

Predicting Sagittal-Plane Swing Hip Kinematics in Response to Trips

Shannon M. Danforth¹, Xinyi Liu², Martin J. Ward³, Patrick D. Holmes¹, and Ram Vasudevan¹

Abstract—State-of-the-art wearable lower-limb robot controllers typically use established baseline human kinematics during common mobility tasks. Unfortunately due to the variability in human response during perturbations, these lower-limb controllers are unable to effectively assist with perturbation recovery. Accurate and quick predictions of kinematic responses to unexpected disturbances during motion can help assistive robotic devices safely aid with an individual's recovery. This paper presents three methods for predicting swing hip kinematics during trip recovery: a Gaussian process regression (GPR) model; a time-series neural network; and a pendulum model with linear feedback. Data were collected in an experiment where 16 subjects were tripped at random percentages of swing phase. The three prediction methods were applied to these data and evaluated for simulation accuracy and computation time. Both subject-specific and generalized models were investigated. Results indicate that the GPR model is the best choice for kinematic predictions due to its low simulation error in both subject-specific and generalized cases and lowest computation time.

Index Terms—Modeling and simulating humans, datasets for human motion, humanoid and bipedal locomotion.

I. INTRODUCTION

WEARABLE lower-limb robotic devices, such as exoskeletons or prostheses, offer advantages for people with mobility limitations by supplying positive work to joints [1], [2]. Control options for these devices typically focus on common tasks (e.g., walking on level ground), because baseline kinematics for these tasks are well-known [3]. If an unexpected event such as a trip occurs during walking, controllers struggle to effectively assist with recovery because it is unclear how the human will respond. We therefore need accurate predictions of human kinematics during these unexpected events. As illustrated in Fig. 1, this paper focuses on predicting sagittal-plane swing hip response during and after trips.

Manuscript received: January 31, 2022; Revised April 22, 2022; Accepted May 22, 2022.

This paper was recommended for publication by Editor Dana Kulic upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the National Science Foundation Career Award #1751093 and by the Office of Naval Research under Award Number N00014-18-1-2575. The authors thank Emma Reznick and Hannah Frame for providing motion capture system training. (Corresponding author: Shannon M. Danforth.)

¹S. M. Danforth, P. D. Holmes, and R. Vasudevan are with the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA. sdanfort@umich.edu, pdholmes@umich.edu, ramv@umich.edu

²X. Liu is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. xinyiww@umich.edu

³M. Ward is with the Department of Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI, USA. muu@umich.edu
Digital Object Identifier (DOI): see top of this page.

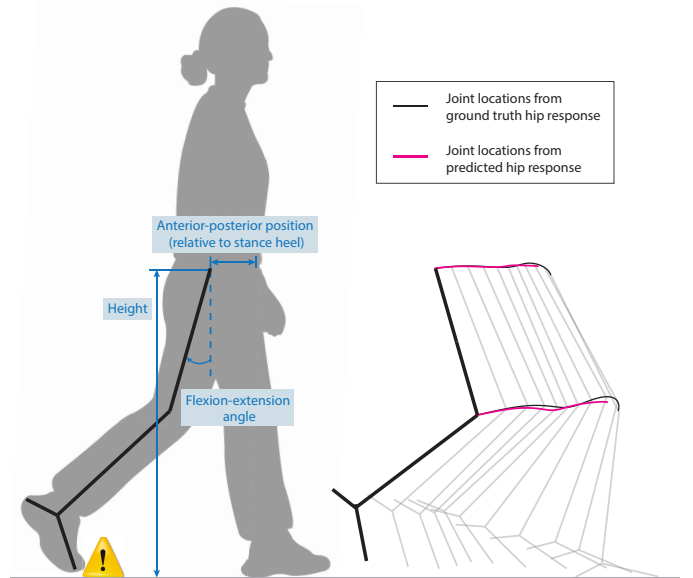


Fig. 1. This paper develops and compares predictions of sagittal-plane swing hip kinematics after a trip occurs. The three hip states of interest (height, anterior-posterior position, and flexion-extension angle) are shown in blue to the left. On the right, the stick figures show an example of one subject's trip-recovery response from our experiment. Note that the knee location is determined by the hip flexion-extension angle and thus is plotted in addition to the hip positions. The black line shows the ground truth hip and knee locations, and the pink line shows the predicted response from our Gaussian process regression model.

When tripped, humans typically exhibit one of three recovery strategies: Elevating, where the swing foot clears the obstacle and the heel strikes in front of the body; delayed lowering, where the swing foot initially elevates, then lowers behind the body with the toe contacting the ground first; and lowering, where the swing foot immediately lowers behind the body with the toe contacting the ground first [4]. These strategies have been studied most often in able-bodied subjects, but have also been observed in studies with lower-limb amputees [5]. Each of the trip-recovery strategies is accompanied by distinct swing hip kinematics, which are of interest if the controller aims to avoid toe scuffs or place the swing foot in a location that ensures stability at the end of the step. Furthermore, if the controller is planning trip-recovery behavior in real-time as in [6], the predictions must be generated before the human responds. Human perturbation reaction time is influenced by many factors, including psychological and neurological limitations [7], but studies have shown stretch reflex responses around 40 ms and voluntary responses around 160 ms [8].

Human trip reactions are difficult to model due to the trip's unpredictability and the variability in the human reaction. Trips can occur throughout the swing phase, and human recoveries can differ based on the perturbation onset and duration [4]. Eveld et al. developed an algorithm for predicting which of the three strategies a human will use after tripping [9], but even within a given recovery strategy, large variation in human kinematics exists. No models have been developed to predict trip-recovery kinematics after a specific strategy is identified. Thatte et al. used a Gaussian process regression (GPR) model conditioned on previously measured points to predict swing hip height and angle in response to "hip drop" perturbations [6]. Because the predictions were not the main focus of their paper, the authors did not provide information on training the model or report the model's prediction accuracy. The authors trained separate GPR models for each hip state, but training one multi-output model could increase accuracy by accounting for correlation between states.

Two other classes of models have been used for predictions of human kinematics. Neural networks are often used to estimate current kinematic variable values given measurements of different variables [10], but a recent study used neural networks to predict future lower-limb kinematics during different walking speeds [11]. The second common models are dynamic models that generate kinematic perturbation-recovery behavior. For example, a pendulum model with linear feedback has been shown to accurately estimate foot placement after a lateral perturbation [12]. Another study induced simulated trips in a sagittal-plane dynamic model and simulated its response using limit cycle optimization methods [13]. These studies, while highlighting the potential of various models for kinematic predictions, do not compare prediction accuracy between model classes or generate the strategy-specific trip-recovery predictions of interest.

In this paper, we evaluate hip response predictions using data from an able-bodied trip experiment with 16 subjects, described in Sec. II. Given Eveld et al.'s previous work in predicting strategy selection [9], this work assumes knowledge of the subject's trip-recovery strategy and focuses on predicting strategy-specific kinematics. First, we implement multi-output GPR models with conditional predictions of trip-recovery swing hip response variables (Sec. III). We compare the GPR models to nonlinear autoregressive exogenous (NARX) neural networks and pendulum models with a moving base (Sec. IV and V, respectively). Recognizing the need for both safety and low computation time, we evaluate all prediction methods for both simulation accuracy and online prediction time. We also evaluate each model's ability to generalize predictions across subjects, which would increase the prediction method's real-world applicability. Out of all simulation methods and training cases, the subject-specific GPR models exhibited the highest simulation accuracy and the lowest computation time. The 16-subject trip dataset is available at <https://doi.org/10.7302/pvhg-q324>, and MATLAB code with examples for training each model is available at https://github.com/roahmlab/swing_hip_trip_prediction.

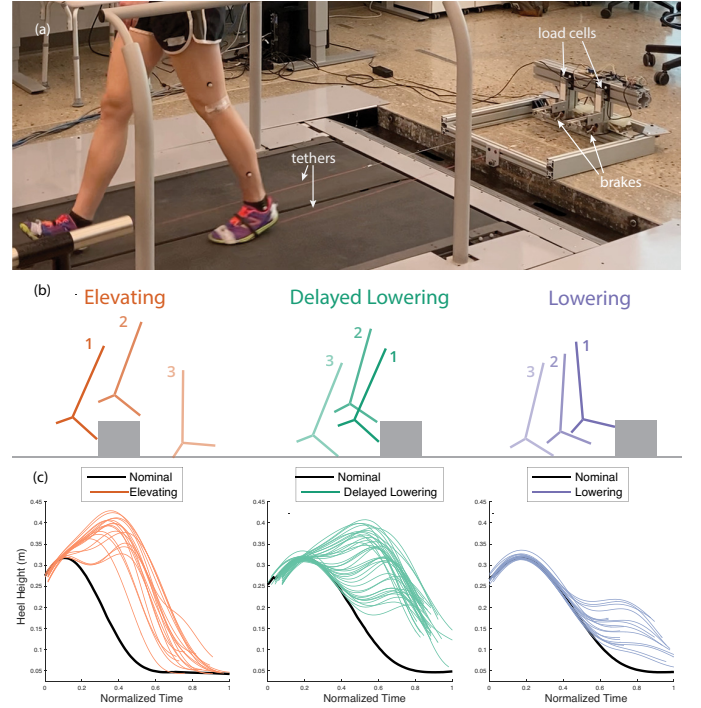


Fig. 2. Our experiment introduced trips via tethers attached to subjects' feet. (a) shows the trip experiment setup, including tethers attached to each foot, brakes applied at short intervals, and load cells to measure the tether force. (b) illustrates the swing leg during each strategy, adapted from [4]. (c) shows example swing heel heights for one subject during each type of recovery compared to nominal data.

II. TRIP-RECOVERY EXPERIMENT AND DATASET

This section describes our trip-recovery experiment with 16 subjects and the resulting dataset used to train and test predictive models.

A. Experiment

In our experiment, we tripped subjects using tethers attached to their feet. The tethers were routed to the back of the treadmill through a custom braking device, shown in Fig. 2, which was activated during the swing phase to interrupt the forward motion of the foot, thus inducing a trip-like perturbation. Shirota et al. found that this technique resulted in similar responses to trips as those from physical obstacles [4].

We collected data from 16 participants: Eight female and eight male; ages 18-29; mean (standard deviation) height 1.77 (0.10) m; mean (standard deviation) body mass 69.6 (12.6) kg. Each subject gave their informed written consent, and had no physical or balance disorders which could affect their ability to walk on a treadmill. The experimental protocol was approved by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board, eResearch ID: HUM00195042.

Subjects were fit with a harness in case of falls, then the tethers were attached to each foot. The subjects selected their desired level-ground treadmill (Bertec) walking speed in a warm-up that started at 1 m/s. After one minute, subjects could request to increase or decrease the speed in 0.2 m/s increments.

The mean (standard deviation) chosen walking speed was 1.14 (0.16) m/s. Then, the perturbed trials commenced. In total, the experiment consisted of 100 20-second walking trials, divided into 4 sessions of 25 trials each. A custom targeting algorithm measured force plate values from the instrumented treadmill in real-time, estimating whether each leg was in stance or swing. The trips were randomized by leg side (right or left); percentage of swing phase (10-80%); percentage of time into each 20-second trial (5-15 seconds); and duration of tether brake (150, 250, or 350 ms) to simulate varying contact duration with an obstacle. A minimum of 10 seconds between each trip allowed time for the subject to recover, and 15 random trials did not include trips. Load cells measured the force in each tether throughout the experiment, and a 26-camera Vicon motion capture system collected kinematic observations of 16 markers at 100 Hz. We applied a Woltring filter and fit lower-body segment models to each subject's data using Nexus Plug-in Gait [14], and we used MATLAB for all subsequent analyses [15]. A 6th-order Butterworth filter with a cut-off frequency of 6 Hz was used to filter joint position trajectories.

B. Dataset

We analyzed these experimental data during the swing phase only, i.e. between toe-off and heel-strike events. We determined each gait event using force plate readings with a threshold of 30 N for the stance phase. In this paper, the swing phase trajectories for the left and right foot are combined into one dataset. We aligned all of a subject's swing phase trajectories by peak heel height. After shifting the trials so the peaks aligned, we found the minimum and maximum shifted time across all trials for one strategy, then used the values to normalize time. The normalized swing time was defined separately for each recovery strategy.

Humans typically use one of three strategies to recover from trips, shown in Fig. 2(b) and (c) and described in Sec. I. Subjects tend to use elevating strategies when trips occur early in the swing phase, delayed lowering when trips occur towards the middle, and lowering when trips occur at the end. The trip-recovery data in this document was sorted into the three strategies by hand. If the trips did not induce a noticeable recovery due to incorrect brake timing, the trials were not analyzed. In this paper, we analyze 963 total trips across the 16 subjects: 433 elevating, 310 delayed lowering, and 220 lowering. The higher number of elevating trials is likely due to the targeting algorithm, which elicited more trips earlier in the swing phase.

Each recovery strategy was associated with distinct hip motion, shown for one subject in Fig. 3. The trip-recovery trials' difference from nominal level-ground walking varies by response variable and recovery strategy. For example, lowering trials most closely resemble nominal motion, while both elevating and delayed lowering trials exhibit more divergence. Within each strategy and response variable, the variation between trip-recovery trajectories also differs. The elevating and delayed lowering hip angles, for example, show a relatively large amount of variation between trials, while the lowering hip angles appear more uniform.

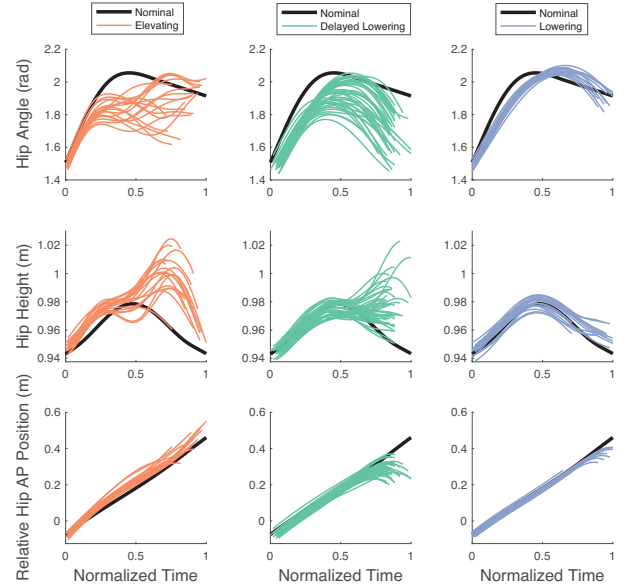


Fig. 3. Hip flexion-extension angle, height, and anterior-posterior (AP) position relative to the stance foot, plotted for the elevating, delayed lowering, and lowering strategies in orange, green, and purple, respectively. The thick black lines show the average nominal hip states. These data are from one subject in our dataset.

III. GAUSSIAN PROCESS REGRESSION MODEL

Given a set of noisy observations of a dependent variable, evaluated at certain values of an independent variable x , our prediction problem aims to find the best estimate of the dependent variable at a new value, x_* . Rather than assuming the underlying function is linear or another specific model, Gaussian process regression (GPR) assumes the observations can be represented as a sample from a multivariate Gaussian distribution. Note, we can always imagine a dataset with m observations to be a single point in $\mathbb{R}^m$ drawn from an m -variate Gaussian distribution [16]. GPR models also provide uncertainty bounds on predictions.

For our predictions, we compute a separate multi-output GPR model for each recovery strategy, with outputs of hip flexion-extension angle, height, and AP position, described in Sec. III-A. The GPR models are formed using a training dataset containing multiple trials for each recovery strategy. We then test trials that are excluded from the training dataset. For each test trial, we condition the trip-recovery prediction on its pre-trip values, described in Sec. III-B. The training and testing procedure is described in Sec. VI.

A. Multi-Output GPR Distribution

Each observation in our training dataset (note, the dataset contains multiple trip trials) can be thought of as related to an underlying function $\mathbf{f}$ through a Gaussian noise model:

$$\mathbf{y}_i = \mathbf{f}(x_i) + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (1)$$

where $x_i \in [0, 1] \subset \mathbb{R}$ is the argument to $\mathbf{f}$ and can be thought of as normalized time in our formulation, $\mathbf{y}_i \in \mathbb{R}^3$ is the output variable given by the hip flexion-extension angle, height, and AP position, and n is the number of points in the training

dataset. The noise term $\epsilon_i \in \mathbb{R}^3$ is a zero mean Gaussian distribution. We denote the covariance matrix of ϵ_i as $\mathbf{S}$.

Let $k(x_i, x_j)$ denote the *covariance function*, which relates one input value x_i to another x_j . We use the rational quadratic kernel, a common covariance function for GPR models [17, Chapter 4], given by

$$k(x_i, x_j) = \left(1 + \frac{\|x_i - x_j\|_2}{2\alpha l^2}\right)^{-\alpha}, \quad (2)$$

where $\|\cdot\|_2$ denotes the ℓ^2 -norm, α is the scale mixture parameter, and l is the length scale of the kernel. Parameters α and l are estimated separately for each recovery strategy. Sec. III-B describes the procedure for estimating α and l .

We calculate the covariance matrix $\mathbf{K}$ by evaluating (2) at all possible combinations of x_i for $i \in \{1, \dots, n\}$:

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}. \quad (3)$$

Recall that we aim to predict hip response values at a test input x_* . We therefore compute (2) between x_* and each observation x_i for $i = 1, \dots, n$:

$$\mathbf{K}_* = \begin{bmatrix} k(x_*, x_1) & k(x_*, x_2) & \cdots & k(x_*, x_n) \end{bmatrix}^T, \quad (4)$$

and the covariance between x_* and x_* :

$$\mathbf{K}_{**} = k(x_*, x_*). \quad (5)$$

In a single-output GPR model, we would now have our data represented as Gaussian distribution with covariance matrix $\mathbf{K}$. The mean and variance of this distribution evaluated at test input x_* would form the best estimate of the output. A multi-output GPR model adds an additional challenge, because our response vector $\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_n^T]^T$ is itself a Gaussian with zero mean and covariance matrix $\mathbf{C} \in \mathbb{R}^{3n \times 3n}$. Let the entry C_{ij}^{pq} be the covariance between the p^{th} and q^{th} response variable of vectors $\mathbf{y}_i$ and $\mathbf{y}_j$, for each $p, q \in \{1, 2, 3\}$ and for each $i, j \in \{1, 2, \dots, n\}$. Recalling the three-dimensional noise term ϵ_i in (1), each entry of $\mathbf{C}$ has two components: one corresponding to the cross-response covariance $\mathbf{C}_{\text{cross}}$, and one corresponding to the covariance of the noise term, $\mathbf{C}_{\text{noise}}$.

Calculating $\mathbf{C}$ requires the estimation of significantly more hyper-parameters. Two assumptions are made to simplify the multi-output formulation[18]:

- 1) Within a cross-response pair indexed by (p, q) , the covariance function is formulated as a scalar-valued covariance matrix $k(x_i, x_j)$ scaled by a cross-response weight B_{pq} . The cross-response covariance is therefore be written as $[C_{\text{cross}}]_{ij}^{pq} = B_{pq}k(x_i, x_j)$. The B_{pq} entries construct a symmetric matrix $\mathbf{B}$, representing the pairwise covariance between different response variables.
- 2) The noise covariance matrix $\mathbf{S}$ is assumed to be diagonal, because the covariances between responses have already been captured in the $\mathbf{C}_{\text{cross}}$ component. We therefore define $[C_{\text{noise}}]_{ij}^{pq} = \delta_{pq}\delta_{ij}S_{ij}$, where δ indicates a Kronecker delta.

With these assumptions, the covariance matrix $\mathbf{C}$ is:

$$\mathbf{C} = \begin{bmatrix} B_{11}\mathbf{K} + S_{11}\mathbf{I}_n & B_{12}\mathbf{K} & B_{13}\mathbf{K} \\ B_{21}\mathbf{K} & B_{22}\mathbf{K} + S_{22}\mathbf{I}_n & B_{23}\mathbf{K} \\ B_{31}\mathbf{K} & B_{32}\mathbf{K} & B_{33}\mathbf{K} + S_{33}\mathbf{I}_n \end{bmatrix}, \quad (6)$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, or $\mathbf{C} = \mathbf{B} \otimes \mathbf{K} + \mathbf{S} \otimes \mathbf{I}_n$, where $\otimes$ denotes the Kronecker product. Note that if $\mathbf{S}, \mathbf{B}$ and $\mathbf{K}$ are all positive definite, then so is $\mathbf{C}$, because both the Kronecker product and the summation on two positive definite matrices maintain positive definiteness.

Because $\mathbf{B}$ is symmetric, constructing $\mathbf{B}$ only requires the six elements from the upper triangle. Adding the three diagonal elements of $\mathbf{S}$, the multi-output formulation adds nine estimation parameters to the training process.

Our training dataset takes the form $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_n^T]^T$ for n data points, and we aim to predict the response $\mathbf{y}_*$ at a test input x_* . The probability of a certain prediction for $\mathbf{y}_*$ follows a Gaussian distribution. Evaluating (4) and (5), we find the distribution's mean and covariance at x_* :

$$\begin{aligned} \boldsymbol{\mu}_* &= \mathbf{C}_*^T \mathbf{C}^{-1} \mathbf{y} \\ \boldsymbol{\Sigma}_* &= \mathbf{B} \otimes \mathbf{K}_{**} + \mathbf{S} - \mathbf{C}_*^T \mathbf{C}^{-1} \mathbf{C}_*, \end{aligned} \quad (7)$$

where $\mathbf{C}_* = \mathbf{B} \otimes \mathbf{K}_*$. The best estimate for $\mathbf{y}_*$ and its associated uncertainty are given by $\boldsymbol{\mu}_*$ and $\boldsymbol{\Sigma}_*$. In our implementation, we evaluate the mean predictions and their covariance at m evenly-spaced values of x_* from 0 to 1, generating a $3m \times 1$ mean vector:

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \dots, \boldsymbol{\mu}_m^T]^T. \quad (8)$$

The associated covariance matrix $\boldsymbol{\Sigma}$ has size $3m \times 3m$.

B. Conditional Prediction

The mean values in (8), computed from a training dataset, provide a prediction of hip responses at discretized points in normalized time. Given a test trip trial, we can use $\boldsymbol{\mu}$ to predict the hip response values after the trip onset. However, $\boldsymbol{\mu}$ does not account for the pre-trip points of a test trial, and therefore does not generate a continuous prediction from pre-trip points. To improve prediction accuracy, we condition the GPR prediction on pre-trip points in the test trial.

Our test trial begins and ends at normalized time values x_1 and x_2 , not necessarily spanning the entire normalized swing time range. To ensure the test trial is the same size as $\boldsymbol{\mu}$, we re-sample the test trial, and pad with NaN values if x_1 does not equal 0 and/or x_2 does not equal 1, to obtain a $3m \times 1$ test trajectory $\mathbf{y}_t = [\mathbf{y}_{t,1}^T, \mathbf{y}_{t,2}^T, \dots, \mathbf{y}_{t,m}^T]^T$.

For a given test trial, we denote the index of trip occurrence as $i_{\text{trip}} \in \{1, \dots, m\}$. We then divide $\mathbf{y}_t$ into *completed* (pre-trip) and *future* (trip-recovery) trajectories:

$$\begin{aligned} \mathbf{y}_{t,c} &= [\mathbf{y}_{t,1}^T, \mathbf{y}_{t,2}^T, \dots, \mathbf{y}_{t,i_{\text{trip}}}^T]^T \\ \mathbf{y}_{t,f} &= [\mathbf{y}_{t,i_{\text{trip}}+1}^T, \mathbf{y}_{t,i_{\text{trip}}+2}^T, \dots, \mathbf{y}_{t,m}^T]^T, \end{aligned} \quad (9)$$

where subscript c indicates completed and f indicates future. We likewise partition the GPR mean and covariance into completed and future parts:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_c \\ \boldsymbol{\mu}_f \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{cc} & \boldsymbol{\Sigma}_{cf} \\ \boldsymbol{\Sigma}_{fc} & \boldsymbol{\Sigma}_{ff} \end{bmatrix}. \quad (10)$$

Then, our predicted future (trip-recovery) hip states and the prediction's uncertainty are given by the conditional mean and covariance:

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \boldsymbol{\mu}_f + \boldsymbol{\Sigma}_{fc} \boldsymbol{\Sigma}_{cc}^{-1} (\mathbf{y}_c - \boldsymbol{\mu}_c) \\ \hat{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma}_{ff} - \boldsymbol{\Sigma}_{fc} \boldsymbol{\Sigma}_{cc}^{-1} \boldsymbol{\Sigma}_{cf}.\end{aligned}\quad (11)$$

The prediction covariance can be used to form a confidence interval. As the number of conditional points grows, the inverse term in (11) takes longer to compute. In practice, to enable real-time predictions, we use at most 15 conditional normalized time points in our predictions.

For each recovery strategy, our multi-output GPR has 11 parameters to train: Parameters α and l from (2) and the nine parameters from the multi-output formulation. During training, we select parameters that minimize the root mean square error (RMSE) between the ground truth trip response and the conditioned GPR prediction across all response variables and training trials. One can solve this optimization problem using a nonlinear programming solver like MATLAB's `fmincon`. Although the training process may be time-consuming, the resulting GPR model can be saved and loaded for real-time conditional prediction.

IV. NARX MODEL

For comparison to the GPR model in Sec. III, we develop nonlinear autoregressive exogenous (NARX) models to predict hip response kinematics. Neural network methods, such as recurrent neural networks (RNNs) and autoregressive networks (ARs), have been popular options for forecasting time series with unknown underlying functions [11]. It has been shown that any RNN network can be equivalent to a NARX neural network with a similar neuron transfer function [19]. Because NARX models exhibit computational advantages in training compared to RNNs, we use NARX models in this work.

We fit separate multi-output NARX models for each recovery strategy. We form the models using a training dataset and then test the models on trials that are excluded from training. Because the NARX model requires sequential time-series data for training, we concatenate the training trials into one time series. When combining the normalized time values from each trial into one vector, we add an offset to each subsequent trial, i.e., the first training trial's normalized time points lie in $[0, 1]$, the second trial's normalized time points lie in $[1, 2]$, and so on. For the hip response variables, we likewise concatenate the trajectories from each training trial into one time series (with no offset). Our concatenated dataset with n points has normalized time values $x_i \in \mathbb{R}, i = 1, \dots, n$. The $3 \times n$ response variable training data takes the form $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$, with each $\mathbf{y}_i \in \mathbb{R}^3$.

The NARX model fits a function $\mathbf{f}$ that relates each entry of the time series, $\mathbf{y}_i$, with previous values, $\mathbf{y}_{i,\text{prev}}$, and exogenous input, $x_{i,\text{ex}}$:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{y}_{i,\text{prev}}, x_{i,\text{ex}}) + \boldsymbol{\varepsilon}_i, \quad (12)$$

where $\boldsymbol{\varepsilon}_i$ is an error term. At the i^{th} step, we specify the exogenous input to be the previous normalized time value, i.e. $x_{i,\text{ex}} = x_{i-1}$, and use N previous response values, i.e. $\mathbf{y}_{i,\text{prev}} = [\mathbf{y}_{i-N}, \mathbf{y}_{i-N+1}, \dots, \mathbf{y}_{i-1}]$.

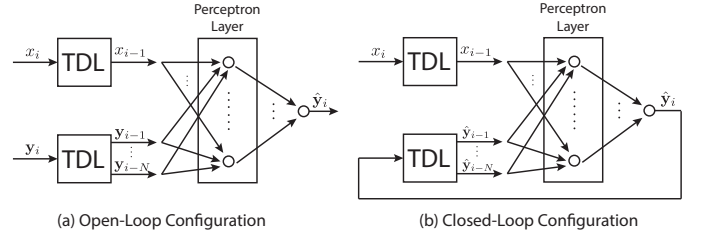


Fig. 4. Diagram of the (a) open-loop configuration and (b) closed-loop configuration of the NARX model, where TDL stands for time delay layer.

We train the NARX networks using MATLAB's `narxnet`. As shown in Fig. 4, there are two configurations, open-loop and closed-loop. When training each network, we use the open-loop configuration. At each index i of our training dataset, we input the delayed ground truth response $\mathbf{y}_{i,\text{prev}}$ and the exogenous input $x_{i,\text{ex}}$. The inputs first pass through a fully-connected layer of perceptrons, also depicted in Fig. 4, which apply the sigmoid operation to yield intermediate outputs. For the j^{th} perceptron, the intermediate output $\mathbf{o}^j$ at step i is given by

$$\mathbf{o}^j = \sigma((\mathbf{w}_x^j)^T \mathbf{x}_{i,\text{ex}} + (\mathbf{w}_y^j)^T \mathbf{y}_{i,\text{prev}} + \mathbf{b}^j), \quad (13)$$

where σ is the sigmoid function, $\mathbf{x}_{i,\text{ex}} = [x_{i,\text{ex}}, x_{i,\text{ex}}, x_{i,\text{ex}}]^T$, $\mathbf{b}^j \in \mathbb{R}^3$ is the bias for the j^{th} perceptron, and $\mathbf{w}_x^j \in \mathbb{R}^3$ and $\mathbf{w}_y^j \in \mathbb{R}^{3 \times N}$ are the weights for the j^{th} perceptron [20]. The intermediate outputs are then passed through the output linear perceptron, yielding a prediction $\hat{\mathbf{y}}_i$. The NARX network selects the bias and weights for each perceptron that minimize the RMSE between the true output $\mathbf{y}_i$ and predicted output $\hat{\mathbf{y}}_i$, for $i \in \{1, \dots, n\}$.

For a test trial with m data points, we denote the normalized time and response as $x_t = [x_{t,1}, x_{t,2}, \dots, x_{t,m}]$ and $\mathbf{y}_t = [\mathbf{y}_{t,1}, \mathbf{y}_{t,2}, \dots, \mathbf{y}_{t,m}]$, respectively. Because we predict future values in the time series with the NARX model, the x_t vector is offset so that it occurs immediately after the concatenated training data. As in Sec. III-B, the test trial has trip onset index i_{trip} and can be divided into completed and future partitions. We use the completed portion of the trial, $\mathbf{y}_{t,c} = [\mathbf{y}_{t,1}, \mathbf{y}_{t,2}, \dots, \mathbf{y}_{t,i_{\text{trip}}}]$, to inform post-trip predictions.

We use the closed-loop configuration, shown in Fig. 4(b), to generate predictions of the test trial for multiple future steps. The optimized bias and weight parameters from the training process are used in the closed-loop model. To predict a future post-trip value $\mathbf{y}_{t,i}$ where $i > i_{\text{trip}}$, the closed loop configuration uses previous predicted values, $\hat{\mathbf{y}}_{t,i,\text{prev}} = [\hat{\mathbf{y}}_{t,i-N}, \hat{\mathbf{y}}_{t,i-N+1}, \dots, \hat{\mathbf{y}}_{t,i-1}]$, as input. However, for the first N points after the trip occurs, there are not yet enough predicted values to form N previous inputs. Therefore, to predict the first post-trip point $\mathbf{y}_{t,i_{\text{trip}}+1}$, we use N previous points of $\mathbf{y}_{t,c}$; to predict $\mathbf{y}_{t,i_{\text{trip}}+2}$, we use $\hat{\mathbf{y}}_{t,i_{\text{trip}}+1}$ and $N-1$ previous points of $\mathbf{y}_{t,c}$; and so on. We set the feedback delay N to be 15 normalized time steps, the same as the number of conditional points used for the GPR model.

V. PENDULUM MODEL WITH MOVING BASE

To describe the dynamics of the swing hip variables, we use a pendulum model with a moving base, shown in Fig. 5.

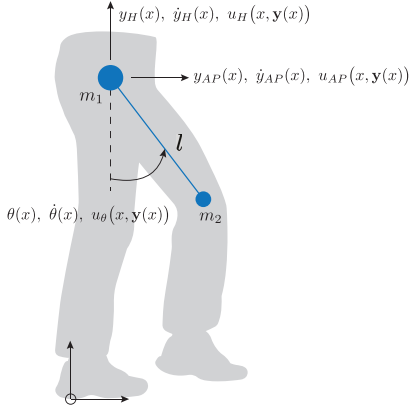


Fig. 5. Our model for predicting hip kinematics has a floating base located at the hip joint and a pendulum representing the swing leg. The model has horizontal and vertical forces and a torque applied at the hip.

Although we focus on kinematics in this paper, dynamic models have been used to effectively predict kinematic variables in previous research [12], [13]. A dynamic model could also prove useful if predictions of hip forces/torques are needed in a wearable robot's control scheme.

Our pendulum model consists of a mass m_1 at the hip joint location, which can translate in the horizontal and vertical direction, connected to a pendulum of length l and mass m_2 . The mass m_1 consists of the estimated torso, arms, and head, and the mass m_2 consists of the estimated swing leg mass. The pendulum length l is the estimated swing thigh length. Masses and lengths were estimated using average distributions from anthropometric data [21, Chapter 4]. Because m_1 is considerably larger than m_2 , our model neglects the effect of the pendulum motion on the moving base. The model's origin is the location of the stance heel.

For this section, we define the normalized time as $x \in [0, 1] \subset \mathbb{R}$. At each point x , our model has anterior-posterior hip position $y_{AP}(x)$, hip height $y_H(x)$, and hip flexion-extension angle $\theta(x)$. The linear and angular velocities (with respect to normalized time) are denoted as $\dot{y}_{AP}(x)$, $\dot{y}_H(x)$, and $\dot{\theta}(x)$. Let $\mathbf{y}(x) = [y_{AP}(x) \ \dot{y}_{AP}(x) \ y_H(x) \ \dot{y}_H(x) \ \theta(x) \ \dot{\theta}(x)]^T$ denote the state vector. The model has three inputs: a horizontal force $u_{AP}(x, \mathbf{y}(x))$ acting on m_1 , a vertical force $u_H(x, \mathbf{y}(x))$ acting on m_1 , and a torque $u_\theta(x, \mathbf{y}(x))$ applied to the pendulum. The acceleration of the horizontal hip position is given by $\ddot{y}_{AP}(x) = \frac{1}{m_1} u_{AP}(x, \mathbf{y}(x))$ and the vertical hip acceleration is given by $\ddot{y}_H(x) = \frac{1}{m_1} u_H(x, \mathbf{y}(x)) - g$, where g is the gravitational acceleration 9.81 m/s^2 . The hip flexion-extension angular acceleration is given as

$$\ddot{\theta}(x) = -\frac{1}{l} \cos(\theta(x)) \ddot{y}_{AP}(x) - \frac{1}{l} \sin(\theta(x)) \ddot{y}_H(x) + \dots - \frac{g}{l} \sin(\theta(x)) + \frac{1}{m_2 l^2} u_\theta(x, \mathbf{y}(x)). \quad (14)$$

We produce six first-order differential equations for the system using these accelerations.

Drawing from work that used pendulum models with linear feedback to accurately predict perturbation-response kinematics [12], our model's control law consists of linear feedback

gains to position and velocity error from each recovery strategy's mean kinematics. One can solve for these feedback gains by using a nonlinear optimization procedure. As in Sec. III-B, we use MATLAB's `fmincon` to solve for feedback gains on position and velocity that minimize the RMSE between simulated and ground truth training trials.

VI. TRAINING AND TESTING THE MODELS

Because we aim to test the significance of mean prediction errors across models, we select our training and testing datasets in accordance with the Combined 5×2 cv F Test framework [22]. This test has been shown to have higher power and lower false positives than previous statistical tests for comparing model performance. The training and testing process for each model is as follows:

- 1) Split each dataset into 50% testing, 50% training.
- 2) Fit the model using the training dataset.
- 3) For each test trial, use the trained model to simulate the trip-recovery hip states.
- 4) Compute the RMSE between predicted and ground truth trip-recovery trajectories. Normalize the RMSE for each response variable by the total range of response variable values in the test trial. Average these normalized RMSE (NRMSE) values across all three response variables.
- 5) Swap training and testing datasets; repeat steps 2)-4).
- 6) Repeat steps 1)-5) for five iterations.

For any two models, we compute an approximate f statistic to investigate whether the difference in NRMSE is statistically significant [22]. For the Combined 5×2 cv F Test, we conclude that the difference in mean NRMSE values between two models is statistically significant with 95% confidence if $f > 4.74$.

Individualized, data-driven models can produce accurate predictions of perturbation response, e.g. [23], but they necessitate collecting experimental (and often perturbative) data. Such experiments are infeasible for widespread use. We therefore conduct the Combined 5×2 cv F Test for both subject-specific and generalized cases. For the subject-specific case, we did not train and test models for strategies where subjects exhibited fewer than five trials (Subject 005 and 015's lowering trials and Subject 002's elevating trials). For the generalized case, we first normalize the hip heights and AP positions by subject leg length, then train and test the models using all subjects' data combined.

VII. RESULTS

Fig. 6 shows a box plot of the subject-specific model results. Due to the extreme values from the NARX model which often generated unstable predictions, we removed the outliers from all models for all following results and analyses. Table I presents the mean and standard deviation of NRMSE values for each strategy, model, and case across all subjects. In each strategy's column, the minimum NRMSE is highlighted for both subject-specific (darker shade) and generalized (lighter shade) models. Example prediction results for each method, plotted for one subject's delayed lowering trial, are shown in Fig. 7. Tab. II shows f statistics computed between each

TABLE I
AVERAGE NRMSE FOR ALL THREE PREDICTION METHODS, WITH LOWEST SUBJECT-SPECIFIC AND GENERALIZED ERRORS HIGHLIGHTED FOR EACH STRATEGY.

Prediction Method	Case	Elevating mean (stdev)	Delayed Lowering mean (stdev)	Lowering mean (stdev)
GPR Model	Subject-Specific	0.110 (0.0333)	0.119 (0.0522)	0.0567 (0.0266)
	Generalized	0.151 (0.0500)	0.160 (0.0630)	0.0879 (0.0362)
NARX Model	Subject-Specific	0.660 (0.482)	0.500 (0.360)	0.320 (0.333)
	Generalized	0.262 (0.119)	0.321 (0.171)	0.141 (0.124)
Pendulum Model	Subject-Specific	0.241 (0.0618)	0.285 (0.0792)	0.277 (0.0782)
	Generalized	0.254 (0.108)	0.262 (0.135)	0.264 (0.147)

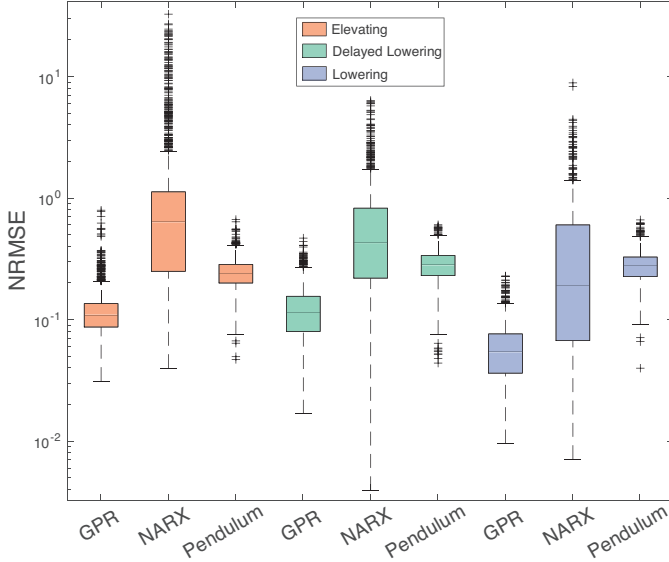


Fig. 6. Visualization of summary statistics for the subject-specific models. The central black line indicates the median NRMSE for each model; the bottom and top edges of each box are the 25th and 75th percentiles, respectively; the dashed whiskers extend to the limits of the data not considered outliers; and the outliers (determined using the interquartile range) are plotted individually as plus signs. The y-axis is a logarithmic scale due to the extreme magnitude of outliers.

model within the subject-specific and generalized cases and their corresponding p -values.

The mean computation time to generate conditional GPR predictions, NARX predictions, and dynamic model predictions was 2.67 ms, 8.81 ms, and 25.1 ms, respectively, on a computer with a 3.5 GHz Intel Core i7 processor. For the GPR models, we checked whether the 95% confidence bounds, shown in Fig. 7(a), encompassed each ground truth trajectory. For the subject-specific models, 77.3% of angle, 100% of height, and 75.1% of AP position trajectories remained inside the confidence bounds. For the generalized models, 56.5% of angle, 98.2% of height, and 67.9% of AP position trajectories remained inside the confidence bounds.

VIII. DISCUSSION AND CONCLUSION

Predicting human trip-recovery kinematics can help wearable lower-limb robots choose appropriate coordinating behavior. Previous studies that investigate predictive models for human kinematics report results for only one class of model [12]; use simulated trip data instead of human subject

TABLE II
APPROXIMATE f STATISTIC BETWEEN MODELS, WITH * INDICATING STATISTICAL SIGNIFICANCE, AND CORRESPONDING p -VALUES.

Models	Case	$f(p)$ Elev.	$f(p)$ Del. Low.	$f(p)$ Lower.
GPR and NARX	Subj.-Spec.	63.7* (<0.001)	82.6* (<0.001)	29.1* (<0.001)
	Gen.	10.4* (0.009)	7.98* (0.02)	4.08 (0.07)
GPR and Pendulum	Subj.-Spec.	275* (<0.001)	83.5* (<0.001)	531* (<0.001)
	Gen.	17.4* (0.003)	82.6* (<0.001)	167* (<0.001)
NARX and Pendulum	Subj.-Spec.	45.6* (<0.001)	18.4* (0.002)	1.43 (0.4)
	Gen.	2.70 (0.1)	2.04 (0.2)	15.0* (0.004)

experimental data [13]; or use unperturbed data, where less variation is likely present [11]. This paper investigates three modeling methods to predict swing hip kinematics in experimental trip data from 16 subjects. We restricted predictions to the sagittal plane, but we expect the models could be extended to include other hip motions (abduction-adduction, interior exterior rotation) in future work.

The subject-specific GPR produced the lowest mean error for each strategy out of all modeling methods. Across all strategies, we found that its NRMSE was statistically significant from that of the other two subject-specific models. The generalized GPR models produced the lowest mean NRMSE across generalized models, which was statistically significant from the NARX and pendulum models in all cases but one. Furthermore, the GPR model's lower computation time is useful for real-time trip-recovery control in wearable robots.

The GPR model's confidence bounds offer another advantage over the other prediction methods. Most ground truth trajectories remained within the 95% confidence bounds, though more often in the subject-specific case. Future work will further investigate the discrepancies between response variables, especially angle and height.

The subject-specific NARX models frequently generated unstable predictions, especially for trips that occurred earlier in the swing phase. We speculate that the considerably large error in elevating and delayed lowering trials is a result of

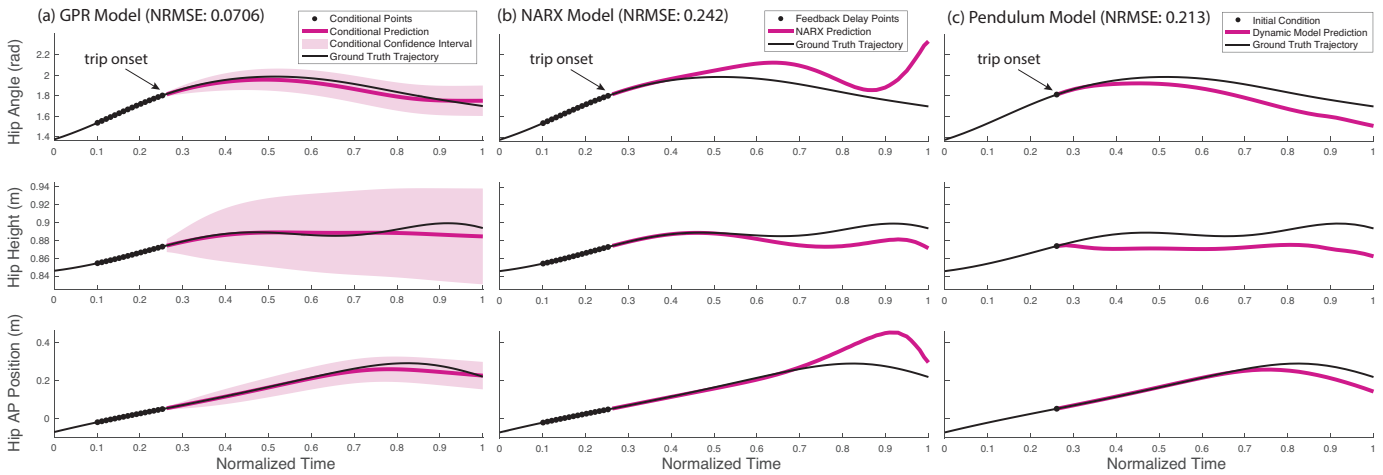


Fig. 7. An example of simulation results for one delayed lowering trial using the subject-specific (a) GPR, (b) NARX, and (c) pendulum models. The GPR model prediction in (a) includes a 95% confidence interval. The entire ground truth trial is shown in the black line, with the trip onset indicated by the arrow. The NRMSE for this trial, averaged across each response variable, is provided for each prediction method. The black dots represent the information provided to each model: 15 conditional points in (a); 15 feedback delay points in (b), and one initial condition for simulation in (c).

overfitting the network to relatively small datasets. Notably, the generalized NARX models' mean NRMSE decreased by approximately 50% from the subject-specific, likely due to larger training datasets. The dynamic pendulum model produced higher mean NRMSE than the GPR model, with differences in NRMSE statistically significant from GPR in both subject-specific and generalized cases. We found that of the three response variables, the hip height contributed most to this high mean NRMSE. Because the dynamic model applied feedback to mean kinematics, the high error could be explained by the relatively large variation in subjects' trip-recovery hip height trajectories.

Our results show the multi-output GPR model's promise for producing accurate predictions of kinematic responses to trips with low computation time and quantified prediction uncertainty. The generalized GPR models exhibited similar simulation accuracy to the subject-specific case, increasing the modeling method's real-time applicability. Although our experiment included only able-bodied subjects, individuals with lower-limb amputation have exhibited similar trip-recovery strategies in experiments [5].

REFERENCES

- [1] J. Kim, J. Wensman, N. Colabianchi, and D. Gates, "The influence of powered prostheses on user perspectives, metabolics, and activity: a randomized crossover trial," *Journal of Neuroengineering and Rehabilitation*, vol. 18, no. 1, p. 49, 2021.
- [2] L. N. Awad, J. Bae, K. O'Donnell, S. M. M. D. Rossi, K. Hendron, L. H. Sloot, P. Kudzia, S. Allen, K. G. Holt, T. D. Ellis, and C. J. Walsh, "A soft robotic exosuit improves walking in patients after stroke," *Science Translational Medicine*, vol. 9, no. 400, 2017.
- [3] F. Sup, A. Bohara, and M. Goldfarb, "Design and control of a powered knee and ankle prosthesis," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 4134–4139.
- [4] C. Shirota, A. M. Simon, and T. A. Kuiken, "Trip recovery strategies following perturbations of variable duration," *Journal of Biomechanics*, vol. 47, no. 11, pp. 2679–2684, 2014.
- [5] C. Shirota, A. M. Simon, and T. A. Kuiken, "Transfemoral amputee recovery strategies following trips to their sound and prosthesis sides throughout swing phase," *Journal of Neuroengineering and Rehabilitation*, vol. 12, 2015.
- [6] N. Thattai, N. Srinivasan, and H. Geyer, "Real-time reactive trip avoidance for powered transfemoral prostheses," in *Proceedings of Robotics: Science and Systems*, 2019.
- [7] B. E. Maki and W. E. McIlroy, "Control of rapid limb movements for balance recovery: age-related changes and implications for fall prevention," *Age and Ageing*, vol. 35, no. 2, pp. ii12–ii18, 09 2006.
- [8] A. Schillings, B. M. van Wezel, T. W. Mulder, and J. Duysens, "Muscular responses and movement strategies during stumbling over obstacles," *Journal of Neurophysiology*, vol. 83, pp. 2093–102, 2000.
- [9] M. E. EVELD, S. T. King, L. G. Vailati, K. E. Zelik, and M. Goldfarb, "On the Basis for Stumble Recovery Strategy Selection in Healthy Adults," *Journal of Biomechanical Engineering*, vol. 143, no. 7, 04 2021.
- [10] M. Gholami, C. Napier, and C. Menon, "Estimating lower extremity running gait kinematics with a single accelerometer: A deep learning approach," *Sensors*, vol. 20, no. 10, 2020.
- [11] A. Zaroug, A. Garofolini, D. T. H. Lai, K. Mudie, and R. Begg, "Prediction of gait trajectories based on the long short term memory neural networks," *PloS One*, vol. 16, no. 8, 2021.
- [12] V. Joshi and M. Srinivasan, "A controller for walking derived from how humans recover from perturbations," *Journal of The Royal Society Interface*, vol. 16, no. 157, 2019.
- [13] B. Miripour Fard, "A comparison between virtual constraint-based and model predictive-based limit cycle walking control in successful trip recovery," *AUT Journal of Mechanical Engineering*, vol. 4, no. 2, pp. 169–182, 2020.
- [14] "Nexus," Vicon, Oxford, UK.
- [15] "MATLAB Statistics and Machine Learning Toolbox," 2017, the MathWorks, Natick, MA, USA.
- [16] M. Ebden, "Gaussian processes: A quick introduction," 2015.
- [17] C. E. Rasmussen and W. C. K. I., *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2008.
- [18] B. Wang and T. Chen, "Gaussian process regression with multiple response variables," *Chemometrics and Intelligent Laboratory Systems*, vol. 142, pp. 159–165, 2015.
- [19] J. Sum, W. K. Kan, and G. H. F. Young, "A note on the equivalence of narx and rnn," *Neural Computing & Applications*, vol. 8, pp. 33–39, 1999.
- [20] Mathworks. (n.d.) Narxnet: Nonlinear autoregressive neural network with external input.
- [21] D. A. Winter, *Biomechanics and Motor Control of Human Movement*. Hoboken, NJ: John Wiley & Sons, 2005.
- [22] E. Alpaydin, "Combined 5 x 2 cv f test for comparing supervised classification learning algorithms," *Neural computation*, vol. 11, no. 8, p. 1885–1892, 1999.
- [23] P. D. Holmes, S. M. Danforth, X.-Y. Fu, T. Y. Moore, and R. Vasudevan, "Characterizing the limits of human stability during motion: perturbative experiment validates a model-based approach for the sit-to-stand task," *Royal Society Open Science*, vol. 7, no. 1, 2020.