifying Their Preferences as the Experiment Progressed

Trial block had a significant effect on all four exploration strategy outcomes (Table 5.2). The models predicted that from the first to the last trial block, subjects would explore 6.6 ($CI_{95\%}$ = [1.9, 11.3]) fewer settings (LMEM, $p = .01$) (Fig. 5.11B), reduce total exploration time by 57.4 seconds ($CI_{95\%}$ = [32.5, 82.2] seconds) (LMEM, $p < .0001$) (Fig. 5.11D), take 0.85 ($CI_{95\%}$ = [0.4, 1.3]) fewer strides per setting (LMEM, $p = .0002$) (Fig. 5.11F), and reduce time per setting by 1.0 seconds ($CI_{95\%}$ = [0.6, 1.5] seconds) (LMEM, $p < .0001$) (Fig. 5.11H).

Walking speed had a significant effect on three of the four exploration strategy outcomes (Table 5.2). The models predicted that increasing speed from 1.0 to 1.4 m/s corresponded to 3.0 ($CI_{95\%}$ = [0.4, 5.6]) fewer settings explored (LMEM, $p = .03$) (Fig. 5.11A), 15.3 seconds ($CI_{95\%}$ = [1.1, 29.4] seconds) less total exploration time (LMEM, $p = .03$) (Fig. 5.11C), and 0.6 ($CI_{95\%}$ =

[0.3, 0.9]) more strides per setting (LMEM, $p < .0001$) (Fig. 5.11E). Increasing speed did not affect the time spent per setting (LMEM, $p = .60$) (Fig. 5.11G).

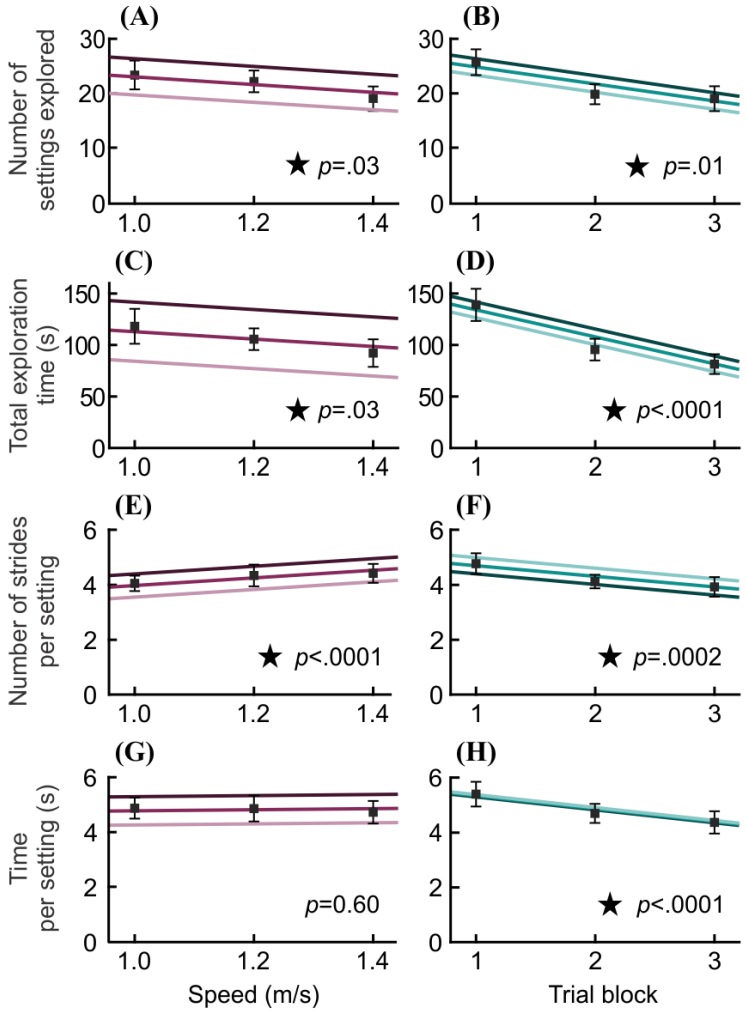

Figure 5.11: Exploration strategy outcomes as a function of speed and trial block. **(A)** Total number of settings explored as a function of walking speed. **(B)** Total number of settings explored as a function of trial block. **(C)** Total exploration time as a function of walking speed. **(D)** Total exploration time as a function of trial block. **(E)** Number of strides taken per setting as a function of walking speed. **(F)** Number of strides taken per setting as a function of trial block. **(G)** Time spent per setting as a function of walking speed. **(H)** Time spent per setting as a function of trial block. For a complete figure legend please refer to Fig. 5.8.

Individuals With Prior Exoskeleton Knowledge Preferred Higher Torque Magnitude

The knowledgeable cohort self-identified as researchers in the field of wearable robotics, yet not all knowledgeable subjects had prior experience walking using a powered exoskeleton. Of the twelve knowledgeable subjects, four reported they had never walked using a robotic exoskeleton before, five estimated 1–10 hours, one estimated 25 hours, and two estimated more than 100 hours (Table B.1). There was no correlation between hours of prior walking experience and preferred magnitude (Pearson's correlation coefficient, $r = .30$), or preferred timing ($r = .10$).

Knowledgeable subjects preferred significantly higher torque magnitude than naive subjects (t-test, $p = .02$, Cohen's $d = 0.95$) (Fig. 5.12A). We did not detect differences between groups in their preferred timing (t-test, $p = .71$, $d = 0.16$) (Fig. 5.12B), standard deviation of preferred magnitude (t-test, $p = .29$, $d = 0.44$) (Fig. 5.12C), or standard deviation of preferred timing (t-test, $p = .49$, $d = 0.29$) (Fig. 5.12D).

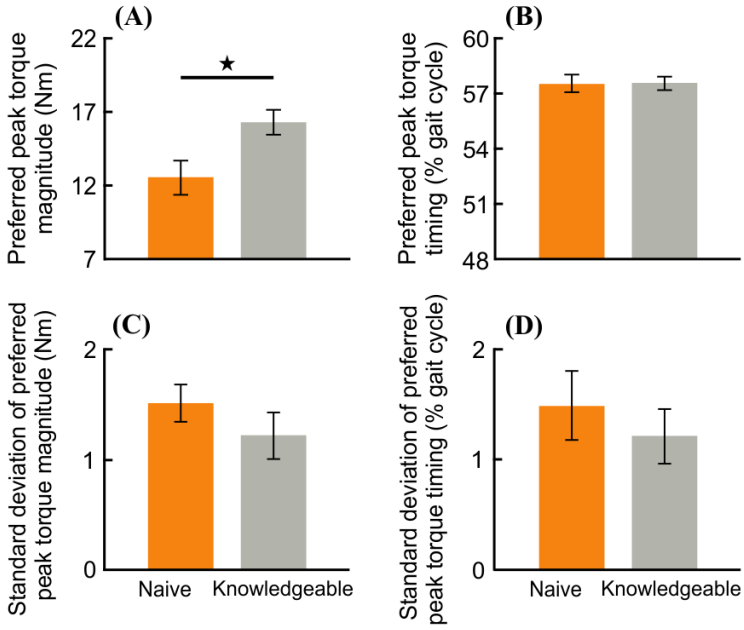

Figure 5.12: Comparison of outcomes between naive and knowledgeable users. **(A)** Preferred magnitude. **(B)** Preferred timing. **(C)** Standard deviation of preferred magnitude. **(D)** Standard deviation of preferred timing. Orange bars indicate the inter-subject mean of the outcome for naive subjects walking at 1.2 m/s. Gray bars indicate the inter-subject mean of the outcome for knowledgeable subjects walking at 1.2 m/s. Black error bars show 1 standard error of the mean (SEM). A black bracket with a star indicates a significant difference between the means, as calculated by a two-tailed t-test.

		Fixed effect = speed (tested range 1-1.4 m/s)			Fixed effect = trial block (tested range 1-3)		
	Outcome	LMEM slope estimate [95% CI]	*p*-value of slope	Expected change over tested range (0.4 m/s)	LMEM slope estimate [95% CI]	*p*-value of slope	Expected change over tested range (2 blocks)
Preference outcomes	Preferred magnitude (Nm)	1.33 [-0.07, 2.72]	.06	0.53	**0.77 [0.21, 1.33]**	**.01**	**1.54**
	Preferred timing (% gait cycle)	-0.54 [-1.90, 0.82]	.44	-0.21	-0.11 [-0.48, 0.26]	.56	-0.22
Precision outcomes	Std. dev. of preferred magnitude (Nm)	**-1.26 [-2.14, -0.38]**	**.01**	**-0.50**	-0.17 [-0.35, 0.003]	.05	-0.35
	Std. dev. of preferred timing (% gait cycle)	-1.26 [-2.62, 0.1]	.07	-0.50	0 [-0.28, 0.27]	.98	-0.01
Exploration strategy outcomes	Number of settings explored (# settings)	**-7.51 [-14.07, -0.94]**	**.03**	**-3.00**	**-3.30 [-5.67, -0.93]**	**.01**	**-6.60**
	Exploration time (s)	**-38.17 [-73.57, -2.78]**	**.03**	**-15.30**	**-28.68 [-41.11, -16.26]**	**< .0001**	**-57.36**
	Number of strides per setting (# strides)	**1.48 [0.76, 2.20]**	**< .0001**	**0.60**	**-0.42 [-0.64, -0.20]**	**.0002**	**-0.85**
	Time per setting (s)	0.21 [-0.58, 1.00]	.60	0.10	**-0.51 [-0.75, -0.28]**	**< .0001**	**-1.03**

Bold text indicates a significant slope with $p < .05$.

Table 5.2: Results from linear mixed effects models (LMEM) for each outcome with speed and trial block as fixed effects.

5.4 Discussion

Participants in this study identified unique preferences (Fig. 5.7B), which corroborates mounting scientific evidence that optimal exoskeleton assistance is subject-specific and requires individualized tuning. We observed that subjects' magnitude preferences were distributed over the range of explorable settings (Fig. 5.7A), illuminating several important factors related to the preferred characteristics of ankle exoskeleton assistance. First, most subjects preferred to feel the exoskeleton assisting them to some degree and did not tune the controller to eliminate the assistive torque. Similarly, individuals did not repeatedly seek the maximum device assistance, which indicates that for this device, the range of exoskeleton torque provided in this study was inclusive of most subjects' preferences. Finally, subjects' preferred magnitude did not scale with body mass or height, suggesting that the choices were based on internalized criteria and not determined by anthropometric constraints. Unlike magnitude, subjects' timing preferences fell closely together, within a narrow band of the explorable grid (Fig. 5.7A). Most subjects' preferences were concentrated between 56–59 % gait cycle. The 3 % gait cycle range that encompassed most subjects' preferences corresponds to a 33 ms time window, or only 6 % of stance phase. These results demonstrate that

users have a unique internalized representation of how they wish to receive exoskeleton assistance.

In this study, individuals precisely identified their preferences in two dimensions simultaneously, thus demonstrating the efficacy of a two-dimensional self-tuning paradigm for preference-based exoskeleton control design. Moreover, as subjects were blinded to the parameters and axes of the gird, this shows that individuals can successfully navigate a two-dimensional interface to identify their preference relying solely on their perception of device assistance. As a metric for precision (*i.e.,* repeatability), we analyzed the standard deviation of a user's preferences across repeated trials. Across naive subjects, the mean standard deviation of preferred timing was 1.53 % gait cycle, which corresponds to just 17 ms for the average stride time observed in this study. The mean standard deviation of preferred magnitude was 1.47 Nm, which represents 1.5 % of the average peak biological ankle torque observed during walking [21]. For comparison to existing literature, it is helpful to consider the coefficient of variation (CV) for preferred magnitude, which represents the subject's standard deviation of preferences as a percentage of their mean. Across naive subjects, the mean (SD) CV was 13.5% (4.1%); this is comparable to the CV for individuals using a one-dimensional dial to self-tune their prosthetic foot stiffness, which was 14.2% (1.7%) [155]. While it is difficult to directly compare the CV between studies because of the different hardware and parameters of interest, these results suggest that adding an additional dimension to a self-tuning interface may not diminish subjects' precision. As subjects were blinded to both the parameters of interest and the axes were randomized between trials, these results may represent the lower bound of subjects' precision.

Subjects were also able to identify their preferred settings quickly. On average, naive users found their preferred settings in only 105 seconds (1.8 minutes). Even if subjects needed to perform 3–5 repeated trials to obtain an average preference, the entire tuning procedure would take between five and ten minutes. This is faster than the pairwise comparison method employed in [173], which required 20 trials for the algorithm to build an approximation of the subject's preference in one dimension. It is also considerably faster than existing human-in-the-loop optimization routines, which take approximately 20–60 minutes to converge on an individual's energetically op-

timal settings in three or four dimensions [35, 189]. Taken together, the speed and precision of subjects' preference identification demonstrate that users are a reliable source of information and that the two-dimensional self-tuning paradigm may be a valuable and efficient method for designing preference-based exoskeleton control systems.

Participants in this study did not prefer exoskeleton assistance that replicated trends in the biological ankle's torque production. We observed only a mild relationship ($p = .06$) between preferred magnitude and walking speed (Fig. 5.8A), and the effect was small; our analyses predicted that preferred magnitude may increase by 0.53 Nm ($CI_{95\%}$ = [-0.03, 1.09] Nm) as walking speed increases from 1.0 to 1.4 $^{m}/_{s}$ (Table 5.2). From the literature, we estimate that the biological ankle's peak torque production increases by approximately 0.12 $^{Nm}/_{kg}$ (8.4 Nm for a 70 kg human) as speed increases from 0.9 to 1.45 $^{m}/_{s}$ [21]. Therefore, the increase in preferred peak torque magnitude observed in this study represents only 6% of the expected increase in biological ankle torque over a similar range of speeds. With regards to the timing of peak exoskeleton assistance, the peak timing that individuals preferred in this study (56–59 % gait cycle) is considerably later in the gait cycle than the peak of the biological ankle torque production, which falls in the range of (45–49 % gait cycle) [21, 182]. Preferred timing also remained constant with respect to speed (Fig. 5.9A). This is in contrast to a recently published dataset that shows the timing of peak biological torque production occurs approximately 2 % gait cycle earlier as walking speed increases over a similar range [21]. While an intuitive starting place for designing ankle exoskeleton assistance is to supplement the biological ankle by replicating its mechanics during walking [119, 129], the observations from our study indicate that this strategy may not be preferred by users. These findings suggest that there are complex and potentially counterintuitive mechanisms driving user preference and future studies are required to elucidate these elements.

One of our key findings is that user preference is not a static quantity. In this study, naive users preferred 1.54 Nm higher torque magnitude after they interacted with the device throughout the duration of the experiment (Fig. 5.8B). There are numerous potential explanations for this result, and we elaborate here on only some of the possibilities. Since our subjects were healthy young adults

and walked for just 22–77 cumulative minutes (similar to [35, 189]), with breaks every 1–3 minutes, we strongly believe the effect is not a result of fatigue. Another possibility is that individuals discovered strategies to adapt their gait biomechanics to comfortably accept higher exoskeleton assistance. As one example, if users were initially de-stabilized by the exoskeleton assistance (as proposed in [124] and [5]) and thus preferred lower torque, they may have subsequently discovered gait strategies to accommodate higher torques while maintaining dynamic stability. Finally, it is possible that a naive user's neuromotor system requires a certain duration of time and/or practice to adapt to exoskeleton assistance, and this adaptation manifests as changes in preferred torque magnitude. Support for this explanation may be found in previous studies that demonstrated that biomechanical quantities (*e.g.,* ankle kinematics, muscle activity) required 20–30 minutes to reach steady state in response to powered exoskeleton assistance [20, 57]. However, Cain *et al.* also concluded that users' neuromotor adaptation changes depending on the control strategy employed, so we cannot definitively extrapolate these results to the exoskeleton and control strategy used in our experiment. To our knowledge, this is the first study to systematically evaluate user preferences over multiple walking conditions, so the timescale for preference adaptation has yet to be determined. An additional study will be required to assess if (or when) a new user's preference reaches steady state. The current study demonstrates that users' preferred torque magnitude increases by 1.54 Nm over the course of 24 self-tuning trials (or 22–77 minutes of active exoskeleton use). These data may be used as a benchmark to determine how often a user might need to re-tune a preference-based exoskeleton controller, or, alternatively, as motivation to design controllers with user preference continuously evaluated in the loop.

As the experiment progressed, in addition to preferring higher torque magnitude, subjects also became more precise and efficient identifying their preferred magnitude. Examining these changes in the users' interaction with the self-tuning interface elucidates important elements related to how naive exoskeleton users learn to navigate the grid and identify their preferences. We observed a moderate downward trend in subjects' standard deviation of preferred magnitude (*i.e.,* increased precision) with respect to trial block ($p = .05$) (Fig. 5.10B). Simultaneously, subjects' efficiency

increased as trial block increased, as quantified by all four exploration strategy metrics (Fig. 5.11, right column). From a practical standpoint, participants likely developed strategies to navigate the grid interface more effectively with practice. Anecdotally, we observed that many subjects streamlined their exploration strategy as the experiment progressed and converged to a 'quadrant-based' strategy; participants would sample 1–2 settings from each quadrant of the grid, select the quadrant that felt best, and then fine-tune within the chosen quadrant (Fig. 5.13). Our models revealed that, over the course of the experiment, subjects would sample 6.6 fewer settings and explore for 57 fewer seconds before confirming their preference (Table 5.2), and we believe it is reasonable to attribute these results to subjects adopting a more efficient exploration strategy. However, subjects' increased precision and efficiency in identifying their preferences could also point to adaptation in the subjects' perception of the exoskeleton assistance. Subjects' increased precision identifying their preferred magnitude may suggest that their ability to differentiate between similar magnitude settings improved, or that they developed a more specific internal definition of their preference. An increased perceptual awareness may also be illustrated by examining the average number of strides individuals took at a given setting before selecting a new setting. Our analyses demonstrate that we would expect subjects to take 0.85 fewer strides (and spend 1 less second) per setting in the third trial block compared to the first (Table 5.2). Therefore, as the experiment progressed, subjects were able to make the same determination (*i.e.,* "how does this feel, and do I like it?") at a particular setting more quickly, and in approximately one fewer stride. At this time, we have not quantified the human's perceptual ability to distinguish between characteristics of exoskeleton assistance (*i.e.,* the just noticeable difference), but we believe this would be a valuable avenue for future research and aid in the interpretation of these results. From a control design perspective, these findings demonstrate that subjects become an even more reliable source of data as they gain experience using the device. With increased exposure, users become more confident in searching for, and deciding upon, their preferred exoskeleton assistance.

An unexpected finding was that as walking speed increased, subjects were more precise in identifying their preferred magnitude (Fig. 5.10A). Based on the model, we would expect sub-

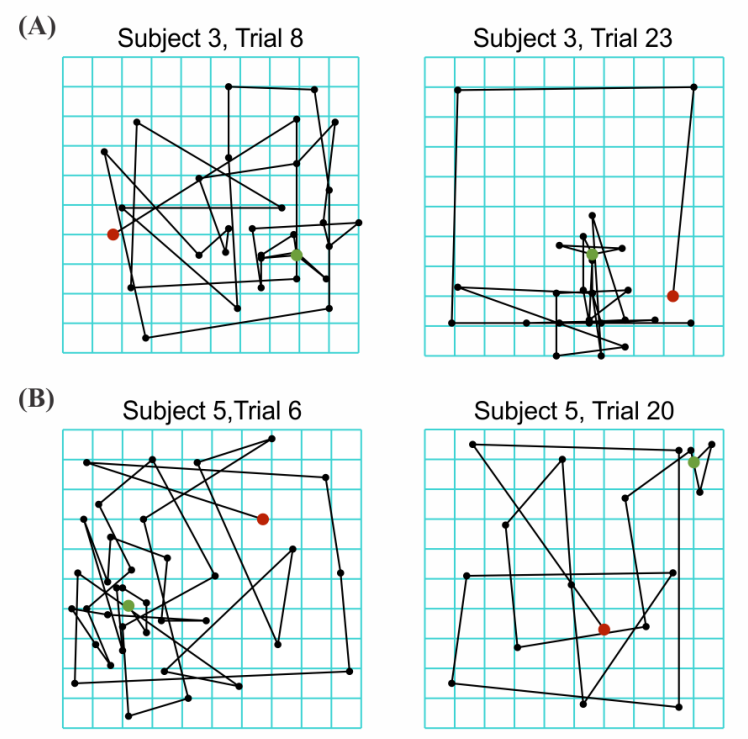

Figure 5.13: Evolution of exploration strategies. As the experiment progressed, many subjects adopted more efficient 'quadrant-based' exploration strategies, as shown here for two subjects, **(A)** Subject 3 and **(B)** Subject 5. Grids in the left column depict representative trials from the first trial block. Grids in the right column depict representative trials from the third trial block. Subjects began at a randomized initial condition (red circle) and explored the grid until they decided upon their preference (green circle). Black circles indicate the settings the participant sampled. Black lines connect settings in the order they were explored; these lines were added for visualization and did not appear on the subject's screen.

jects' standard deviation of preferred magnitude to decrease by 32 % (0.5 Nm) as speed increases from 1.0–1.4 m/s (Table 5.2). We additionally observed a moderate downward trend in the standard deviation of subjects' preferred timing ($p = .07$) (Fig. 5.10C). These results are intriguing, especially considering that participants also became more efficient as speed increased—our analyses predict that, at faster speeds, we would expect subjects to explore 3 fewer total settings and spend 15 seconds less time exploring (Figs. 5.11A, 5.11C) (Table 5.2). As the order the speeds a subject performed was block-randomized (and evenly distributed within the subject pool), we do not believe these are order effects. One explanation for the increased precision is that subjects simply took more strides at faster speeds and therefore had increased exposure to the exoskeleton assistance. Subjects did, in fact, take slightly more strides per setting at faster speeds (Fig. 5.11E), but

did not spend more time per setting (Fig. 5.11G), which indicates that the increased stride count per setting is likely a result of faster treadmill belt speed. However, subjects did not take more total strides per trial as speed increased. In fact, the average total number of strides taken per trial decreased monotonically with increasing walking speed, which also corresponds to the decreased total exploration time. Therefore, it does not seem plausible that subjects' increased precision with increased walking speed is attributable to more time spent walking or more total strides taken at faster speeds. An interesting hypothesis to explain these results is that an individual's "preference gradient" is steeper at faster walking speeds. In essence, this would mean that subjects are more sensitive to the settings that feel good to them, and therefore select a more concentrated set of preferred settings. It could be that, at faster walking speeds, subjects feel less stable or less comfortable at settings farther away from their preference. Alternatively, if individuals were more sensitive to the *power* delivered by the exoskeleton than the *torque*, they may be more precise in identifying their preferences in conditions with higher exoskeleton power (*i.e.,* the same torque burst applied in concert with faster ankle angular velocity [117]). The observed increase in precision at faster walking speeds is surprising because the assistive torque burst (with a fixed width of 20 % gait cycle) occurs over a shorter duration of time as walking speed increases and stride time decreases. This suggests that users are not only sensing the torque burst itself to decide on their preference, but that they are actually evaluating the impact of the assistance over the entire gait cycle.

As discussed so far, user satisfaction and comfort are likely key elements that will affect the adoption of augmentation-focused robotic exoskeleton technology in the future. Another critical piece, however, may be user education surrounding the potential benefits of such devices. In this study, we observed that preferences differed between knowledgeable subjects (*i.e.,* researchers in the field of wearable robotics), and naive subjects. During the nominal speed condition, the knowledgeable cohort preferred 30 % (3.8 Nm) higher peak torque magnitude than the naive cohort (Fig. 5.12A). Naive and knowledgeable users were both read an identical script at the beginning of the experiment, so initial subject education did not bias our results. It is interesting that not

all members of the knowledgeable cohort actually had prior experience walking using a powered exoskeleton, and there was no correlation between the previous number of hours spent walking using a powered exoskeleton and preferred assistance settings. Therefore, we would not expect the knowledgeable users' preferences to represent an asymptote of the naive users' preferences as they gained walking experience. Rather, it appears that knowledgeable subjects' personal experiences as researchers in the field primed them to prefer higher assistance from the robotic exoskeleton. One possibility is that knowledgeable users were more aware of the potential benefits of robotic assistance (*e.g.,* energy cost reduction) and thus preferred higher magnitudes. These findings speak to the importance of educating new users about the potential benefits of exoskeleton technology, as this knowledge may change the way the user interacts with the device and the assistance they prefer. For exoskeleton researchers, the preference differences between the knowledge and naive cohorts also expose the critical need to expand our subject recruitment pools beyond members of our lab or research community—if we only perform experiments on knowledgeable individuals, we will likely miss out on important data that will inform how naive users learn to use exoskeletons for able-bodied augmentation.

In this study, we were limited to providing approximately 22 Nm of exoskeleton torque, which was the maximum our controller could reliably provide step-to-step without saturating the current limits of the device. Compared to similar studies (*e.g.,* [119, 189]), this is a relatively low amount of torque (equivalent to roughly 0.31 $^{\text{Nm}}/_{\text{kg}}$ for a 70 kg subject). As discussed earlier, however, the maximum torque limit was high enough to capture all the naive subjects' preferences. Because we were limited by the total maximum torque we could provide, we chose to allow subjects to self-tune their exoskeleton magnitude in Nm, as opposed to the more common body-mass-normalized $^{\text{Nm}}/_{\text{kg}}$. This choice ensured that all subjects could explore roughly the same limits of the grid, and we did not need to reduce the maximum limit for heavier subjects. We performed the same analyses presented in this study using subjects' magnitude preferences normalized by body mass (post-processed), and this did not impact the interpretation of any of our observed results (Appendix B). For comparisons between naive and knowledgeable cohorts, we designed this experiment to com-

pare knowledgeable subjects' preferences (walking at nominal speed) to naive subjects' preferences as they performed the speed-matched condition. Alternatively, we could have compared the knowledgeable subjects' preferences (as they performed eight trials) to the naive subjects' preferences as they performed the first trial block (*i.e.,* trials 1–8). We performed this additional analysis, and it did not change the observed differences between groups (Appendix B). In this study, subjects only performed one, self-directed practice session (approximately 5 minutes long) before performing the experimental trials. We intentionally chose naive subjects for this experiment to assess how preferences change as individuals learn to use the exoskeleton, but at this time, we cannot draw conclusions about how the results might be different if these individuals had been given longer familiarization periods. The demographics of our naive cohort were well-distributed for gender (6 men, 6 women), and age (19–51 years), but our study was limited by homogeneity in subjects' body mass index (BMI: 22–30). The demographics of the knowledgeable cohort were less well-distributed than the naive cohort in gender (9 men, 3 women) and age (22–33 years) and were similarly distributed for BMI (15–31).

In this study, we demonstrated that users have unique preferences in their robotic ankle exoskeleton assistance, and that they can precisely and quickly identify these preferences using a two-dimensional self-tuning interface. Because users were blinded to the parameters they were tuning, this novel paradigm encouraged participants to customize the device assistance using their own perception and internalized representation of optimal device assistance. This study is the first of its kind to rigorously analyze the characteristics of user preference, including how preference changes with walking speed, time, and exposure to the device, as well as how preferences differ between groups with different levels of prior technical knowledge and expertise. In conclusion, the results from this study demonstrate that robotic exoskeleton users are a rich and reliable source of information related to their experience wearing the device. We advocate for the incorporation of user preference into the design of future control systems in order to promote synergistic human-robot interaction and accelerate the translation of augmentative exoskeleton systems.

CHAPTER 6

Conclusions and Future Directions

As a whole, this dissertation evaluates the notion of optimality in the control of wearable robotic systems and lays the foundation for translating optimization protocols outside the laboratory environment. The work described in the preceding chapters explored different cost functions for the optimization of wearable robots and presented robust measurement tools to obtain accurate estimates of necessary physiological and subjective quantities. This dissertation presented the results of three full experiments conducted with human subjects, and also contributed detailed scientific evaluation regarding how interaction with wearable robotic systems affects individuals' energetics, biomechanics, and subjective preferences. By addressing key challenges in the design and assessment of optimal controllers for wearable robots, this research advances the field closer to the translation of these devices outside the laboratory environment.

In the context of the generalized framework for the optimization and evaluation of wearable robots I presented in Chapter 1 (Fig. 1.1), each chapter of this dissertation contributed to specific phases of the pipeline (Fig. 6.1). In Chapter 2, I investigated how changing the *design* of the BiOM's controller (*i.e.,* modifying the power setting parameter) impacts individuals wearing the device. This experiment was designed as a parameter study, and this structure allowed us to *evaluate* the relationship between the BiOM's power setting and the wearer's energetics by constructing full cost landscapes for each individual. In Chapters 3 and 4, I described a new way to *measure* metabolic energy expenditure using physiological signals collected from portable wearable sensors. This work was targeted towards predicting instantaneous metabolic energy cost outside the

laboratory environment for body-in-the-loop optimization algorithms, which iteratively update the *design* of the controller in response to real-time measurements of energy cost. And finally, in Chapter 5, I assessed a novel *objective* for the optimization of wearable robots—user preference. I developed a touch-screen tablet interface to *measure* individuals' preferences as they self-tuned two parameters of the assistive torque provided by bilateral ankle exoskeletons. The work presented in this dissertation contributes to each phase of the optimization and evaluation pipeline, and paves the way for future research advances in each of these areas.

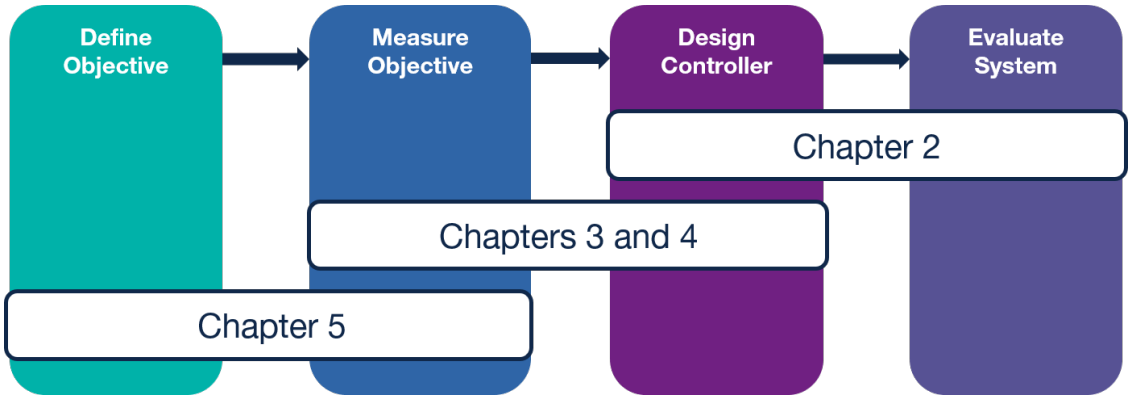

Figure 6.1: Contributions of this dissertation in the context of the general optimization framework. Each chapter of this dissertation contributes to individual phases of the optimization and evaluation framework. This framework is described in detail in Chapter 1 (Fig. 1.1).

6.1 Discussion of Contributions

The following sections summarize the most significant contributions from each of the previous chapters, as well as describe how the work fills a gap in the field's current body of knowledge.

Chapter 2: Significance and Contributions

In Chapter 2, I described the experimental study we conducted that investigated the relationship between power setting and the cost of transport (COT) of individuals with transtibial amputation walking wearing the BiOM powered ankle prosthesis. This work has been published in *Scientific Reports* [74]. At the time of publication, previous studies had investigated how metabolic cost and

walking mechanics differed when wearing the BiOM (tuned by an expert) compared to wearing a passive elastic foot (*e.g.,* [52, 63, 145]). Previous studies had also characterized the relationship between prosthetic net ankle work and metabolic cost, but used a tethered prosthetic ankle emulator system, which can not fully replicate the experience of wearing a powered device [23, 131]. Therefore, this study filled a gap that existed in the literature; to our knowledge, this was the first parameter study to be conducted using the clinically and commercially available BiOM ankle. All around the country, prosthetists are actively prescribing and fitting patients with this device, yet there is very little information regarding how their tuning choices affect the patients' biomechanics, physiology, and experience wearing the BiOM.

This study directly informs clinical tuning and prescription practices in two main ways. The first contribution of this work is that it demonstrates that changing the BiOM power setting impacts individuals' cost of transport. On average, subjects exhibited a lower COT at higher power settings, yet this relationship varied considerably between individual subjects. The second contribution of this work is that it shows that, in order to minimize their COT, all individuals required *higher* power settings than the setting chosen by the prosthetist (*i.e.,* the setting that approximated biological ankle kinetics). While replicating biological ankle kinetics is a reasonable place to start for the control of a powered prosthetic ankle, our results show that this strategy does not produce the prosthetic ankle mechanics necessary to reduce or minimize a user's energy expenditure. The clinical tuning of powered prostheses is a complicated and time-consuming process that relies heavily on the expertise of the prosthetist and the communication between patient and clinician. For this reason, the field of wearable robotics is moving away from expert tuning and towards more objective, physiologically-informed, or human-in-the-loop tuning practices. As this exciting research area is relatively new, there are still many open questions related to how to choose the 'best' tuning objective for a particular device or individual, and relevant candidate objectives must be carefully considered. In this study, we evaluate only one possibility—minimizing cost of transport. This objective has clinical significance because individuals with amputation expend more energy during walking than non-amputees, and therefore it is plausible that researchers and/or prosthetists

may wish to tune a powered device to reduce this energetic burden. The work described in this Chapter 2 presents several practical takeaways to inform the design of control systems for powered ankle prostheses that reduce or minimize an individual's cost of transport.

Chapters 3 and 4: Significance and Contributions

In Chapters 3 and 4, I described a method to estimate instantaneous metabolic cost using physiological signals collected from portable wearable sensors. Chapter 3 has been published in the *Journal of Applied Physiology* [77], and Chapter 4 has been published in the *IEEE Robotics and Automation Magazine* [81]. In itself, the idea to estimate energy expenditure from wearable sensors is not new—researchers and scientists from the fields of physiology, exercise science, clinical nutrition, and pediatric medicine have been developing such algorithms for decades. However, estimation algorithms are discipline-specific, and thus have focused on predicting different forms of energy expenditure, depending on the goal. For example, prior studies used black-box algorithms to predict the metabolic equivalent of a task (MET) (*e.g.,* [19, 32]) or the breath-by-breath response of energy expenditure or oxygen uptake (*e.g.,* [3, 13, 143]). To our knowledge, the two studies presented here in Chapters 3 and 4 were the first to use physiological signals to predict *instantaneous* or *steady-state* metabolic cost. We chose to estimate instantaneous metabolic cost because it is the quantity that is required for body-in-the-loop or human-in-the-loop optimization algorithms that minimize energy expenditure. There is a major gap to be filled in the development of body-in-the-loop optimization algorithms because current techniques are fundamentally limited by their reliance on indirect calorimetry for estimates of instantaneous metabolic cost. In addition to the data-related challenges, indirect calorimetry is unsuitable for long-term data collections outside the laboratory environment because of the obtrusive rubber mask. Therefore, by estimating instantaneous metabolic cost without the use of the mask, this work represents the first step towards bridging the gap between human-in-the-loop optimization performed in the confined laboratory setting and translating these algorithms into real-world situations.

In Chapter 3, I described the experiment I performed to collect the necessary dataset for build-

ing predictive algorithms, which is a major contribution of this work. In designing this experiment, we took great care to choose combinations of tasks that would give any predictive algorithm the best chance at predicting metabolic energy expenditure, and not an intermediate biomechanical or physiological correlate. For example, in a dataset that contains only walking at different speeds, we know that walking speed is highly correlated with metabolic energy expenditure. Given that many of the sensors (*e.g.,* accelerometers) are designed to measure changes in walking speed or kinematics, a predictive algorithm may erroneously associate such patterns with actual changes in energy cost. For this reason, we chose to include tasks such as backwards walking in our dataset, which have a similar metabolic cost to forward walking, but very distinct kinematics. Conversely, we chose to include tasks such as incline walking, which have similar kinematics to level ground walking, but higher energetic demands. The cycling task (performed on a stationary exercise bike) also presented an opportunity to include redundancy in the dataset. We collected data from individuals cycling at 70 rpm, but under three different resistance conditions, which means that the kinematics of the task were identical, but the energetic demands changed. I believe that the careful consideration of tasks and built-in redundancy (coupled with the number of physiological signals included) make this dataset a unique and valuable resource to the research community. For this reason, the dataset is available open-source [78], and it is my hope that other researchers will utilize it to build their own predictive algorithms for a variety of applications.

Chapter 3 also presents the systematic evaluation of the 'salience', or predictive capability, of physiological signals used in a linear regression algorithm to predict instantaneous metabolic cost. In this chapter, I documented the detailed analyses comparing the predictive capability of raw vs. processed signals, filtered vs. unfiltered signals, and individual vs. groups of signals. For this study, we prioritized simplicity over optimal performance and used linear regression as a common algorithm to compare predictive power across signals. A key contribution of this work is the evaluation of signal performance using leave-one-subject-out or leave-one-task-out cross validation. While an 80%–20% or 90%–10% cross validation is a common approach for validating machine learning algorithms, we specifically chose leave-one-subject-out and leave-one-task-out

validations because of their potential to inform body-in-the-loop optimization methods. Wearing adaptive assistive robotic devices might very well constitute an unknown task and might lead to changes in an individual subject's physiology, yet we still need to obtain accurate estimates of metabolic cost. Thus, for body-in-the-loop optimization algorithms to be successful, we need to incorporate signals that are robust to potentially "unknown" tasks and subjects. In this work, we found that accelerometer signals were particularly affected by unknown tasks, while EMG and global signals (*e.g.,* heart rate, electrodermal activity) were more sensitive to unknown subjects. Another contribution of this work is the evaluation of the 'best' signals to include in a multiple linear regression model when only a small number of signals (fewer than 8) are included. This analysis is crucial for the practical implementation of these algorithms because it is unlikely that researchers will have access to or wish to use the large number of sensors included in this study. In fact, our results demonstrated that with only one signal (minute ventilation) we could achieve a reasonably low prediction error. Adding additional sensors improved performance, but with diminishing return. Based on the results of this study, for body-in-the-loop optimization and other applications that require predicting energy cost from portable sensors, we recommend combining accelerometers (ideally located close to the center of mass) with heart rate, breath frequency, and/or minute ventilation to provide the most robust prediction of instantaneous energy cost.

The work presented in Chapter 4 builds upon the work presented in Chapter 3 by improving both the *speed* and *accuracy* of predicting instantaneous metabolic cost from physiological signals collected from portable sensors. This work presented in this chapter is significant to the optimization of wearable robots because the faster we can obtain accurate estimates of instantaneous metabolic cost, the more quickly we can iterate through control parameters in a human-in-the-loop optimization routine. The work described in Chapter 3 showed that filtered signals like heart rate, minute ventilation, and limb accelerations were highly correlated with instantaneous metabolic cost. However, we noticed that using dynamically delayed input signals (*e.g.,* heart rate, minute ventilation) in a linear regression algorithm produced dynamically delayed predictions of instantaneous energy cost. To improve the speed of instantaneous metabolic cost predictions using wear-

able sensors, we included the time derivatives of physiological signals the black-box estimation algorithm. We expected that this information would be useful because a linear regression performed on a signal and its derivative has the same mathematical structure as an explicit first-order dynamic model. One of the contributions of this work is that we demonstrated that including derivatives in the linear regression models improved both the speed and accuracy of predicting instantaneous energetic cost, especially when the input signals were dynamically delayed. Furthermore, the signal derivative information can be included essentially "for free", because it requires no additional sensor hardware and minimal signal processing, yet it can enhance the predictive capability of the regression model. An additional contribution of this work is the mathematical framework we developed that allows us to compare the coefficients of a regression model with those identified from an explicit first-order model. This analysis showed that, in the absence of noise, the coefficients of the linear regression model are the same as the explicit model, but in the presence of noise, the coefficients differ. At this time, it is very difficult to make a one-to-one comparison between the predictive performance of a regression model and an explicit model, but expanding the mathematical framework to systematically compare the predictive performance of explicit dynamic models and regression models is a worthwhile and valuable avenue for future work.

Chapter 5: Significance and Contributions

In Chapter 5, I presented a study that evaluated user preference as a holistic objective for the control of robotic exoskeletons. At the time of this writing, the pilot experiment I conducted has been published in the conference proceedings for the *2020 IEEE Conference on Biomedical Robotics & Biomechatronics* [80], and the manuscript presented in Chapter 5 is currently under review. To date, metabolic cost has been the near-universal objective for the optimization and tuning of control systems, particularly for augmentative robotic exoskeletons [151]. Human-in-the-loop optimization methods have emerged as promising techniques to automatically tune the parameters of a control system to minimize the user's energy cost (*e.g.,* [35, 92, 184, 189]). However, current human-in-the-loop optimization methods are fundamentally limited due to their reliance on one,

predefined, measurable physiological objective as a metric of device efficacy. In reality, it is likely that users prioritize many different metrics simultaneously while wearing a robotic exoskeleton, such as comfort, stability, pain, symmetry, or perceived effort. This observation brings to light two gaps in our current knowledge: 1) how do we define a cost function that has relevance during all activities the user performs during their daily lives? and 2) how do we robustly measure the relevant user-specific and subjective quantities? As a starting place to fill these gaps, researchers have recently begun to investigate *user preference* as an alternative metric for the control of wearable robots, because it may encode the multi-factorial nature of exoskeleton use. Particularly, prior work in lower-limb prostheses established robust self-tuning methods for determining a patient's preferred prosthetic characteristics using a one-dimensional dial [28, 156] or three dials that were adjusted independently [22].

The work presented in Chapter 5 advances this foundational work in three key ways. First, to our knowledge, this is the first study to demonstrate that users can quickly and precisely identify unique preferences in the characteristics of their ankle exoskeleton assistance in two dimensions *simultaneously*. This result has major implications for the field because it motivates new strategies for the control of lower-limb exoskeletons in which users customize assistance according to their unique preferences, leveraging their internalized representation of optimality. Additionally, the two-dimensional self-tuning interface I designed for this study is a generalizable tool that may be used with different assistive device hardware and any desired control parameters. Second, this study evaluated user preference in autonomous (*i.e.,* untethered) robotic ankle exoskeletons, which were designed for able-bodied augmentation during walking or running. This is novel because prior studies have evaluated preference in users wearing lower-limb prostheses (*e.g.,* [22, 155]), or fully-actuated 12-DOF exoskeletons (*e.g.,* [172, 173])—devices that serve very different purposes than augmentative ankle exoskeletons. Therefore, this study defines a reference point for user preference in augmentation-focused exoskeletons, which may be used in the future to asses the impact of exoskeleton hardware (and assistance goals) on users' preferred control characteristics. The third contribution of this work is that it is the first study of its kind to rigorously ana-

lyze the characteristics of user preference, including how preference changes with walking speed, time, and exposure to the device, as well as how preferences differ between groups with different levels of prior technical knowledge and expertise. We found that, as the experiment progressed and participants gained experience, naive users (with no prior exoskeleton experience) preferred higher torque magnitude and became more precise identifying their preferences. At faster walking speeds, these individuals were more precise identifying both the magnitude and timing of their preferred assistance. Knowledgeable users (with prior exoskeleton experience) preferred higher torque magnitude than naive users. These results highlight that while user preference is a complex and dynamic quantity, individuals can repeatably and reliably identify their preferences. This study is a first step towards understanding the characteristics of preference in the context of exoskeleton control and discovering how individuals perceive, interact with, and learn from the assistance provided. The results from this study provide insight into how users interact with exoskeletons, and establish important benchmarks for researchers, designers, and future consumers.

6.2 Recommendations for Future Work

This dissertation presents key developments in the field of optimizing controllers for wearable robotic systems, yet there are an infinite number of exciting research paths to continue exploring in this field. In this section, I will discuss my recommendations for future work surrounding the two main themes in this dissertation: first, the translation of experimental methods and measurement tools outside the laboratory environment and second, the evaluation of context-specific objectives for the optimization of wearable robotic devices.

Translation Outside the Laboratory Environment

From my perspective, one of the most important directions for future research is moving beyond the confines of constrained laboratory testing. The vast majority of experimental research conducted over the last two decades (including the work presented in this dissertation) has featured individuals wearing lower limb assistive robotic devices walking on a treadmill or perform-

ing long bouts of steady-state activities in a laboratory setting. These past research efforts have resulted in a mountain of groundbreaking research that has answered many fundamental questions about the interactions between human users and wearable robots. Yet, we know that most activities that individuals perform during their jobs or daily lives are highly transient and non-steady-state (*e.g.,* [127]). High-performance experimental hardware continues to become untethered, lighter, and more compact, which is a crucial step towards translation. Thus, the next big hurdle for the field is the translation of experimental techniques and measurement tools from the laboratory environment into non-steady-state and "real world" conditions.

This dissertation (Chapters 3 and 4) presents initial steps towards the translation of body-in-the-loop optimization algorithms, by predicting instantaneous metabolic cost from physiological signals collected from portable, wearable sensors. In this context, I recommend several future directions for this work. First, we must improve the accuracy in which we can predict instantaneous metabolic cost using wearable sensors thorough model optimization (*e.g.,* feature selection and exploring a variety of sensor fusion algorithms). One possible route for this research would be to leverage state-of-the-art tools in machine learning and utilize black-box or deep learning models to predict energy cost from physiological signals. The benefit to a data-driven approach is that it may identify a variety of undetectable nonlinear relationships in the data that improve model performance; the downside to this approach is that it is fundamentally black-box, and does not provide us with and intuitive understanding of the physiological processes involved. Alternatively, one could explore model-based approaches, and determine dynamic relationships between key physiological signals (*e.g.,* heart rate, electrodermal activity, EMG, etc.) and energetic cost. The intersection of these approaches provides and interesting avenue for future research it itself—we began to explore the relationship between a black-box linear regression algorithm and an explicit first order model in Chapter 4, but future work is needed. In reality, it is likely a combination of data-driven and physiological model-based approaches that will yield the most informative and accurate algorithms for estimating instantaneous metabolic cost from wearable sensors.

Second, we must build algorithms that enable us to predict changes in metabolic cost more

quickly. From an algorithmic perspective, many of the same recommendations from the previous paragraph also apply here. For example, in Chapter 4 of this dissertation, I demonstrated that including a particular feature (*i.e.,* the signal derivative) improved the speed of instantaneous metabolic cost predictions using linear regression algorithms. If we also prioritize the speed of prediction when optimizing the predictive model structures, it is likely that similar algorithmic improvements (*e.g.,* feature selection) will result in better accuracy and faster predictions. To inform the prediction algorithms, we must also answer the question, "how fast is fast enough?" The answer to this question is likely closely tied to the specific application of the device. For an assistive device worn at home, the required speed of metabolic cost estimation is probably quite fast because the controller must continuously adapt to rapid and transient changes in activities performed throughout the day. For an application such as in-clinic device prescription, the necessary speed of estimation may determined by how many iterations of the optimization algorithm are required within a certain duration of time, which may also be dictated by the physical abilities of the patient.

Third, for maximum acceptance and translational potential, the future implementation of such algorithms should ideally only rely on an individual wearing a smart watch or smart textile (*e.g.,* the Hexoskin [25]). As such, future researchers in this space should always be looking for opportunities to use widely-accepted commercial products or develop miniaturized, unobtrusive, and portable versions of the necessary sensors. Finally, achieving these goals will inherently require large, diverse, and carefully-collected datasets in order to build and evaluate predictive algorithms. As such, I advocate for researchers and scientists to share their collected datasets open-source, to foster collaboration and accelerate the speed of translation.

Evaluation of Context-Specific Objectives

The work presented in this dissertation seeks to answer the question, "how do we provide optimal assistance from robotic exoskeletons and prostheses outside the laboratory environment?" The interpretation of this question is necessarily predicated on the definition of the word *optimal*. As I have discussed, in my work, I define the optimal controller as the one that maximizes the

objective of the device or assistance. This means that there is no one all-encompassing definition, and instead, optimality is *context-specific*. To this end, I advocate for the evaluation of context-specific objectives for the optimization of wearable robots as a future research direction.

In a practical sense, this future research path may look like critically evaluating different cost functions for wearable robots, with the goals of the task in mind. For certain applications, such as augmenting walking and running performance or increasing stamina and load-carrying potential for industrial workers or military personnel, minimizing the user's energetic cost using a robotic exoskeleton is an intuitive and valuable objective. For the elderly population or individuals with muskuloskeletal weakness, perhaps a more appropriate objective for exoskeleton assistance is to increase walking speed or support upright joint biomechanics during walking. In Chapter 5, I discuss one such potential objective for the optimization of robotic exoskeletons—user preference. While our work demonstrated that exoskeleton users have distinct preferences, we have not yet evaluated how the control settings that individuals prefer relate to those that optimize other physiological cost functions. Therefore, in addition to evaluating new outcomes, future research should also seek to compare control strategies that achieve different objectives. This research direction would provide fundamental scientific information regarding the impact that robotic assistance has on the entire human-robot system. Choosing appropriate cost functions is an exceptionally challenging task, and the individuals who will be using the exoskeleton or prosthesis should always be included in the design process. Continued research into the desired goals of end-users and the complex biomechanical and physiological impacts of robotic assistance will help inform future objectives to explore.

While the first phase of the general framework I proposed (Fig. 1.1) is *define objective*, the second is *measure objective*. We can define almost any goal imaginable as the objective of the robotic assistance, but without being able to robustly measure the objective (particularly outside the lab or clinic), it is very challenging to design controllers to meet the stated objective. Therefore, as we work to evaluate different context-specific objectives, we must simultaneously design accurate sensing and measurement tools. This may look like improving the design of existing sensors by

making them smaller and more robust (*e.g.,* EMG sensors). It may also look like developing new ways to measure or estimate the quantities of interest (*e.g.,* using wearable sensors and machine-learning approaches to build predictive algorithms). The only objectives we can use in our control systems are those we can reliably measure (or approximate), and so improved measurement tools are necessary for the furtherance of this research.

Finally, as optimization methods continue to evolve, we will be able to explore multi-objective optimization. Given the complexities of human locomotion and neuromotor control, it is likely that the best assistive robotic controller considers multiple objectives simultaneously, and these objectives may change in certain situations. As a long-term vision, the culmination of this research path will result in control systems for wearable robots that use high-level intent recognition or classification systems to identify the user's task, and seamlessly switch between the optimization of context-specific multi-objective cost functions. To achieve this goal, an exciting path for research is the mathematical representation of physiologically-based (*e.g.,* metabolic cost) and socially-motivated (*e.g.,* user comfort) criteria as multi-objective cost functions for use in optimization algorithms. As an example of this type of work, Kumar *et al.* recently described a convex objective function for continuous-phase control of powered prosthetic legs, which prioritizes kinematic tracking error while incorporating the patient's comfort level [98]. Multi-objective optimization may not always require the definition of one convex cost function, and instead may take the form of coactive feedback. Recently, Tucker *et al.* implemented an optimization algorithm to maximize subject comfort in a 12-DOF lower-body exoskeleton, and after every iteration, subjects were able to directly suggest improvements to the controller (*i.e.,* 10% longer step length). This type of coactive feedback could be utilized to allow the user to choose the exoskeleton assistance profile they prefer from a set of choices that produce similar metabolic economy.

6.3 Concluding Remarks

Wearable robotic devices, like prostheses and exoskeletons, have the potential to be life-changing technologies for millions of people who may use them for work, recreation, and rehabilitation. It is my hope that the contributions described herein will pave the way for future translational research related to the optimization and evaluation of control systems for wearable robotic systems.

APPENDICES

Appendix A

Comparison of Coefficients: Regression Model vs. Explicit Dynamic Model

Let $x(t)$ and $y(t)$ be two zero-mean signals which are subject to the dynamic constraint $\dot{y}(t) = \frac{1}{\tau}(x(t) - y(t))$. That is, $x(t) = \begin{bmatrix} y(t) & \dot{y}(t) \end{bmatrix} \begin{bmatrix} 1 \\ \tau \end{bmatrix}$. Now let $\hat{y}(t_i)$ be a series of n measurements of $y(t)$ that have been taken at times t_i and that are subject to a zero-mean additive noise $w(t)$: $\hat{y}(t_i) = y(t_i) + w(t_i)$. The vector of regression coefficients that optimally approximate $x(t_i)$ from these measurements is given by:

$$\mathbf{b} = \left(\hat{\mathbf{Y}}^T \hat{\mathbf{Y}}\right)^{-1} \hat{\mathbf{Y}}^T \mathbf{X}$$

where

$$\hat{\mathbf{Y}} = \mathbf{Y} + \mathbf{W} = \begin{bmatrix} y(t_1) & \dot{y}(t_1) \\ \vdots & \vdots \\ y(t_n) & \dot{y}(t_n) \end{bmatrix} + \begin{bmatrix} w(t_1) & \dot{w}(t_1) \\ \vdots & \vdots \\ w(t_n) & \dot{w}(t_n) \end{bmatrix}$$

and

$$\mathbf{X} = \begin{bmatrix} x(t_1) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} y(t_1) & \dot{y}(t_1) \\ \vdots & \vdots \\ y(t_n) & \dot{y}(t_n) \end{bmatrix} \begin{bmatrix} 1 \\ \tau \end{bmatrix} = \mathbf{Y} \begin{bmatrix} 1 \\ \tau \end{bmatrix}.$$

This leads to:

$$\begin{aligned}\mathbf{b} &= \left(\mathbf{Y}^T\mathbf{Y} + \mathbf{Y}^T\mathbf{W} + \mathbf{W}^T\mathbf{Y} + \mathbf{W}^T\mathbf{W}\right)^{-1}\left(\mathbf{Y}^T\mathbf{Y} + \mathbf{W}^T\mathbf{Y}\right)\begin{bmatrix}1\\ \tau\end{bmatrix} \\ &\approx \left(\mathrm{cov}(\mathbf{Y}) + \mathrm{cov}(\mathbf{W})\right)^{-1}\mathrm{cov}(\mathbf{Y})\begin{bmatrix}1\\ \tau\end{bmatrix},\end{aligned}$$

as the signals y and w have zero mean ($\mathrm{cov}(\mathbf{Y}) = \frac{1}{n}\mathbf{Y}^T\mathbf{Y}$, $\mathrm{cov}(\mathbf{W}) = \frac{1}{n}\mathbf{W}^T\mathbf{W}$) and are uncorrelated ($\mathbf{Y}^T\mathbf{W} \approx \mathbf{W}^T\mathbf{Y} \approx \mathbf{0}$). In the absence of noise, that is, for $\mathrm{cov}(\mathbf{W}) = \mathbf{0}$, this equation simplifies to $\mathbf{b} = \begin{bmatrix}1\\ \tau\end{bmatrix}$, but in the presence of noise, the coefficients $\mathbf{b}$ determined through the regression will differ from the theoretical model $\begin{bmatrix}1\\ \tau\end{bmatrix}$.

Appendix B

Chapter 5 Supplementary Analyses

B.1 Analyses Performed Using Preferred Magnitude Normalized to Body Mass

In the experiment presented in Chapter 5, we chose to have subjects self-tune their peak torque magnitude setting in Nm so that each subject explored the same range of settings. After data collection, we performed post-hoc analyses using preferred magnitude normalized by body mass ($^{\text{Nm}}/_{\text{kg}}$). For this normalization, we used subjects' measured weight while wearing the exoskeletons. Please note that the weights presented in Table 5.1 have had the mass of the exoskeleton (5 kg) removed. Using normalized preferred magnitude did not influence the interpretation of the results observed in this study. For comparison to the un-normalized analyses, please refer to Figs. 5.7–5.10, Fig. 5.12, and Table 5.2.

Across naive subjects, the minimum normalized preferred magnitude was 0.085 $^{\text{Nm}}/_{\text{kg}}$ ($\text{CI}_{95\%} =$ [0.08, 0.09] $^{\text{Nm}}/_{\text{kg}}$) (Fig. B.1); this minimum corresponded to the heaviest subject in our study (Subject 1) and was not the same subject with the lowest preference in un-normalized magnitude (Subject 4). The maximum normalized preferred magnitude was 0.221 $^{\text{Nm}}/_{\text{kg}}$ ($\text{CI}_{95\%} = [0.215, 0.22]$ $^{\text{Nm}}/_{\text{kg}}$) (Fig. B.1); this subject (Subject 12) preferred the highest magnitude in both normalized and un-normalized analyses.

The optimized linear mixed effects model (LMEM) structures for normalized preferred magnitude and standard deviation of normalized preferred magnitude were the same as their un-

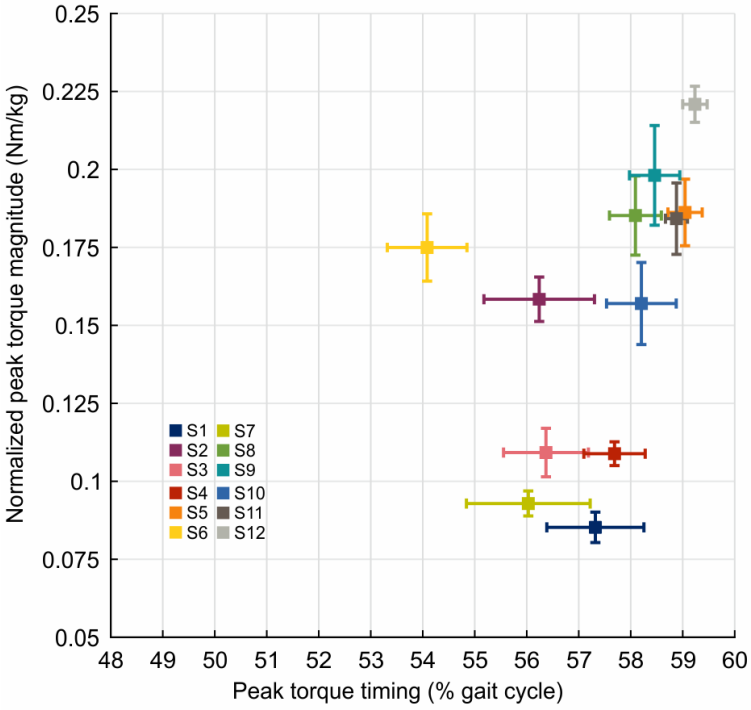

Figure B.1: User preference of normalized peak torque magnitude. Colored squares depict a subject's mean preferred timing and normalized magnitude across all 24 preference-identification trials. Colored error bars represent the 95% confidence interval about the mean

normalized versions. The optimized LMEM for normalized preferred magnitude had fixed effects of speed and trial block ($m = 3$) and included a subject-specific slope for trial block and a subject-specific intercept ($q = 2$, $K = 12$). The optimized LMEM for standard deviation of normalized preferred magnitude was the base model with fixed effects of speed and trial block ($m = 3$) and included a subject-specific intercept ($q = 1$, $K = 12$).

There was a significant effect of trial block on normalized preferred magnitude (LMEM, $p = .01$) (Fig. B.2B). According to the slope of the model, we would expect a subject's normalized preferred magnitude to increase by 0.02 $^{\text{Nm}}/_{\text{kg}}$ ($\text{CI}_{95\%} = [0.005, 0.034]$ $^{\text{Nm}}/_{\text{kg}}$) from the first to the third trial block (Table B.1). We did not observe a relationship between normalized preferred magnitude and walking speed (LMEM, $p = .10$) (Fig. B.2A).

The standard deviation of normalized preferred magnitude decreased significantly with increasing walking speed (LMEM, $p = .01$) (Fig. B.2C). According to the slope of the model, we would expect a subject's standard deviation of normalized preferred magnitude to decrease by 0.006 $^{\text{Nm}}/_{\text{kg}}$

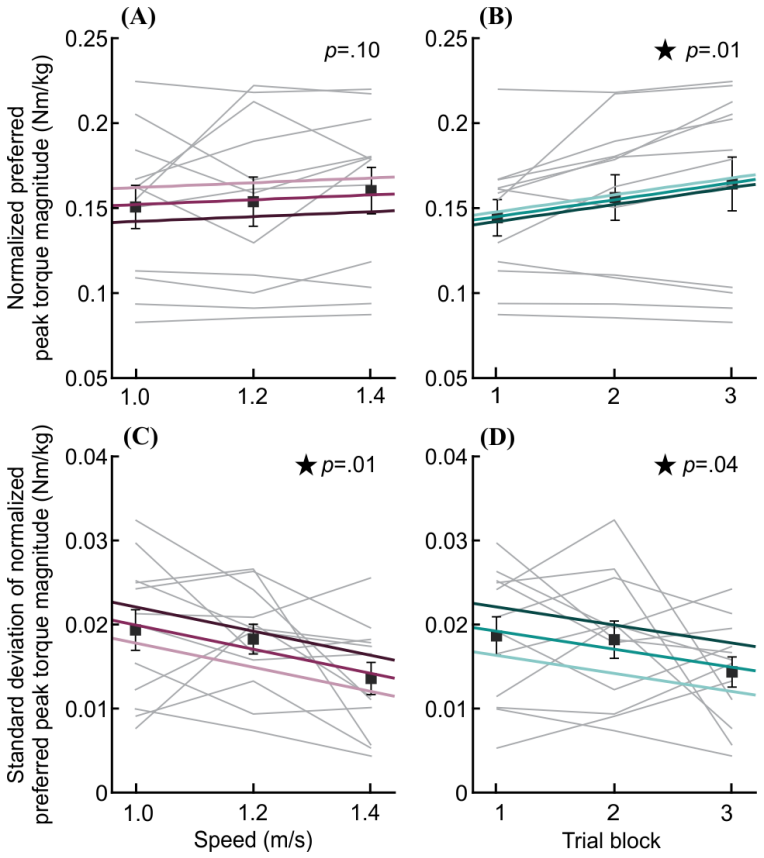

Figure B.2: Normalized peak torque magnitude preference and standard deviation as a function of speed and trial block. **(A)** Normalized preferred magnitude as a function of walking speed. **(B)** Normalized preferred magnitude as a function of trial block. **(C)** Standard deviation of normalized preferred magnitude as a function of walking speed. **(D)** Standard deviation of normalized preferred magnitude as a function of trial block. Black squares represent the inter-subject mean, and black error bars denote 1 standard error of the mean (SEM). Gray lines connect individual subjects' means. Each linear mixed effects model (LMEM) generates a surface that characterizes the relationship between the outcome and the two fixed effects (speed and trial block); a cartoon example of an LMEM surface is illustrated in Fig. 5.8. Left and right individual panels are 2-dimensional views of the same LMEM surface. Purple lines (left) show the slope of the model with respect to walking speed and represent slices of the LMEM surface at trial block 1 (dark), trial block 2 (medium), and trial block 3 (light). Teal lines (right) show the slope of the model with respect to trial block and represent slices of the LMEM surface at 1.0 m/s (dark), 1.2 m/s (medium), and 1.4 m/s (light). The LMEM p-value for each fixed effect's slope is depicted at the top right of each panel; a significant effect is denoted with a black star.

($\text{CI}_{95\%} = [0.002, 0.01]$ $^{\text{Nm}}/_{\text{kg}}$) as speed increases from 1–1.4 $^{\text{m}}/_{\text{s}}$. Additionally, the standard deviation of subjects' normalized preferred magnitude significantly decreased as trial block increased (LMEM, $p = .04$) (Fig. B.2D). According to the slope of the model, we would expect a subject's standard deviation of normalized preferred magnitude to decrease by 0.004 $^{\text{Nm}}/_{\text{kg}}$ ($\text{CI}_{95\%} = [0.0003, 0.008]$ $^{\text{Nm}}/_{\text{kg}}$) from the first to the third trial block (Table B.1). The only difference between the nor-

malized and un-normalized analyses was that this relationship was statistically significant while the relationship between un-normalized preferred magnitude and trial block did not quite surpass the threshold for significance (LMEM, $p = .05$); yet given the moderate downward trend we observed, this difference did not influence the interpretation of our results.

Knowledgeable subjects preferred significantly higher normalized peak torque magnitude (*t*-test, $p = .002$) than naive subjects (Fig. B.3A). There was no difference between groups in the standard deviation of normalized preferred magnitude (*t*-test, $p = .77$) (Fig. B.3B).

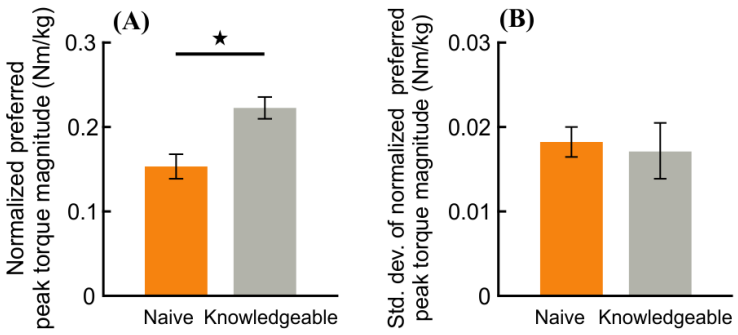

Figure B.3: Comparison of normalized peak torque magnitude and standard deviation between naive and knowledgeable users. **(A)** Normalized preferred magnitude. **(B)** Standard deviation of normalized preferred magnitude. Orange bars indicate the inter-subject mean of the outcome for naive subjects walking at 1.2 m/s. Gray bars indicate the inter-subject mean of the outcome for knowledgeable subjects walking at 1.2 m/s. Black error bars show 1 standard error of the mean (SEM). The black bracket with the star indicates a significant difference between the means, as calculated by a two-tailed *t*-test.

	Fixed effect = speed (tested range 1-1.4 m/s)			**Fixed effect = trial block (tested range 1-3)**		
Outcome	LMEM slope estimate [95% CI]	*p*-value of slope	Expected change over tested range (0.4 m/s)	LMEM slope estimate [95% CI]	*p*-value of slope	Expected change over tested range (2 blocks)
Normalized preferred magnitude (Nm/kg)	0.014 [-0.003, 0.031]	.10	0.01	**0.01 [0.003, 0.017]**	**.01**	**0.02**
Std. dev. of normalized preferred magnitude (Nm/kg)	**-0.014 [-0.024, -0.004]**	**.01**	**-0.006**	**-0.002 [-0.004,-0.0002]**	**.04**	**-0.004**

Bold text indicates a significant slope with $p < .05$.

Table B.1: Results from linear mixed effects models (LMEM) for normalized preference and precision outcomes with speed and trial block as fixed effects.

B.2 Comparisons Between Knowledgeable and Naive Cohorts During the Trial-Matched Condition

In the study presented in Chapter 5, we compared the preferences of the knowledgeable cohort (who performed one trial block while walking at 1.2 m/s), to the preferences of the naive cohort during the *speed-matched* condition of 1.2 m/s. These results are presented in Fig. 5.12. Alternatively, we analyzed comparisons between the knowledgeable cohort and the naive cohort during the *trial-matched* condition of trial block 1 (*i.e.,* trials 1–8). Compared to the analysis performed with data from naive subjects' speed-matched condition, all statistical relationships remained the same. Our main finding that knowledgeable subjects preferred higher magnitude than naive subjects remained unchanged as a result of this analysis (Fig. B.4A).

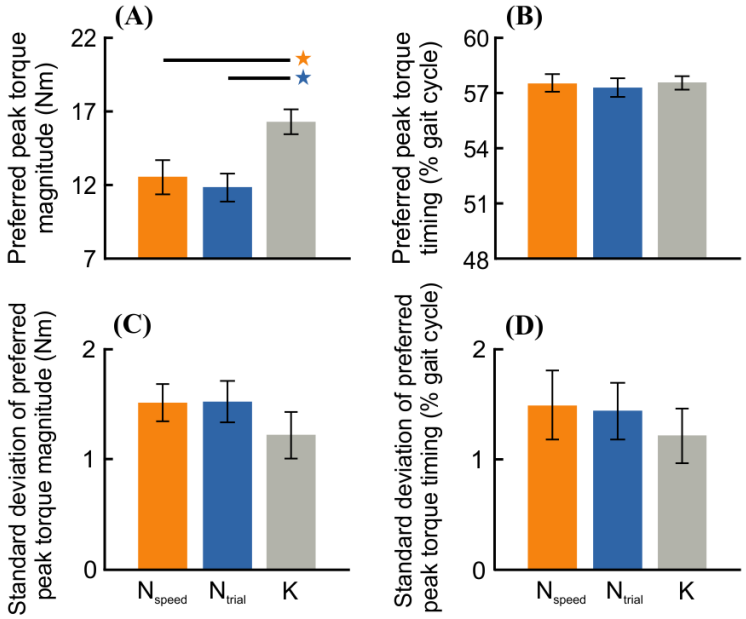

Figure B.4: Comparison of outcomes between knowledgeable users and naive users during speed-matched or trial-matched conditions. **(A)** Preferred magnitude. **(B)** Preferred timing. **(C)** Standard deviation of preferred magnitude. **(D)** Standard deviation of preferred timing. Orange bars indicate the inter-subject mean of the outcome for naive subjects during the speed-matched condition (1.2 m/s). Blue bars indicate the inter-subject mean of the outcome for naive subjects during the trial-matched condition (trial block 1, trials 1–8). Gray bars indicate the inter-subject mean of the outcome for knowledgeable subjects walking at 1.2 m/s. Black error bars show 1 standard error of the mean (SEM). A black bracket with an orange star indicates a significant difference between the knowledgeable (K) and naive speed-matched (N_{speed}) groups. A black bracket with a blue star indicates a significant difference between the knowledgeable (K) and naive trial-matched (N_{trial}) groups. Significance was determined using two-tailed *t*-tests.

BIBLIOGRAPHY

[1] M. A. Akkermans, M. J. Sillen, E. F. Wouters, and M. A. Spruit. Validation of the Oxycon Mobile metabolic system in healthy subjects. *Journal of Sports Science & Medicine*, 11(1):182, 2012.

[2] M. Altini, P. Casale, J. F. Penders, and O. Amft. Personalization of energy expenditure estimation in free living using topic models. *IEEE Journal of Biomedical and Health Informatics*, 19(5):1577–1586, 2015.

[3] M. Altini, J. Penders, R. Vullers, and O. Amft. Combining wearable accelerometer and physiological data for activity and energy expenditure estimation. In *Proceedings of the 4th Conference on Wireless Health*, pages 1–8, 2013.

[4] M. Altini, J. Penders, R. Vullers, and O. Amft. Estimating energy expenditure using body-worn accelerometers: A comparison of methods, sensors number and positioning. *IEEE Journal of Biomedical and Health Informatics*, 19(1):219–226, 2015.

[5] P. Antonellis, S. Galle, D. De Clercq, and P. Malcolm. Altering gait variability with an ankle exoskeleton. *PLoS One*, 13(10):e0205088, 2018.

[6] A. T. Asbeck, S. M. M. De Rossi, I. Galiana, Y. Ding, and C. J. Walsh. Stronger, smarter, softer: next-generation wearable robots. *IEEE Robotics & Automation Magazine*, 21(4):22–33, 2014.

[7] A. T. Asbeck, S. M. M. De Rossi, K. G. Holt, and C. J. Walsh. A biologically inspired soft exosuit for walking assistance. *The International Journal of Robotics Research*, 34(6):744–762, 2015.

[8] S. K. Au and H. Herr. Powered ankle-foot prosthesis. *IEEE Robotics & Automation Magazine*, 15(3):52–59, 2008.

[9] S. K. Au, J. Weber, and H. Herr. Powered ankle-foot prosthesis improves walking metabolic economy. *IEEE Transactions on Robotics*, 25(1):51–66, 2009.

[10] P. Beckerle, G. Salvietti, R. Unal, D. Prattichizzo, S. Rossi, C. Castellini, S. Hirche, S. Endo, H. B. Amor, M. Ciocarlie, F. Mastrogiovanni, B. D. Argall, and M. Bianchi. A human-robot interaction perspective on assistive and rehabilitation robotics. *Frontiers in Neurorobotics*, 11(1):24, 2017.

[11] K. M. Bedigrew, J. C. Patzkowski, J. M. Wilken, J. G. Owens, R. V. Blanck, D. J. Stinner, K. L. Kirk, and J. R. Hsu. Can an integrated orthotic and rehabilitation program decrease pain and improve function after lower extremity trauma? *Clinical Orthopaedics and Related Research*, 472(10):3017–3025, 2014.

[12] T. Beltrame, R. Amelard, R. Villar, M. J. Shafiee, A. Wong, and R. L. Hughson. Estimating oxygen uptake and energy expenditure during treadmill walking by neural network analysis of easy-to-obtain inputs. *Journal of Applied Physiology*, 121(5):1226–1233, 2016.

[13] T. Beltrame, R. Amelard, A. Wong, and R. L. Hughson. Prediction of oxygen uptake dynamics by machine learning analysis of wearable sensors during activities of daily living. *Scientific Reports*, 7(1):45738, 2017.

[14] Bionx Medical Technologies Inc. *Instructions for Use BiOM T2 Ankle*, 2016.

[15] O. M. Blake and J. M. Wakeling. Estimating changes in metabolic power from EMG. *SpringerPlus*, 2(1):229, 2013.

[16] S. Brage, N. Brage, P. W. Franks, U. Ekelund, and N. J. Wareham. Reliability and validity of the combined heart rate and movement sensor Actiheart. *European Journal of Clinical Nutrition*, 59(4):561–570, 2005.

[17] S. Brage, N. Brage, P. W. Franks, U. Ekelund, M.-Y. Wong, L. B. Andersen, K. Froberg, and N. J. Wareham. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *Journal of Applied Physiology*, 96(1):343–351, 2004.

[18] J. Brockway. Derivation of formulae used to calculate energy expenditure in man. *Human Nutrition. Clinical Nutrition*, 41(6):463–471, 1987.

[19] A. G. Brooks, S. M. Gunn, R. T. Withers, C. J. Gore, and J. L. Plummer. Predicting walking METs and energy expenditure from speed or accelerometry. *Medicine and Science in Sports and Exercise*, 37(7):1216–1223, 2005.

[20] S. M. Cain, K. E. Gordon, and D. P. Ferris. Locomotor adaptation to a powered ankle-foot orthosis depends on control method. *Journal of NeuroEngineering and Rehabilitation*, 4(1):48, 2007.

[21] J. Camargo, A. Ramanathan, W. Flanagan, and A. J. Young. A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions. *Journal of Biomechanics*, 119(1):110320, 2021.

[22] J. M. Caputo. *Informing ankle-foot prosthesis design and prescription through systematic experimentation with a tethered robotic prosthesis*. PhD thesis, Carnegie Mellon University, 2015.

[23] J. M. Caputo and S. H. Collins. Prosthetic ankle push-off work reduces metabolic rate but not collision work in non-amputee walking. *Scientific Reports*, 4(1):7213, 2014.

[24] J. M. Caputo and S. H. Collins. A universal ankle-foot prosthesis emulator for human locomotion experiments. *Journal of Biomechanical Engineering*, 136(3):035002, 2014.

[25] Carre Technologies Inc. Hexoskin Smart Shirts, 2018. URL www.hexoskin.com.

[26] Centers for Medicare and Medicaid Services, Springfield, VA. *Healthcare Common Procedure Coding System (HCPCS)*, 2010.

[27] P. Cherelle, V. Grosu, A. Matthys, B. Vanderborght, and D. Lefeber. Design and validation of the ankle mimicking prosthetic (AMP-) foot 2.0. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(1):138–148, 2013.

[28] T. R. Clites, M. K. Shepherd, K. A. Ingraham, and E. J. Rouse. Patient preference in the selection of prosthetic joint stiffness. In *Proceedings of the 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pages 1073–1079. IEEE, 2020.

[29] S. H. Collins, M. B. Wiggin, and G. S. Sawicki. Reducing the energy cost of human walking using an unpowered exoskeleton. *Nature*, 522(7555):212–215, 2015.

[30] C. Compher, D. Frankenfield, N. Keim, L. Roth-Yousey, and Evidence Analysis Working Group. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. *Journal of the American Dietetic Association*, 106(6):881–903, 2006.

[31] Cosmed. K5 Wearable Metabolic System for both Field and Laboratory Testing. URL https://www.cosmed.com/en/products/cardio-pulmonary-exercise-test/k5.

[32] S. E. Crouter, J. R. Churilla, and D. R. Bassett Jr. Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology*, 98(6):601–612, 2006.

[33] S. E. Crouter, K. G. Clowers, and D. R. Bassett Jr. A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology*, 100(4):1324–1331, 2006.

[34] A. Davidson, E. S. Gardinier, and D. H. Gates. Within and between-day reliability of energetic cost measures during treadmill walking. *Cogent Engineering*, 3(1):1251028, 2016.

[35] Y. Ding, M. Kim, S. Kuindersma, and C. J. Walsh. Human-in-the-loop optimization of hip assistance with a soft exosuit during walking. *Science Robotics*, 3(15):eaar5438, 2018.

[36] J. M. Donelan, R. Kram, and A. D. Kuo. Mechanical and metabolic determinants of the preferred step width in human walking. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1480):1985–1992, 2001.

[37] J. M. Donelan, R. Kram, and A. D. Kuo. Simultaneous positive and negative external mechanical work in human walking. *Journal of Biomechanics*, 35(1):117–124, 2002.

[38] M. F. Eilenberg, H. Geyer, and H. Herr. Control of a powered ankle-foot prosthesis based on a neuromuscular model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(2):164–173, 2010.

[39] K. Ellis, J. Kerr, S. Godbole, G. Lanckriet, D. Wing, and S. Marshall. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiological Measurement*, 35(11):2191–2203, 2014.

[40] A. Erdogan and B. D. Argall. Prediction of user preference over shared-control paradigms for a robotic wheelchair. In *Proceedings of the 15th IEEE/RAS-EMBS International Conference on Rehabilitation Robotics (ICORR)*, pages 1106–1111. IEEE, 2017.

[41] R. G. Eston, A. V. Rowlands, and D. K. Ingledew. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. *Journal of Applied Physiology*, 84(1):362–371, 1998.

[42] C. T. Farley and C. R. Taylor. A mechanical trigger for the trot-gallop transition in horses. *Science*, 253(5017):306–308, 1991.

[43] R. J. Farris, H. A. Quintero, and M. Goldfarb. Preliminary evaluation of a powered lower limb orthosis to aid walking in paraplegic individuals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(6):652–659, 2011.

[44] W. Felt, J. C. Selinger, J. M. Donelan, and C. D. Remy. "Body-in-the-loop": Optimizing device parameters using measures of instantaneous energetic cost. *PLoS One*, 10(8):e0135342, 2015.

[45] A. E. Ferris, J. M. Aldridge, C. A. Rábago, and J. M. Wilken. Evaluation of a powered ankle-foot prosthetic system during walking. *Archives of Physical Medicine and Rehabilitation*, 93(11):1911–1918, 2012.

[46] D. P. Ferris. The exoskeletons are here. *Journal of NeuroEngineering and Rehabilitation*, 6(1):17, 2009.

[47] D. P. Ferris, K. E. Gordon, G. S. Sawicki, and A. Peethambaran. An improved powered ankle-foot orthosis using proportional myoelectric control. *Gait & Posture*, 23(4):425–428, 2006.

[48] D. P. Ferris, G. S. Sawicki, and M. A. Daley. A physiologist's perspective on robotic exoskeletons for human locomotion. *International Journal of Humanoid Robotics*, 4(3):507–528, 2007.

[49] M. Franceschini, M. Massucci, L. Ferrari, M. Agosti, and C. Paroli. Effects of an ankle-foot orthosis on spatiotemporal parameters and energy cost of hemiparetic gait. *Clinical Rehabilitation*, 17(4):368–372, 2003.

[50] S. Galle, P. Malcolm, S. H. Collins, and D. De Clercq. Reducing the metabolic cost of walking with an ankle exoskeleton: Interaction between actuation timing and power. *Journal of NeuroEngineering and Rehabilitation*, 14(1):35, 2017.

[51] S. Galle, P. Malcolm, W. Derave, and D. De Clercq. Uphill walking with a simple exoskeleton: Plantarflexion assistance leads to proximal adaptations. *Gait & Posture*, 41(1):246–251, 2015.

[52] E. S. Gardinier, B. M. Kelly, J. Wensman, and D. H. Gates. A controlled clinical trial of a clinically-tuned powered ankle prosthesis in people with transtibial amputation. *Clinical Rehabilitation*, 32(3):319–329, 2018.

[53] D. H. Gates, J. M. Aldridge, and J. M. Wilken. Kinematic comparison of walking on uneven ground using powered and unpowered prostheses. *Clinical Biomechanics*, 28(4):467–472, 2013.

[54] D. H. Gates, B. J. Darter, J. B. Dingwell, and J. M. Wilken. Comparison of walking overground and in a computer assisted rehabilitation environment (CAREN) in individuals with and without transtibial amputation. *Journal of NeuroEngineering and Rehabilitation*, 9(1):81, 2012.

[55] D. H. Gates, J. B. Dingwell, S. J. Scott, E. H. Sinitski, and J. M. Wilken. Gait characteristics of individuals with transtibial amputations walking on a destabilizing rock surface. *Gait & Posture*, 36(1):33–39, 2012.

[56] D. Gopinath, S. Jain, and B. D. Argall. Human-in-the-loop optimization of shared autonomy in assistive robotics. *IEEE Robotics and Automation Letters*, 2(1):247–254, 2017.

[57] K. E. Gordon and D. P. Ferris. Learning to walk with a robotic ankle exoskeleton. *Journal of Biomechanics*, 40(12):2636–2644, 2007.

[58] A. M. Grabowski and S. D'Andrea. Effects of a powered ankle-foot prosthesis on kinetic loading of the unaffected leg during level-ground walking. *Journal of NeuroEngineering and Rehabilitation*, 10(1):49, 2013.

[59] L. J. Hargrove, A. M. Simon, R. Lipschutz, S. B. Finucane, and T. A. Kuiken. Non-weight-bearing neural control of a powered transfemoral prosthesis. *Journal of NeuroEngineering and Rehabilitation*, 10(1):62, 2013.

[60] H. A. Haugen, E. L. Melanson, Z. V. Tran, J. T. Kearney, and J. O. Hill. Variability of measured resting metabolic rate. *The American Journal of Clinical Nutrition*, 78(6):1141–1144, 2003.

[61] D. P. Heil. Predicting activity energy expenditure using the Actical activity monitor. *Research Quarterly for Exercise and Sport*, 77(1):64–80, 2006.

[62] H. Herr. Exoskeletons and orthoses: Classification, design challenges and future directions. *Journal of NeuroEngineering and Rehabilitation*, 6(1):21, 2009.

[63] H. Herr and A. M. Grabowski. Bionic ankle-foot prosthesis normalizes walking gait for persons with leg amputation. *Proceedings of the Royal Society B: Biological Sciences*, 279(1728):457–464, 2012.

[64] D. Hill and H. Herr. Effects of a powered ankle-foot prosthesis on kinetic loading of the contralateral limb: A case series. In *Proceedings of the 13th IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6. IEEE, 2013.

[65] J. K. Hitt, R. Bellman, M. Holgate, T. G. Sugar, and K. W. Hollander. The SPARKy (SPring Ankle with Regenerative Kinetics) project: Design and analysis of a robotic transtibial prosthesis with regenerative kinetics. In *Proceedings of the ASME 2007 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 1587–1596. American Society of Mechanical Engineers Digital Collection, 2007.

[66] K. G. Holt, S. F. Jeng, R. Ratcliffe, and J. Hamill. Energetic cost and stability during human walking at the preferred stride frequency. *Journal of Motor Behavior*, 27(2):164–178, 1995.

[67] T. Hortobágyi, A. Finch, S. Solnik, P. Rider, and P. DeVita. Association between muscle activation and metabolic cost of walking in young and old adults. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 66(5):541–547, 2011.

[68] H. Houdijk, E. Pollmann, M. Groenewold, H. Wiggerts, and W. Polomski. The energy cost for the step-to-step transition in amputee walking. *Gait & Posture*, 30(1):35–40, 2009.

[69] M.-J. Hsu, D. H. Nielsen, S.-J. Lin-Chan, and D. Shurr. The effects of prosthetic foot design on physiologic measurements, self-selected walking velocity, and physical activity in people with transtibial amputation. *Archives of Physical Medicine and Rehabilitation*, 87(1):123–129, 2006.

[70] H. Huang, D. L. Crouch, M. Liu, G. S. Sawicki, and D. Wang. A cyber expert system for auto-tuning powered prosthesis impedance control parameters. *Annals of Biomedical Engineering*, 44(5):1613–1624, 2016.

[71] S. Huang, J. P. Wensman, and D. P. Ferris. An experimental powered lower limb prosthesis using proportional myoelectric control. *Journal of Medical Devices*, 8(2), 2014.

[72] S. Huang, J. P. Wensman, and D. P. Ferris. Locomotor adaptation by transtibial amputees walking with an experimental powered prosthesis under continuous myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(5):573–581, 2015.

[73] K. J. Hunt, S. E. Fankhauser, and J. Saengsuwan. Identification of heart rate dynamics during moderate-to-vigorous treadmill exercise. *Biomedical Engineering Online*, 14(1):117, 2015.

[74] K. A. Ingraham, H. Choi, E. S. Gardinier, C. D. Remy, and D. H. Gates. Choosing appropriate prosthetic ankle work to reduce the metabolic cost of individuals with transtibial amputation. *Scientific Reports*, 8(1):15303, 2018.

[75] K. A. Ingraham, D. P. Ferris, and C. D. Remy. Using portable physiological sensors to estimate energy cost for 'body-in-the-loop' optimization of assistive robotic devices. In *Proceedings of the 5th IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 413–417. IEEE, 2017.

[76] K. A. Ingraham, D. P. Ferris, and C. D. Remy. Using wearable physiological sensors to predict energy expenditure. In *Proceedings of the 15th IEEE/RAS-EMBS International Conference on Rehabilitation Robotics (ICORR)*, pages 340–345. IEEE, 2017.

[77] K. A. Ingraham, D. P. Ferris, and C. D. Remy. Evaluating physiological signal salience for estimating metabolic energy cost from wearable sensors. *Journal of Applied Physiology*, 126(3):717–729, 2019.

[78] K. A. Ingraham, D. P. Ferris, and C. D. Remy. Predicting energy cost from wearable sensors: A dataset of energetic and physiological wearable sensor data from healthy individuals performing multiple physical activities, 2019. URL https://doi.org/10.6084/m9.figshare.7473191.v4.

[79] K. A. Ingraham, N. P. Fey, A. M. Simon, and L. J. Hargrove. Assessing the relative contributions of active ankle and knee assistance to the walking mechanics of transfemoral amputees using a powered prosthesis. *PLoS One*, 11(1):e0147661, 2016.

[80] K. A. Ingraham, C. D. Remy, and E. J. Rouse. User preference of applied torque characteristics for bilateral powered ankle exoskeletons. In *Proceedings of the 8th IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 839–845, 2020.

[81] K. A. Ingraham, E. J. Rouse, and C. D. Remy. Accelerating the estimation of metabolic cost using signal derivatives: Implications for optimization and evaluation of wearable robots. *IEEE Robotics & Automation Magazine*, 27(1):32–42, 2020.

[82] R. W. Jackson and S. H. Collins. An experimental comparison of the relative benefits of work and torque assistance in ankle exoskeletons. *Journal of Applied Physiology*, 119(5):541–557, 2015.

[83] R. W. Jackson and S. H. Collins. Heuristic-based ankle exoskeleton control for co-adaptive assistance of human locomotion. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(10):2059–2069, 2019.

[84] S. Jain, A. Farshchiansadegh, A. Broad, F. Abdollahi, F. Mussa-Ivaldi, and B. Argall. Assistive robotic manipulation through shared autonomy and a body-machine interface. In *Proceedings of the 14th IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 526–531. IEEE, 2015.

[85] H. S. Kaye, T. Kang, and M. LaPlante. Mobility device use in the United States. In *Disability Statistics Report*, volume 14. U.S. Department of Education, National Institute on Disability and Rehabilitation Research, 2000.

[86] L. Keytel, J. Goedecke, T. D. Noakes, H. Hiiloskorpi, R. Laukkanen, L. van der Merwe, and E. Lambert. Prediction of energy expenditure from heart rate monitoring during submaximal exercise. *Journal of Sports Sciences*, 23(3):289–297, 2005.

[87] D.-J. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang, and A. Behal. How autonomy impacts performance and satisfaction: Results from a study with spinal cord injured subjects using an assistive robot. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, 42(1):2–14, 2011.

[88] J. Kim, G. Lee, R. Heimgartner, D. A. Revi, N. Karavas, D. Nathanson, I. Galiana, A. Eckert-Erdheim, P. Murphy, D. Perry, N. Menard, D. K. Choe, P. Malcolm, and C. J. Walsh. Reducing the metabolic rate of walking and running with a versatile, portable exosuit. *Science*, 365(6454):668–672, 2019.

[89] M. Kim and S. H. Collins. Once-per-step control of ankle-foot prosthesis push-off work reduces effort associated with balance during walking. *Journal of NeuroEngineering and Rehabilitation*, 12(1):43, 2015.

[90] M. Kim, Y. Ding, P. Malcolm, J. Speeckaert, C. J. Siviy, C. J. Walsh, and S. Kuindersma. Human-in-the-loop Bayesian optimization of wearable device parameters. *PLoS One*, 12(9):e0184054, 2017.

[91] C. R. Kinnaird and D. P. Ferris. Medial gastrocnemius myoelectric control of a robotic ankle exoskeleton. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17(1):31–37, 2008.

[92] J. R. Koller, D. H. Gates, D. P. Ferris, and C. D. Remy. 'Body-in-the-loop' optimization of assistive robotic devices: A validation study. In *Robotics: Science and Systems (RSS)*, pages 1–10, 2016.

[93] J. R. Koller, D. H. Gates, D. P. Ferris, and C. D. Remy. Confidence in the curve: Establishing instantaneous cost mapping techniques using bilateral ankle exoskeletons. *Journal of Applied Physiology*, 122(2):242–252, 2017.

[94] J. R. Koller, D. A. Jacobs, D. P. Ferris, and C. D. Remy. Learning to walk with an adaptive gain proportional myoelectric controller for a robotic ankle exoskeleton. *Journal of NeuroEngineering and Rehabilitation*, 12(1):97, 2015.

[95] J. R. Koller, C. D. Remy, and D. P. Ferris. Biomechanics and energetics of walking in powered ankle exoskeletons using myoelectric control versus mechanically intrinsic control. *Journal of NeuroEngineering and Rehabilitation*, 15(1):42, 2018.

[96] S. Kramer, L. Johnson, J. Bernhardt, and T. Cumming. Energy expenditure and cost during walking after stroke: A systematic review. *Archives of Physical Medicine and Rehabilitation*, 97(4):619–632, 2016.

[97] J. Kulkarni, W. Gaine, J. Buckley, J. Rankine, and J. Adams. Chronic low back pain in traumatic lower limb amputees. *Clinical Rehabilitation*, 19(1):81–86, 2005.

[98] S. Kumar, A. Mohammadi, D. Quintero, S. Rezazadeh, N. Gans, and R. D. Gregg. Extremum seeking control for model-free auto-tuning of powered prosthetic legs. *IEEE Transactions on Control Systems Technology*, 28(6):2120–2135, 2019.

[99] N. Lamarra, B. J. Whipp, S. A. Ward, and K. Wasserman. Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. *Journal of Applied Physiology*, 62(5):2003–2012, 1987.

[100] B. E. Lawson, J. Mitchell, D. Truex, A. Shultz, E. Ledoux, and M. Goldfarb. A robotic leg prosthesis: Design, control, and implementation. *IEEE Robotics & Automation Magazine*, 21(4):70–81, 2014.

[101] B. E. Lawson, H. A. Varol, A. Huff, E. Erdemir, and M. Goldfarb. Control of stair ascent and descent with a powered transfemoral prosthesis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(3):466–473, 2012.

[102] S. Lee, J. Kim, L. Baker, A. Long, N. Karavas, N. Menard, I. Galiana, and C. J. Walsh. Autonomous multi-joint soft exosuit with augmentation-power-based control parameter tuning reduces energy cost of loaded walking. *Journal of NeuroEngineering and Rehabilitation*, 15(1):66, 2018.

[103] U. H. Lee, C.-W. Pan, and E. J. Rouse. Empirical characterization of a high-performance exterior-rotor type brushless DC motor and drive. In *Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8018–8025. IEEE, 2019.

[104] C. L. Lewis and D. P. Ferris. Walking with increased ankle pushoff decreases hip muscle moments. *Journal of Biomechanics*, 41(10):2082–2089, 2008.

[105] K. Li, M. Tucker, E. Bıyık, E. Novoseller, J. W. Burdick, Y. Sui, D. Sadigh, Y. Yue, and A. D. Ames. ROIAL: Region of interest active learning for characterizing exoskeleton gait preference landscapes. *arXiv preprint arXiv:2011.04812*, 2020.

[106] R. Li, P. Deurenberg, and J. G. A. J. Hautvast. A critical evaluation of heart rate monitoring to assess energy expenditure in individuals. *The American Journal of Clinical Nutrition*, 58(5):602–607, 1993.

[107] G. Lv, H. Zhu, and R. D. Gregg. On the design and control of highly backdrivable lower-limb exoskeletons: A discussion of past and ongoing work. *IEEE Control Systems Magazine*, 38(6):88–113, 2018.

[108] V. Y. Ma, L. Chan, and K. J. Carruthers. Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: Stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. *Archives of Physical Medicine and Rehabilitation*, 95(5):986–995, 2014.

[109] D. J. Macfarlane. Open-circuit respirometry: A historical review of portable gas analysis systems. *European Journal of Applied Physiology*, 117(12):2369–2386, 2017.

[110] M. J. Major, J. Scham, and M. Orendurff. The effects of common footwear on stance-phase mechanical properties of the prosthetic foot-shoe system. *Prosthetics and Orthotics International*, 42(2):198–207, 2018.

[111] P. Malcolm, W. Derave, S. Galle, and D. De Clercq. A simple exoskeleton that assists plantarflexion can reduce the metabolic cost of human walking. *PLoS One*, 8(2):e56137, 2013.

[112] P. Malcolm, S. Lee, S. Crea, C. Siviy, F. Saucedo, I. Galiana, F. A. Panizzolo, K. G. Holt, and C. J. Walsh. Varying negative work assistance at the ankle with a soft exosuit during loaded walking. *Journal of NeuroEngineering and Rehabilitation*, 14(1):62, 2017.

[113] P. Malcolm, R. E. Quesada, J. M. Caputo, and S. H. Collins. The influence of push-off timing in a robotic ankle-foot prosthesis on the energetics and mechanics of walking. *Journal of NeuroEngineering and Rehabilitation*, 12(1):21, 2015.

[114] J. Markowitz, P. Krishnaswamy, M. F. Eilenberg, K. Endo, C. Barnhart, and H. Herr. Speed adaptation in a powered transtibial prosthesis controlled with a neuromuscular model. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1570):1621–1631, 2011.

[115] MC10 Inc. BioStampRC. URL www.mc10inc.com.

[116] S. A. McClave, D. A. Spain, J. L. Skolnick, C. C. Lowen, M. J. Kleber, P. S. Wickerham, J. R. Vogt, and S. W. Looney. Achievement of steady state optimizes results when performing indirect calorimetry. *Journal of Parenteral and Enteral Nutrition*, 27(1):16–20, 2003.

[117] B. F. Mentiplay, M. Banky, R. A. Clark, M. B. Kahn, and G. Williams. Lower limb angular velocity during walking at various speeds. *Gait & Posture*, 65(1):190–196, 2018.

[118] P. G. Montgomery, D. J. Green, N. Etxebarria, D. B. Pyne, P. U. Saunders, and C. L. Minahan. Validation of heart rate monitor-based predictions of oxygen uptake and energy expenditure. *The Journal of Strength & Conditioning Research*, 23(5):1489–1495, 2009.

[119] L. M. Mooney and H. Herr. Biomechanical walking mechanisms underlying the metabolic reduction caused by an autonomous exoskeleton. *Journal of NeuroEngineering and Rehabilitation*, 13(1):4, 2016.

[120] L. M. Mooney, E. J. Rouse, and H. Herr. Autonomous exoskeleton reduces metabolic cost of human walking. *Journal of NeuroEngineering and Rehabilitation*, 11(1):151, 2014.

[121] L. M. Mooney, E. J. Rouse, and H. Herr. Autonomous exoskeleton reduces metabolic cost of human walking during load carriage. *Journal of NeuroEngineering and Rehabilitation*, 11(1):80, 2014.

[122] National Health Interview Survey. Percentage of difficulty walking or climbing steps for adults aged 18 and over, United States, 2019. Generated interactively. URL www.cdc.gov/NHISDataQueryTool/SHS_2019_ADULT3.

[123] R. R. Neptune, S. A. Kautz, and F. E. Zajac. Contributions of the individual ankle plantar flexors to support, forward progression and swing initiation during walking. *Journal of Biomechanics*, 34(11):1387–1398, 2001.

[124] J. A. Norris, A. P. Marsh, K. P. Granata, and S. D. Ross. Positive feedback in powered exoskeletons: Improved metabolic efficiency at the cost of reduced stability? In *Proceedings of the ASME 2007 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 1619–1626. American Society of Mechanical Engineers Digital Collection, 2007.

[125] D. C. Norvell, J. M. Czerniecki, G. E. Reiber, C. Maynard, J. A. Pecoraro, and N. S. Weiss. The prevalence of knee pain and symptomatic knee osteoarthritis among veteran traumatic amputees and nonamputees. *Archives of Physical Medicine and Rehabilitation*, 86(3):487–493, 2005.

[126] D. Novak, M. Mihelj, and M. Munih. Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction. *Robotica*, 29(3):367–374, 2011.

[127] M. S. Orendurff, J. A. Schoen, G. C. Bernatz, A. D. Segal, and G. K. Klute. How humans walk: Bout duration, steps per bout, and rest duration. *Journal of Rehabilitation Research & Development*, 45(7), 2008.

[128] Ottobock. Empower. URL https://www.ottobockus.com/prosthetics/lower-limb-prosthetics.

[129] F. A. Panizzolo, I. Galiana, A. T. Asbeck, C. Siviy, K. Schmidt, K. G. Holt, and C. J. Walsh. A biologically-inspired multi-joint soft exosuit that can reduce the energy cost of loaded walking. *Journal of NeuroEngineering and Rehabilitation*, 13(1):43, 2016.

[130] J. C. Patzkowski, R. V. Blanck, J. G. Owens, J. M. Wilken, K. L. Kirk, J. C. Wenke, J. R. Hsu, and Skeletal Trauma Research Consortium (STReC). Comparative effect of orthosis design on functional performance. *Journal of Bone & Joint Surgery*, 94(6):507–515, 2012.

[131] R. E. Quesada, J. M. Caputo, and S. H. Collins. Increasing ankle push-off work with a powered prosthesis does not necessarily reduce metabolic rate for transtibial amputees. *Journal of Biomechanics*, 49(14):3452–3459, 2016.

[132] B. T. Quinlivan, S. Lee, P. Malcolm, D. M. Rossi, M. Grimmer, C. Siviy, N. Karavas, D. Wagner, A. Asbeck, I. Galiana, and C. J. Walsh. Assistance magnitude versus metabolic cost reductions for a tethered multiarticular soft exosuit. *Science Robotics*, 2(2):eaah4416, 2017.

[133] D. Quintero, E. Reznick, D. J. Lambert, S. Rezazadeh, L. Gray, and R. D. Gregg. Intuitive clinician control interface for a powered knee-ankle prosthesis: A case study. *IEEE Journal of Translational Engineering in Health and Medicine*, 6:2600209, 2018.

[134] E. C. Ranz, E. R. Esposito, J. M. Wilken, and R. R. Neptune. The influence of passive-dynamic ankle-foot orthosis bending axis location on gait performance in individuals with lower-limb impairments. *Clinical Biomechanics*, 37:13–21, 2016.

[135] G. W. Reed and J. O. Hill. Measuring the thermic effect of food. *The American Journal of Clinical Nutrition*, 63(2):164–169, 1996.

[136] M. M. Reeves, P. S. Davies, J. Bauer, and D. Battistutta. Reducing the time period of steady state does not affect the accuracy of energy expenditure measurements by indirect calorimetry. *Journal of Applied Physiology*, 97(1):130–134, 2004.

[137] K. L. Rennie, S. J. Hennings, J. Mitchell, and N. J. Wareham. Estimating energy expenditure by heart rate monitoring without individual calibration. *Medicine and Science in Sports and Exercise*, 33(6):939–945, 2001.

[138] S. Rezazadeh, D. Quintero, N. Divekar, E. Reznick, L. Gray, and R. D. Gregg. A phase variable approach for improved rhythmic and non-rhythmic control of a powered knee-ankle prosthesis. *IEEE Access*, 7:109840–109855, 2019.

[139] C. P. Richter. Physiological factors involved in the electrical resistance of the skin. *American Journal of Physiology–Legacy Content*, 88(4):596–615, 1929.

[140] R. A. Robergs, D. Dwyer, and T. Astorino. Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Medicine*, 40(2):95–111, 2010.

[141] H. Rosdahl, L. Gullstrand, J. Salier-Eriksson, P. Johansson, and P. Schantz. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. *European Journal of Applied Physiology*, 109(2):159–171, 2010.

[142] J. Rose and J. G. Gamble. *Human Walking*. Lippincott Williams & Wilkins, 2006.

[143] M. P. Rothney, M. Neumann, A. Béziat, and K. Y. Chen. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *Journal of Applied Physiology*, 103(4):1419–1427, 2007.

[144] E. J. Rouse, L. M. Mooney, and H. M. Herr. Clutchable series-elastic actuator: Implications for prosthetic knee design. *The International Journal of Robotics Research*, 33(13):1611–1625, 2014.

[145] E. Russell Esposito, J. M. Aldridge Whitehead, and J. M. Wilken. Step-to-step transition work during level and inclined walking using passive and powered ankle-foot prostheses. *Prosthetics and Orthotics International*, 40(3):311–319, 2016.

[146] E. Russell Esposito, K. M. Rodriguez, C. A. Ràbago, and J. M. Wilken. Does unilateral transtibial amputation lead to greater metabolic demand during walking? *Journal of Rehabilitation Research and Development*, 51(8):1287–1296, 2014.

[147] E. Russell Esposito, K. A. Schmidtbauer, and J. M. Wilken. Experimental comparisons of passive and powered ankle-foot orthoses in individuals with limb reconstruction. *Journal of NeuroEngineering and Rehabilitation*, 15(1):111, 2018.

[148] H. Sadeghi, P. Allard, and M. Duhaime. Muscle power compensatory mechanisms in below-knee amputee gait. *American Journal of Physical Medicine & Rehabilitation*, 80(1):25–32, 2001.

[149] B. Saltin and A. Gagge. Sweating and body temperatures during exercise. *International Journal of Biometeorology*, 15(2-4):189–194, 1971.

[150] D. J. Sanderson and P. E. Martin. Lower extremity kinematic and kinetic adaptations in unilateral below-knee amputees during walking. *Gait & Posture*, 6(2):126–136, 1997.

[151] G. S. Sawicki, O. N. Beck, I. Kang, and A. J. Young. The exoskeleton expansion: improving walking and running economy. *Journal of NeuroEngineering and Rehabilitation*, 17(1):25, 2020.

[152] G. S. Sawicki and D. P. Ferris. Mechanics and energetics of level walking with powered ankle exoskeletons. *Journal of Experimental Biology*, 211(9):1402–1413, 2008.

[153] G. S. Sawicki, C. L. Lewis, and D. P. Ferris. It pays to have a spring in your step. *Exercise and Sport Sciences Reviews*, 37(3):130, 2009.

[154] J. C. Selinger and J. M. Donelan. Estimating instantaneous energetic cost during non-steady-state gait. *Journal of Applied Physiology*, 117(11):1406–1415, 2014.

[155] M. K. Shepherd, A. F. Azocar, M. J. Major, and E. J. Rouse. Amputee perception of prosthetic ankle stiffness during locomotion. *Journal of NeuroEngineering and Rehabilitation*, 15(1):99, 2018.

[156] M. K. Shepherd and E. J. Rouse. The VSPA foot: A quasi-passive ankle-foot prosthesis with continuously variable stiffness. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12):2375–2386, 2017.

[157] M. K. Shepherd and E. J. Rouse. Comparing preference of ankle-foot stiffness in below-knee amputees and prosthetists. *Scientific Reports*, 10(1):16067, 2020.

[158] M. K. Shepherd, A. M. Simon, J. Zisk, and L. Hargrove. Patient-preferred prosthetic ankle-foot alignment for ramps and level-ground walking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:52–59, 2020.

[159] A. Silder, T. Besier, and S. L. Delp. Predicting the metabolic cost of incline walking from muscle activity and walking mechanics. *Journal of Biomechanics*, 45(10):1842–1849, 2012.

[160] A. K. Silverman, N. P. Fey, A. Portillo, J. G. Walden, G. Bosker, and R. R. Neptune. Compensatory mechanisms in below-knee amputee gait in response to increasing steady-state walking speeds. *Gait & Posture*, 28(4):602–609, 2008.

[161] A. M. Simon, K. A. Ingraham, N. P. Fey, S. B. Finucane, R. D. Lipschutz, A. J. Young, and L. J. Hargrove. Configuring a powered knee and ankle prosthesis for transfemoral amputees within five specific ambulation modes. *PLoS One*, 9(6):e99387, 2014.

[162] D. C. Simonson and R. A. DeFronzo. Indirect calorimetry: Methodological and interpretative problems. *American Journal of Physiology–Endocrinology and Metabolism*, 258(3):E399–E412, 1990.

[163] J. Staudenmayer, D. Pober, S. Crouter, D. R. Bassett Jr., and P. Freedson. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, 107(4):1300–1307, 2009.

[164] F. Sup, H. A. Varol, and M. Goldfarb. Upslope walking with a powered knee and ankle prosthesis: initial results with an amputee subject. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(1):71–78, 2011.

[165] F. Sup, H. A. Varol, J. Mitchell, T. J. Withrow, and M. Goldfarb. Preliminary evaluations of a self-contained anthropomorphic transfemoral prosthesis. *IEEE/ASME Transactions on Mechatronics*, 14(6):667–676, 2009.

[166] A. M. Swartz, S. J. Strath, D. R. Bassett Jr., W. L. O'Brien, G. A. King, and B. E. Ainsworth. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and Science in Sports and Exercise*, 32(9):450–456, 2000.

[167] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *Proceedings of the 11th IEEE International Symposium on Wearable Computers*, pages 37–40. IEEE, 2007.

[168] N. Thatte, H. Duan, and H. Geyer. A sample-efficient black-box optimizer to train policies for human-in-the-loop systems with user preferences. *IEEE Robotics and Automation Letters*, 2(2):993–1000, 2017.

[169] N. Thatte, N. Srinivasan, and H. Geyer. Real-time reactive trip avoidance for powered transfemoral prostheses. In *Robotics: Science and Systems (RSS)*, 2019.

[170] O. Tikkanen, S. Kräkkäinen, P. Haakana, M. Kallinen, T. Pullinen, and T. Finni. EMG, heart rate, and accelerometer as estimators of energy expenditure in locomotion. *Medicine and Science in Sports and Exercise*, 46(9):1831–1839, 2014.

[171] L. Torburn, C. M. Powers, R. Gutierrez, and J. Perry. Energy expenditure during ambulation in dysvascular and traumatic below-knee amputees: A comparison of five prosthetic feet. *Journal of Rehabilitation Research and Development*, 32(2):111–119, 1995.

[172] M. Tucker, M. Cheng, E. Novoseller, R. Cheng, Y. Yue, J. W. Burdick, and A. D. Ames. Human preference-based learning for high-dimensional optimization of exoskeleton walking gaits. *arXiv preprint arXiv:2003.06495*, 2020.

[173] M. Tucker, E. Novoseller, C. Kann, Y. Sui, Y. Yue, J. W. Burdick, and A. D. Ames. Preference-based learning for exoskeleton gait optimization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2351–2357. IEEE, 2020.

[174] M. P. Tulppo, A. J. Hautala, T. H. Makikallio, R. T. Laukkanen, S. Nissila, R. L. Hughson, and H. V. Huikuri. Effects of aerobic training on heart rate dynamics in sedentary subjects. *Journal of Applied Physiology*, 95(1):364–372, 2003.

[175] R. Villar, T. Beltrame, and R. L. Hughson. Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities. *Applied Physiology, Nutrition, and Metabolism*, 40(10):1019–1024, 2015.

[176] D. J. Villarreal and R. D. Gregg. Controlling a powered transfemoral prosthetic leg using a unified phase variable. In *Wearable Robotics*, pages 487–506. Elsevier, 2020.

[177] N. Vyas, J. Farringdon, D. Andre, and J. I. Stivoric. Machine learning and sensor fusion for estimating continuous energy expenditure. *AI Magazine*, 33(2):55–66, 2012.

[178] J. M. Wakeling, O. M. Blake, I. Wong, M. Rana, and S. S. Lee. Movement mechanics as a determinate of muscle structure, recruitment and coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1570):1554–1564, 2011.

[179] R. L. Waters and S. Mulroy. The energy expenditure of normal and pathologic gait. *Gait & Posture*, 9(3):207–231, 1999.

[180] Y. Wen, J. Si, X. Gao, S. Huang, and H. H. Huang. A new powered lower limb prosthesis control framework based on adaptive dynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 28(9):2215–2220, 2017.

[181] B. J. Whipp, S. A. Ward, N. Lamarra, J. A. Davis, and K. Wasserman. Parameters of ventilatory and gas exchange dynamics during exercise. *Journal of Applied Physiology*, 52(6):1506–1513, 1982.

[182] D. A. Winter. Biomechanics of normal and pathological gait: Implications for understanding human locomotor control. *Journal of Motor Behavior*, 21(4):337–355, 1989.

[183] D. A. Winter and S. E. Sienko. Biomechanics of below-knee amputee gait. *Journal of Biomechanics*, 21(5):361–367, 1988.

[184] K. A. Witte, P. Fiers, A. L. Sheets-Singer, and S. H. Collins. Improving the energy economy of human running with powered and unpowered ankle exoskeleton assistance. *Science Robotics*, 5(40), 2020.

[185] A. J. Young and D. P. Ferris. State of the art and future directions for lower limb robotic exoskeletons. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(2):171–182, 2016.

[186] A. J. Young, J. Foss, H. Gannon, and D. P. Ferris. Influence of power delivery timing on the energetics and biomechanics of humans wearing a hip exoskeleton. *Frontiers in Bioengineering and Biotechnology*, 5:4, 2017.

[187] K. E. Zelik. Quantifying physical interface dynamics: Human-prosthesis and human-exoskeleton power transmission. In *Proceedings of the 40th Annual Meeting of the American Society of Biomechanics (ASB)*. ASB, 2016.

[188] K. E. Zelik and E. C. Honert. Ankle and foot power in gait analysis: Implications for science, technology and clinical assessment. *Journal of Biomechanics*, 75:1–12, 2018.

[189] J. Zhang, P. Fiers, K. A. Witte, R. W. Jackson, K. L. Poggensee, C. G. Atkeson, and S. H. Collins. Human-in-the-loop optimization of exoskeleton assistance during walking. *Science*, 356(6344):1280–1284, 2017.

[190] R. J. Zmitrewicz, R. R. Neptune, and K. Sasaki. Mechanical energetic contributions from individual muscles and elastic prosthetic feet during symmetric unilateral transtibial amputee walking: A theoretical study. *Journal of Biomechanics*, 40(8):1824–1831, 2007.