Policy Design for an Ankle-Foot Orthosis Using Simulated Physical Human–Robot Interaction via Deep Reinforcement Learning

Jong In Han, Jeong-Hoon Lee, Ho Seon Choi,
Jung-Hoon Kim, and Jongeun Choi, *Member, IEEE*

***Abstract*—This paper presents a novel approach for designing a robotic orthosis controller considering physical human-robot interaction (pHRI). Computer simulation for this human-robot system can be advantageous in terms of time and cost due to the laborious nature of designing a robot controller that effectively assists humans with the appropriate magnitude and phase. Therefore, we propose a two-stage policy training framework based on deep reinforcement learning (deep RL) to design a robot controller using human-robot dynamic simulation. In Stage 1, the optimal policy of generating human gaits is obtained from deep RL-based imitation learning on a healthy subject model using the musculoskeletal simulation in OpenSim-RL. In Stage 2, human models in which the right soleus muscle is weakened to a certain severity are created by modifying the human model obtained from Stage 1. A robotic orthosis is then attached to the right ankle of these models. The orthosis policy that assists walking with optimal torque is then trained on these models. Here, the elastic foundation model is used to predict the pHRI in the coupling part between the human and robotic orthosis. Comparative analysis of kinematic and kinetic simulation results with the experimental data shows that the derived human musculoskeletal model imitates a human walking. It also shows that the robotic orthosis policy obtained from two-stage policy training can assist the weakened soleus muscle. The proposed approach was validated by applying the learned policy to ankle orthosis, conducting a gait experiment, and comparing it with the simulation results.**

***Index Terms*—Computer simulation, exoskeletons, human-robot interaction, orthotics, reinforcement learning.**

Manuscript received 21 January 2022; revised 30 June 2022; accepted 29 July 2022. Date of publication 4 August 2022; date of current version 11 August 2022. This work was supported in part by the National Research Foundation of Korea (NRF) through the Korea Government [Ministry of Science and ICT (MSIT)] under Grant 2021R1A2B5B01002620 and in part by the Yonsei-Korea Institute of Science and Technology (KIST) Convergence Research Program. *(Corresponding author: Jongeun Choi.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Yonsei University under IRB No. 7001988-202205-HR-1560-02.

Jong In Han, Jeong-Hoon Lee, and Jongeun Choi are with the School of Mechanical Engineering, Yonsei University, Seoul 03722, South Korea (e-mail: gmiilp1318@yonsei.ac.kr; ljh_0921@yonsei.ac.kr; jongeunchoi@yonsei.ac.kr).

Ho Seon Choi is with the Center for Healthcare Robotics, Korea Institute of Science and Technology, Seoul 02792, South Korea (e-mail: ghtjs3607@gmail.com).

Jung-Hoon Kim is with the School of Civil and Environmental Engineering, Yonsei University, Seoul 03722, South Korea (e-mail: junghoon@yonsei.ac.kr).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2022.3196468, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2022.3196468

I. Introduction

ASSISTIVE robots have been developed to support various types of human body movement [1], [2]. Specifically, lower limb exoskeleton robots provide stability and support during gait cycles [3] and are widely used in the field of rehabilitation [4]–[8]. Because the lower limb exoskeleton robot operates in conjunction with the human body, the effect of the robot on the human body must be analyzed for effective assistance [9]. The assistive performance of exoskeleton robots on the human body is often evaluated through laborious experiments. Thus, kinematics/kinetics simulation of a human-robot system can be advantageous by minimizing the number of experiments required to verify the effect of the assistive device on human bodies [3], [10]. To show the credibility of controlled motion in the simulated human-robot interaction, it is crucial to evaluate whether the controller is designed in a plausible manner [11]. Simulations can be made more reliable by using controllers designed in a manner similar to how humans generate motion. Humans use sensory feedback during gait [12]. A neuromechanical controller resembling the structure of human nervous system has been used for controlling muscle-driven human models [13], [14], which can be utilized in the human-robot simulation.

The human-robot system has a structural feature in which the human body and the robot contact at physical interfaces such as a straps, through which interaction forces and torques are transmitted, referred to as the physical human-robot interaction (pHRI). The pHRI can be used to assess motor control ability during balancing [15]. The pHRI can provide important information about the behavior of an exoskeleton, such as assessing motor control ability during balancing [15], or predicting comfort of the interface based on direction and magnitude of forces which may loosen the attachment or cause pain to the skin. The pHRI can be measured using force sensors [16], [17] or spring-based distance measurement [18] at the robot/human interface, or by using complex dynamic models [19].

In designing control policy for assistive devices, the human-robot simulation must be exploited to consider pHRI efficiently. Zhang *et al.* performed human-robot simulations including pHRI to test assistive strategies of knee exoskeletons [20]. In the simulation study of the knee exoskeleton, pHRI increased further when assistive force was applied. Ankle orthosis has a different target joint compared to the knee exoskeleton. Therefore, the assistive timing or magnitude will differ, and different aspects of pHRI will appear. Vree *et al.* designed a controller for a human model wearing a prosthesis through deep learning-based human-robot simulation [21]. However, there has been less effort to design a patient-specific ankle-foot orthosis controller for muscle weakened patients considering pHRI using human-robot simulation.

Therefore, our study focuses on designing a control policy for an ankle-foot orthosis using simulated pHRI. We will show how to design a robotic orthosis controller that assists a patient with weakened plantar flexor through human-robot simulation. The simulated pHRI can be used to produce the coordination and its resulting interaction forces between a human model controller and a robotic orthosis controller.

Previous studies related to human musculoskeletal simulation used static optimization or Computed Muscle Control (CMC) tools [22]–[24] to calculate muscle force or muscle excitations satisfying given kinematic data. The optimization-based problems were able to calculate the muscle force that meets the given kinematic and external force data. However, if conditions change, such as a slight initial change of posture, the problem must be re-solved under that setting. Unlike optimization, if we use deep RL to solve the biomechanics problem, we get the policy which is the system's controller. Using the policy, we may obtain a robust solution for different conditions such as slight initial attitude changes.

Previous studies related to human-robot integrated simulation have designed assistive control strategies [20], [25]–[27], and investigated changes in recruitment and force of individual muscles according to robot assistance [22], [23]. These studies have used prescribed kinematic data [20], [22], [23], reference kinematic data tracking using proportional and derivative (PD) controller [26], or simple mass models [25] to drive the human musculoskeletal system. When only prescribed kinematic data is used, it is difficult to simulate kinematic adaptation, and manual effort, including system modeling, is required when designing a controller. Although such methods can generate the gait motion of a human body model, it will not be a simulation in a plausible manner because it does not design a neuromechanical controller based on sensory feedback like an actual human. Humans control gait through neural circuits that use sensory feedback [28].

Studies have considered a neuromechanical controller design to control human models, in a similar manner to humans [13], [14]. The neuromechanical controller, which generates gait motion of the human body from sensory feedback similar to the human nervous system, includes reflex-based control and central pattern generator (CPG). Neuromechanical control enabled the generation of human-like motions in terrain such as slopes and curves [13] and adaptation of motion against perturbations [14]. Most neuromechanical simulation studies have manually designed controllers based on prior knowledge of human motion. Recently, studies have been conducted on designing a controller that uses sensory feedback based on deep reinforcement learning (deep RL) to drive a muscle-driven human model [11], [29]. Although deep artificial neural networks (ANNs) used to learn the human nervous system have a simpler structure than real neural networks, neural controllers designed based on deep RL can aid in studies related to human motor control [11], [30].

Studies related to simulating pHRI have used spring, and damper elements [31] or artificial muscles [20] to model the interaction forces generated in a human-robot system. However, the elastic foundation model (EFM), which simulates contact between arbitrarily complex surface geometries represented by meshes [32], can also be considered to model such interaction force. EFM tends to overestimate the contact pressure compared to the finite element model because it calculates the deformation and force using a simplified elastic model. However, owing to its computational cost-effectiveness, EFM has been widely used in the study of contact simulation of knee joint bones [33].

The contributions of our paper are as follows. First, we introduce EFM to predict pHRI, i.e., the interaction force between the human and the robotic orthosis, and develop a suitable model for human-robot simulation where relative motion and interaction forces occur continuously. Although several human-robot simulation studies have been conducted [20], [31], those studies model the interaction force as force occurring at several specified points. As our proposed method using EFM models the contact between object surfaces, contact force can be generated at any point on the object surface. Our deep RL-based orthosis controller design builds on this EFM. Second, we propose two-stage policy training based on reinforcement learning (RL) to design a robot controller that can assist patients. In particular, the human model controller is designed in advance. The objective of the human controller is to generate a human-like gait through sensory feedback [34]; it is then used to simulate a human model with weakened soleus. The goal of the robot controller is to make a human model with a weakened soleus walk normally with robotic orthosis assistance. The observation used in designing the robot controller consists of human and robot's joint angles, angular velocities, and ground reaction forces (GRFs) that can be measured in real robots. Therefore, the robot controller can be used to control the real robot through sim-to-real transfer [35]. We validated the kinematic/kinetic data generated by each designed policy by comparing it with the experimental data. Finally, we analyze changes in ankle moments and assistive strategies of an orthosis policy pertaining to soleus weakness. Our analysis reveals the impact of soleus weakness on human gait and the intention of robotic orthosis policies to support the human. Our framework could be adapted to design a patient-specific robotic assistive device using the patient's data.

The remainder of this paper is organized as follows. Section II provides preliminaries for deep RL and imitation learning. Section III introduces musculoskeletal simulation

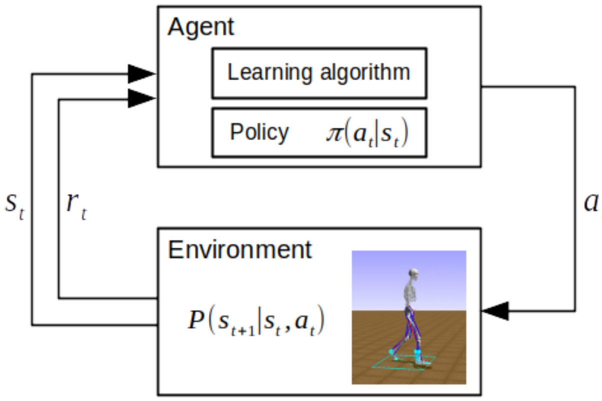

Fig. 1. **Overview of reinforcement learning with OpenSim-RL.** In reinforcement learning, the agent aims to find a policy π to maximize the expected sum of discounted rewards of an MDP. In OpenSim-RL, OpenSim provides an environment for reinforcement learning: a musculoskeletal model and a physics-based simulation environment [11].

tools, two-stage policy training, the musculoskeletal model used in the simulation, the reward function, and the gait experiment for validation. In Section IV, we present and discuss results for the forward dynamics simulation using policies trained at each stage. In addition, we compare and discuss the results of gait experiments and simulations for validation. Section V provides concluding remarks.

II. Preliminaries

In the human-robot simulation, deep RL and imitation learning methods were used to learn a policy for generating human gait motions and an robotic orthosis policy to support the patient model. This section describes the methods used.

A. Deep Reinforcement Learning

We assume our environment to be a Markov decision process (MDP), a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$ consisting of sets of states and actions, system dynamics, a reward function, and a discount factor, respectively. An MDP includes a policy π, which governs the decision-making process. We denote the policy as $\pi(a_t|s_t)$, as it is regarded a conditional probability distribution of an action $a_t \in \mathcal{A}$, given a state $s_t \in \mathcal{S}$ at time t, as shown in Fig. 1. At each time step, an action a_t is sampled from the π under s_t and applied to the environment to induce the next state s_{t+1} following the system dynamics, $\mathcal{P}(s_{t+1}|s_t, a_t)$, along with the corresponding scalar reward $r_t = \mathcal{R}(s_t, a_t)$. Our task is then formalized to a standard RL problem for obtaining for the optimal policy π^* such that maximizes the expected sum of discounted rewards of an MDP.

RL attempts to recover such a solution by iteratively evaluating and improving the policy with trajectory samples obtained from the past and/or the current policy [36]. Therefore, sample efficiency should be considered first when the simulation cannot be run faster than in real-time, such as in an environment simulating with a musculoskeletal model. We use Soft-Actor-Critic (SAC) that has shown superior performance regarding data-expensive environments where the agent has to interact with a real-world, or a simulation without acceleration capabilities [37]. Unlike other optimization methods, SAC modifies the conventional policy objective by augmenting rewards with additional policy entropy as follows:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\tau\sim p(\tau_0)}\left[\sum_{t=0}^{\infty}[\gamma^t r_t + \mathcal{H}(\pi(\cdot|s_t))]\right] \qquad (1)$$

where $p(\tau_0) = p(s_0)\prod_{t=0}^{\infty}\mathcal{P}(s_{t+1}|s_t, a_t)\pi(a_t|s_t)$ denotes the distribution over trajectories starting from the initial state s_0 under the policy π, and $\mathcal{H}(\pi)$ denotes the entropy of the policy π which measures the average amount of uncertainty intrinsic to π. Consequently, the exploration of the stochastic policy and robustness to the additional perturbations residing in the model are improved.

B. Imitation Learning

Although deep RL has shown promising performances in achieving the solution in an MDP, maximizing the policy objective does not guarantee the optimal policy to behave naturally like a real human. Furthermore, it is known to be notorious to obtain proper learning signals under sparsely rewarded environments where the non-zero reward is only given as an indication of successful termination, known as the *credit-assignment problem* [36]. Imitation learning tackles these two problems by leveraging reference trajectories of the expert [38].

By learning from the reference, imitation learning can be branched out into three categories: a) behavior cloning, b) inverse reinforcement learning (IRL), and c) auxiliary reward learning. a) Behavior cloning is a learning method using state-action pairs obtained from expert demonstrations [39], [40] applied to autonomous driving [41] or manipulator operations [42]. b) IRL aims to learn a maximized cost function following the reference trajectory. Albeit IRL methods can avoid covariate shift problems caused by fitting single-timestep decisions, these are computationally costly due to the RL process in an inner loop [43], [44]. A Gaussian process regression is used to predict a reward function with a small number of expert demonstrations [45]. c) Auxiliary reward learning is an example-guided learning framework proposed by combining a reference imitation objective with a task objective [46]; the study showed that the proposed auxiliary reward term could reduce the optimal policy of natural-looking behaviors. Accordingly, we leveraged reference imitation objective into our work, both in the human policy and orthosis policy training stages.

III. Methods

This section describes the musculoskeletal simulation tools used, human and robotic orthosis models, and detailed methods to learn policies for human-robot simulations. In addition, the process of performing a gait experiment on a human subject is described to verify the applicability of the learned orthosis policy to actual orthosis and the validity of simulated pHRI.

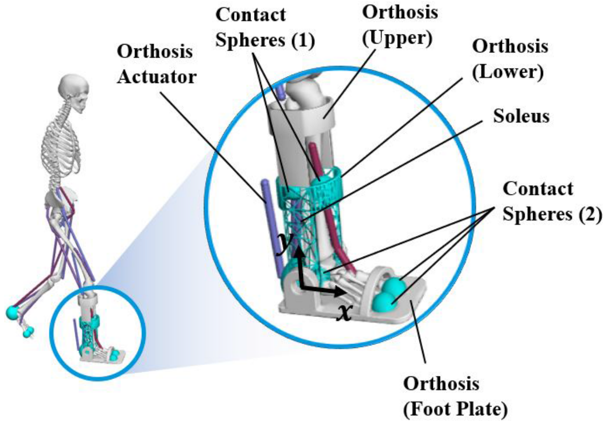

Fig. 2. **Human-robot model used in the simulation.** We use a musculoskeletal model that moves in the sagittal plane. pHRI is generated by the EFM between two contact spheres (Contact Spheres (1)) located in the shank of the human model and an orthosis part modeled as a mesh (Orthosis (Lower)). Orthosis (Upper) is attached to the upper part of the human shank. The x, y coordinate system is shown to represent the direction of pHRI and is a local coordinate system on orthosis. For the GRF, three contact spheres are used on the foot of the human body model (Contact Spheres (2)), and the forces are generated by the Hunt-Crossley contact model.

A. OpenSim-RL

We used OpenSim-RL [11] to learn policies for human-robot simulation. OpenSim-RL, built using OpenSim [47] and OpenAI Gym [48], can train human motor controllers using the deep RL method [29], which is provided as a Conda package [11]. RL aims to train an agent to perform actions that maximize rewards in its environment (Fig. 1). The RL environment we used in this study comprises a musculoskeletal model and a physics-based simulation environment in the OpenSim-RL.

B. Model

We used the ‘gait10dof18musc’ model distributed with OpenSim as the human musculoskeletal system model after modification. We removed the degree of freedom of the lumbar joint to use only the hip, knee, and ankle joints in the model. The model was scaled using 3-dimensional marker set data distributed with OpenSim and had a height of approximately 1.8 m, a mass of 72 kg, and 9 degrees of freedom. The left and right hip, knee, and ankle joints have 1 degree of freedom. Pelvic translation and pelvic rotation have 2 degrees of freedom and 1 degree of freedom, respectively. The model moves in the sagittal plane, and the degrees of freedom in other directions are constrained. Hill-type muscle, widely used in muscle-driven simulations, is used to model the relationship between muscle length, velocity, and force [49], [50]. Eighteen Hill-type muscles attached to the lower extremities were used for movement. The maximum isometric force of the soleus was weakened to create the patient model (Soleus in Fig. 2). A one-degree-of-freedom robotic orthosis consists of three parts. One is a part fixed to the foot of the model (Orthosis (Foot Plate) in Fig. 2), and another (Orthosis (Lower) in Fig. 2) can rotate relative to the fixed part. The other one (Orthosis (Upper) in Fig. 2) was attached to the top of the shank. WeldConstraint constrains the Orthosis (Upper) and the Orthosis (Lower). Bushing Force was added to the orthosis ankle joint and the Orthosis (Upper).

TABLE I
CONTACT PARAMETERS USED IN SIMULATION

Parameter	Hunt–Crossley	EFM
Stiffness (N/m^2)	3067776	10^9
Dissipation (s/m)	2.0	5.0
Static friction coefficient	0.8	0.7
Dynamic friction coefficient	0.8	0.5
Viscous friction coefficient	0.5	0.3
Transition velocity (m/s)	0.2	0.2

The Hunt–Crossley contact model in OpenSim [32] was used to generate the ground reaction force between the human foot and the ground during model walking simulation. Contact spheres, one on the heel and two on the forefoot, were used to measure ground reaction forces (Contact Spheres (2) in Fig. 2). To predict the pHRI, we used EFM in OpenSim, which is computationally more efficient than the finite element model [32], [33]. The pHRI between the orthosis mesh (Orthosis (Lower) in Fig. 2) and the contact spheres located on the shank (Contact Spheres (1) in Fig. 2) was generated using EFM. The parameters for the Hunt–Crossley contact model were similar to those used in the study of Falisse *et al.* [51]. The parameters of the EFM are empirically determined within the range used in the study of Hast *et al.* [33]. Detailed values are shown in Table I.

The meshlab software [52] was used to convert the CAD file of the orthosis to a mesh file with a triangular mesh grid recognizable in OpenSim, which is required in the EFM contact model. The orthosis actuator was modeled as a Hill-type muscle, and muscle activation dynamics were used. We neglected the weight of the actuator. The Foot Plate, Orthosis (Upper), and Orthosis (Lower) mass is approximately 0.84 kg, 0.38 kg, and 0.29 kg, respectively.

Reference motion data for imitation learning was obtained from normal level walking data included in OpenSim. The gait data were obtained on the self-selected speed gait of one healthy subject [53]. The data were obtained for one gait cycle and extended to three cycles using cubic spline interpolation to make smooth cyclic gait data. The joint angle and angular velocity were obtained using inverse kinematics. Muscle excitation data which will be used as reference data in the imitation learning, were obtained through inverse dynamics, Residual Reduction Algorithm, and CMC tools in OpenSim [23], [24]. This study uses the deep RL-based imitation learning method to learn human policy. Unlike the existing prescribed kinematic data research that requires GRF data, only kinematic data is enough as reference data for policy learning since the GRF can be estimated through the contact spheres on the musculoskeletal model's feet.

C. Two-Stage Policy Training

To train the orthosis policy in a patient-specific fashion, we propose a two-stage policy training based on RL with imitation objectives to derive the human and the orthosis policies to be used as the controllers in the human-robot simulation (Fig. 3). In Stage 1, the human policy (π^{human})

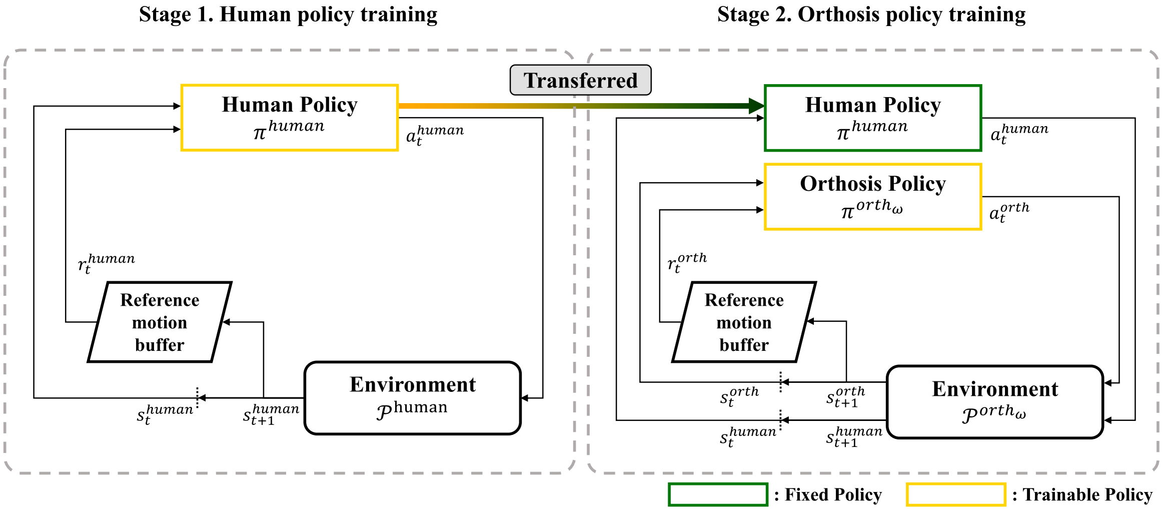

Fig. 3. **Overall schematic of the two–stage policy training framework.** In Stage 1, the human policy (π^{human}) is trained in a way that minimizes the kinematic and muscle activation differences between the current model and the reference data. The observations (s_t^{human}) of π^{human} are muscle activations, joint angles, angular velocities and GRFs states of the human model, and the action (a_t^{human}) is muscle activations. The musculoskeletal model used in this stage is a healthy subject ($\mathcal{P}^{human}$). In Stage 2, we weaken the right soleus of the musculoskeletal model by ω% and attach a ankle orthosis to the ipsilateral side ($\mathcal{P}^{orth_\omega}$). While the transferred policy from Stage 1 drives the musculoskeletal model, the orthosis policy (π^{orth_ω}) is trained to output the required torque for the orthosis to assist the weakened model remain in balance. The observations (s_t^{orth}) of π^{orth_ω} are human right ankle joint angle, angular velocity, orthosis joint angle and angular velocity, and GRFs states, and the action (a_t^{orth}) is a muscle activation.

for generating human gaits was trained, and in Stage 2, the orthosis policy (π^{orth_ω}) for the orthosis assistance was trained. In Stage 2, we reduce the maximum isometric force of Soleus to model the patient and use the human policy learned from healthy gait data to drive the human model. At this time, we assumed a patient who could perform a healthy gait when wearing an ankle orthosis.

In Stage 1, the human policy was trained to imitate the human reference trajectories (Stage 1 in Fig. 3). The reference trajectories can be replaced with the patient's gait to learn the patient-specific gait. The agent receives the states, i.e., the 37-dimensional vector consist of muscle activations, joint angles, angular velocities, and GRFs from the environment. The agent then sends muscle activations as an action with values between 0 and 1. The action forms an 18-dimensional continuous action space. Environment of Stage 1 receives action a_t^{human}, performs forward dynamics for a given time step, and sends observation s_{t+1}^{human}. A healthy subject is used as a human model, and the system dynamics can be represented as $\mathcal{P}^{human}(s_{t+1}^{human}|s_t^{human}, a_t^{human})$. π^{human} has an input/output structure similar to human sensory feedback. π^{human} uses muscle, joint, and GRF states as input and muscle activation as output.

In Stage 2, the orthosis policy is trained for optimal torque of an orthosis to support the human in walking (Stage 2 in Fig. 3). Referring to Ong *et al.*, we generated three models by weakening the maximum isometric force of the right Soleus muscle to 75% (mild), 87.5% (moderate), and 93.75% (severe), respectively [54]. In addition, a one-degree-of-freedom orthosis was added to the right ankle to support the soleus weakened patient model. EFM was added between the orthosis and the human model to generate pHRI (Fig. 2). The system dynamics of Stage 2 can be represented as $\mathcal{P}^{orth_\omega}(s_{t+1}^{orth_\omega}|s_t^{orth_\omega}, a_t^{orth_\omega})$ where ω is 75, 87.5, and 93.75 which indicates the percentage with respect to the maximum isometric force of the healthy model. Imitation learning using reference motion for the joint angle and angular velocity of the human right ankle was used to train π^{orth_ω}. π^{human} derived from Stage 1 was used as a fixed policy to control the weakened musculoskeletal system in Stage 2. Stage 2 agent does not update π^{human} during the training. In Stage 2, 19-dimensional actions are input to the human-robot model, of which 18 are actions to drive human muscles, and one is to input orthosis actuators. As the learning policy is not a π^{human} but an π^{orth_ω}, the action space size of Stage 2 is set to 1. While the soleus weakened musculoskeletal model is controlled by π^{human}, the orthosis policy π^{orth_ω} for assistance is trained by Stage 2 agent. An RL agent for π^{orth_ω} training receives the following observations from the environment: human right ankle angle, human right ankle angular velocity, robot joint angle, robot joint angular velocity, and GRFs. The RL agent sends an action a_t^{orth} to the environment to maximize reward r_t^{orth}. The action is the muscle activation, which has a value between 0 and 1. The environment of Stage 2 takes the actions of the fixed policy a_t^{human} and the trainable policy a_t^{orth} as inputs. It then performs integration for a given time step and outputs the observation. The observation consists of s_{t+1}^{human} and s_{t+1}^{orth} for the human and the orthosis policies, respectively. π^{orth_ω} receives joint kinematics data and GRFs, which are measurable values by sensors in ankle-foot orthosis. π^{orth_ω} outputs torque as action. We use SAC as an RL algorithm for two-stage policy training. The hyperparameters of the SAC used in each stage are presented in Table II. The values of the parameters were selected empirically.

TABLE II
HYPERPARAMETERS OF THE SAC LEARNING ALGORITHM USED IN TWO-STAGE POLICY TRAINING

Parameter	**Value (Stage 1)**	**Value (Stage 2)**
Optimizer	Adam	Adam
Learning rate	lin7.3×10^{-4}	lin7.3×10^{-4}
Discount (γ)	0.98	0.99
Number of hidden layers	2	2
Number of hidden units (layer 1)	400	16
Number of hidden units (layer 2)	300	16
Nonlinearity	ReLU	ReLU
Target smoothing coefficient (τ)	0.02	0.01
Train frequency	64	32
Gradient steps	64	32
Entropy	Auto	Auto
Batch size	256	256
Buffer size	300000	50000

TABLE III
NUMERICAL VALUES OF WEIGHTS USED IN REWARD FUNCTIONS

	Human Policy Reward				**Orthosis Policy Reward**		
Weights	$w_{pos,h}$	$w_{vel,h}$	w_{pvel}	w_{ma}	$w_{pos,o}$	$w_{vel,o}$	w_{sd}
Values	0.6	0.05	0.15	0.2	0.8	0.2	-10

D. Reward Function

1) Human Policy Reward: For constructing the reward function for imitation learning, we refer to the method used in [55]. The reward function r_t^h used in π^{human} training is as follows:

$$r_t^h = w_{pos,h} r_t^{pos,h} + w_{vel,h} r_t^{vel,h} + w_{pvel} r_t^{pvel} + w_{ma} r_t^{ma}, \tag{2}$$

where $w_{pos,h} = 0.6$, $w_{vel,h} = 0.05$, $w_{pvel} = 0.15$, and $w_{ma} = 0.2$. The r_t^{pos} is set as follows to give a high reward when the j^{th} joint angle q_t^j of the human model and the joint angle $\hat{q}_t^j$ of the reference motion are close at time t.

$$r_t^{pos,h} = \exp\left[-10 \sum_j (\hat{q}_t^j - q_t^j)^2\right] \tag{3}$$

The term r_t^{vel} makes the joint angular velocity of the model $\dot{q}_t^j$ track the reference data $\hat{\dot{q}}_t^j$.

$$r_t^{vel,h} = \exp\left[-0.5 \sum_j (\hat{\dot{q}}_t^j - \dot{q}_t^j)^2\right] \tag{4}$$

The r_t^{pvel} encourages the model to walk with the speed of the reference data. $\hat{\dot{p}}_t$ and $\dot{p}_t$ are the forward pelvic velocities of the reference data and model, respectively.

$$r_t^{pvel} = \exp\left[-50(\hat{\dot{p}}_t - \dot{p}_t)^2\right] \tag{5}$$

The r_t^{ma} encourages the muscle activations of the musculoskeletal model to track the reference data. $\hat{x}_t^j$ and x_t^j are the muscle activations of muscle j in the reference data and model, respectively.

$$r_t^{ma} = \exp\left[-2 \sum_j (\hat{x}_t^j - x_t^j)^2\right]. \tag{6}$$

We used Eq. (6) to track muscle activations obtained from the CMC tool under given kinematics to accelerate the imitation learning process.

2) Orthosis Policy Reward: The reward function r_t^o used in π^{orth_ω} training is as follows:

$$r_t^o = w_{pos,o} r_t^{pos,o} + w_{vel,o} r_t^{vel,o} + w_{sd} r_t^{sd}, \tag{7}$$

where $w_{pos,o} = 0.8$, $w_{vel,o} = 0.2$, and $w_{sd} = -10$. The r_t^o provides a high value when the ankle angle and angular velocity of the weakened model follow the reference data using the ankle-foot orthosis. $r_t^{pos,o}$ generates a high reward when the right ankle joint angle $q_t^{ankle,r}$ of the human model matches the joint angle $\hat{q}_t^{ankle,r}$ of the reference motion at time t.

$$r_t^{pos,o} = \exp\left[-100(\hat{q}_t^{ankle,r} - q_t^{ankle,r})^2\right] \tag{8}$$

Similarly, the joint angular velocity reward term is provided. The $\hat{\dot{q}}_t^{ankle,r}$ and $\dot{q}_t^{ankle,r}$ are the angular velocity of the right ankle joint from the reference data and the human model, respectively.

$$r_t^{vel,o} = \exp\left[-2 \sum_j (\hat{\dot{q}}_t^{ankle,r} - \dot{q}_t^{ankle,r})^2\right] \tag{9}$$

$r_t^{sd,o}$ is added for smoothing of orthosis torque input. It generates a high reward when minimizing the variance of the last ten orthosis actions. $a_{o,j}$ represent the j^{th} previous input from the current time step among the recent N orthosis actions, and a value of $N = 10$ was used.

$$r_t^{sd} = \exp\left[\frac{1}{N-1} \sum_{j=0}^{N-1} (a_{o,j} - \bar{a}_{o,N-1})^2\right] \tag{10}$$

E. Gait Experiment

To show the feasibility of whether orthosis policy can be applied to an actual ankle orthosis, we fabricated an ankle orthosis and performed a gait experiment on a healthy subject. In the experiment, the ankle orthosis was controlled using the learned orthosis policy.

We fabricated the ankle orthosis by modifying a part of the 1-degree of freedom ankle orthosis design used in the study of Choi *et al.* [56] (Fig. 4). The ankle orthosis has one degree of freedom in the direction of dorsi-plantar flexion. A pneumatic actuator drives ankle orthosis (Pneumatic Actuator in Fig. 4). The actuator is connected to the heel of the orthosis through a steel wire and assists in plantar flexion during contraction. The load cell attached to the pneumatic actuator measures the force when the actuator contracts (Loadcell in Fig. 4). The encoder measures the ankle joint angle (Encoder in Fig. 4). Part 1 (Part 1 in Fig. 4) and Part 2 (Part 2 in Fig. 4) are connected via the F/T sensor (F/T Sensor in Fig. 4), a sensor

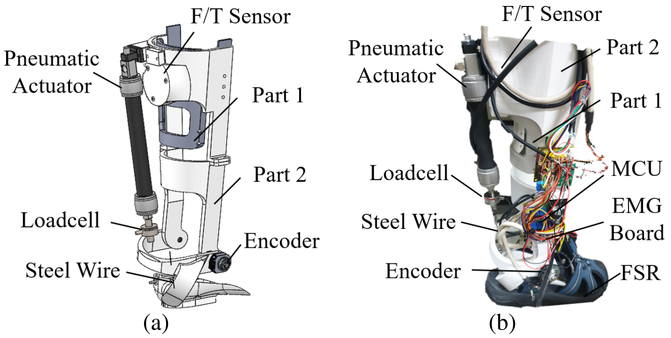

Fig. 4. **Design and fabrication of an ankle orthosis for gait experiment.** (a) Orthosis CAD Model (b) Fabricated Ankle Orthosis.

that measures 3-axis force and torque. In other words, when a relative force or torque occurs between Part 1 and Part 2, it can be measured through the F/T sensor. Part 2 is connected to the actuator, and Part 1 is engaged with the human shank. If a force is applied from the actuator, pHRI between the human and robot can be measured through the F/T sensor. Force sensing resistors (FSR) for measuring vertical GRF are attached to the shoe insole (FSR in Fig. 4) [7]. To obtain the EMG of the plantar flexor, we measured the EMG of the gastrocnemius, a muscle that can be measured by surface EMG (EMG board in Fig. 4). The measured sensor data is collected through a microcontroller unit (MCU) (MCU in Fig. 4). For data transmission, we used Controller Area Network (CAN) communication.

The experiment was performed on one healthy subject with a height of 1.8 m and a weight of 83 kg. The test subject walked on the treadmill at 1.2 m/s. The experiment was performed in 3 sets of 50 steps each for both cases with and without orthosis assistance. The experimental protocol was approved by the institutional review board of Yonsei University (7001988-202205-HR-1560-02). We performed a gait experiment using the orthosis policy of Stage 2 to validate the policy. We verified the relevance of each feature on the action by alternately masking each index to 0. It has been shown that some of the features dominate the action mapping. For example, in the case of the moderate orthosis policy, 2 out of 8 observations - ankle angle and vertical ground reaction force - had a significant effect on action. From that feature, we used the values measured from the FSR and encoder as observations. The experimental results were then compared with the simulation results using a model where Soleus was weakened by 20%.

IV. Results and Discussion

To evaluate the results of our two-stage policy training, forward dynamics simulation are performed using policies trained at each stage, namely π^{human} and $\pi^{orth_{\omega}}$. The kinematic and kinetic results of the simulation are compared with the reference motion. It is also compared with experimental data gathered by motion capture system of healthy subjects [57] for validation. The results of the gait experiment performed using the manufactured ankle orthosis with the orthosis policy are presented together with the simulation results. We have included a supplementary file that contains eight multimedia MOV format movie clips, which show forward dynamic simulation results for the two stages and the robustness of the learned policies.

A. Stage 1 Training

First, Fig. 5 shows the forward dynamics simulation results for Stage 1, where the healthy subject model is controlled by π^{human}. Joint angles, moments, and GRFs are shown to evaluate kinematics and kinetics of gait generated by π^{human}. The left and right joint angles generated by π^{human} are similar to the reference data given as the goal of imitation learning. The joint angles are also located approximately within 2 standard deviations (σ) of the experimental data [57] (Fig. 5a). Data falling outside of the 2σ range of the experimental data and some oscillations are observed in the joint moment results (Fig. 5b). The possible cause of the discrepancy includes the dynamic inconsistency between the reference data used for imitation learning and the RL environment. We used the muscle activation term in the reward function to make the muscle activation pattern of human gait be learned.

The reference data used for imitation learning was obtained from the CMC tool in OpenSim. The musculoskeletal model and environment used in the CMC are different from those used in π^{human} training. For example, CMC inputs external forces from recorded data to apply GRF to the model and use reserve actuators to drive the model. On the other hand, the model in Stage 1 uses a contact sphere and a fixed lumbar angle, respectively. These differences would make it difficult for the agent to train a policy to generate data consistent with the reference muscle activation, resulting in differences from the reference data in joint moments.

The simulation and reference data for GRF show a similar trend except for certain points: gait cycles of 1.9% to 5.6% in vertical R, 1.3% to 3.9% in horizontal R, and 75% to 80% in vertical and horizontal L. GRF is a state predicted through forward simulation using π^{human}, not a state included in the reward function for imitation learning. Nevertheless, it shows good agreement with the experimental data except for these points. The dynamic inconsistency between the reference data and the deep RL environment might be the cause of the high peak values or oscillation. A deep RL-based human motor controller design can generate a variety of motions, such as adaptive motion with orthosis assistance in addition to walking, depending on how the human policy is learned [46], [58]. Therefore, future work could be a simulation of human adaptive behavior for orthosis assistance.

B. Stage 2 Training

Second, forward dynamics simulation results for Stage 2, that is, when a weakened soleus model wearing orthosis on the right ankle is controlled by π^{human} and $\pi^{orth_{\omega}}$ are shown in Fig. 6. In this simulation, π^{human} controls the musculoskeletal model in which the maximum force of the soleus, which generates propulsion during walking, is weakened by 75%, 87.5%, and 93.75%. At the same time, $\pi^{orth_{\omega}}$ assists the

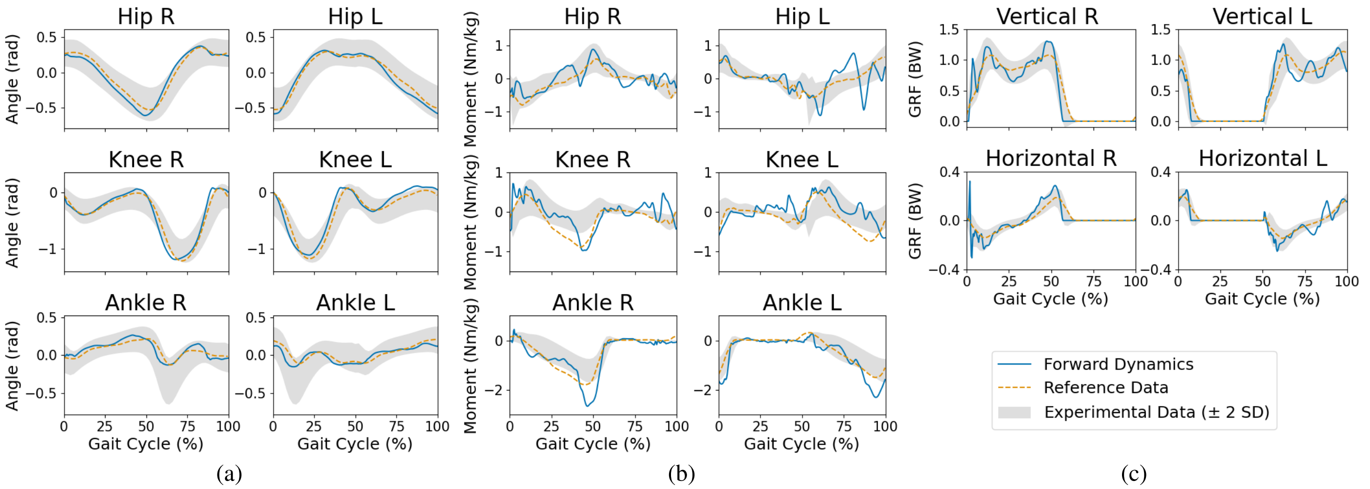

Fig. 5. **Results of forward dynamics simulation in Stage 1.** A healthy subject model is controlled by π^{human}. Joint angles (a), moments (b), and GRFs (c) obtained from forward dynamics simulation are shown (R and L in the subtitles are right side and left side, respectively). Simulation results (blue solid line), reference data used for imitation learning (orange dashed line), and experimental data [57] (shaded area) are compared together. Hip flexion, knee extension, and ankle dorsiflexion are positive. The hip angles of the experimental data are shifted by -0.35 radians to offset the difference between the experimental setup and the simulation model as in [54].

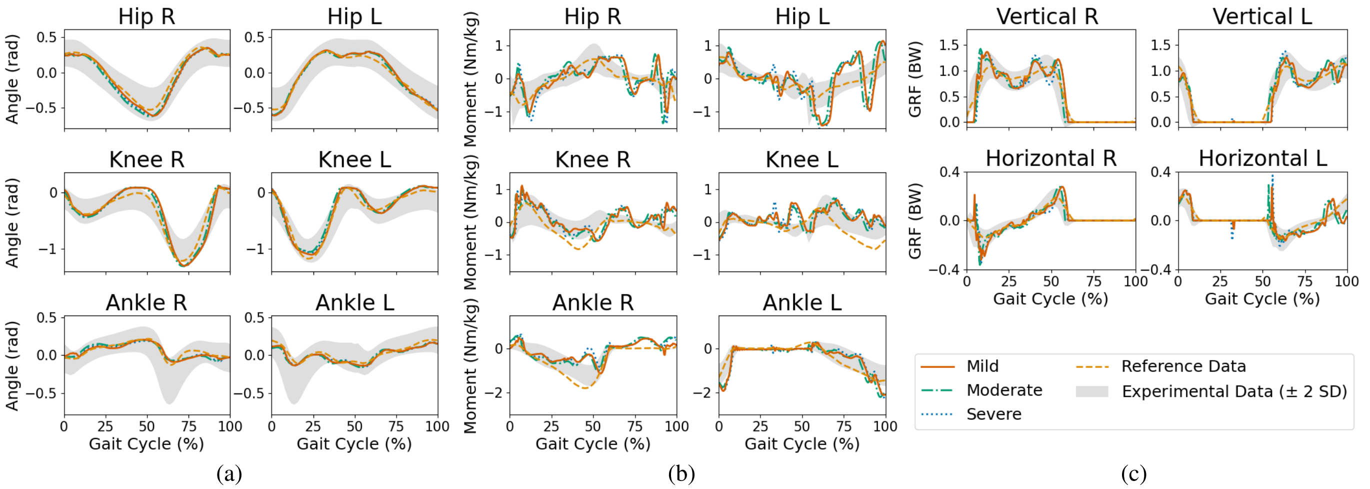

Fig. 6. **Results of forward dynamic simulation in Stage 2.** A soleus weakened model wearing an orthosis is controlled by π^{human} and π^{orth_ω}. Joint angles (a), moments (b), and GRFs (c) obtained from forward dynamics simulation are shown (R and L in the subtitles are right side and left side, respectively). Simulated data for the musculoskeletal model in which the maximum isometric force of Soleus muscle is weakened by 75%, 87.5%, and 93.75% (red solid line, green dash-dotted line, and blue dotted line, respectively), reference data used for imitation learning (orange dashed line), and experimental data [57] (shaded area) are compared together. Hip flexion, knee extension, and ankle dorsiflexion are positive. The hip angles of the experimental data are shifted by -0.35 radians to offset the difference between the experimental setup and the simulation model as in [54].

model in which the plantarflexion moment on the right side is insufficiently generated through an orthosis.

The joint angles of both legs for hip, knee, and ankle show good agreement with the reference data for all muscle weakness severity and lie within 2σ of the experimental data (Fig. 6a). It is observed that oscillations in the hip joint moment of Stage 2 are increased compared to Stage 1 (Fig. 6b). In particular, the right ankle moments significantly deviate from normal gait (Ankle R in Fig. 6b).

In the nonaffected leg, a moment close to zero is observed in 10–50% of the gait cycle, which corresponds to the swing phase (Ankle L in Fig. 6b). In contrast, a moment in the dorsiflexion direction occurred in the affected leg during the same phase (Ankle R in Fig. 6b). This may be due to the moment imbalance of the antagonistic pair, the dorsiflexor, and the plantar flexor, spanning the ankle. During the swing phase of a healthy subject's gait, co-contraction of the antagonistic pair occurs. The co-contraction of antagonistic pair is known to improve the stability of motion by increasing joint stiffness [59], [60]. We are also able to observe that this co-contraction occurs during the swing phase in simulation of healthy models (Fig. 7).

For Stage 2 simulation, the weakened maximal isometric force of the plantar flexor reduces the plantar flexors' contribution to the ankle joint moment during co-contraction of the antagonistic pair in the swing phase. This results in a net

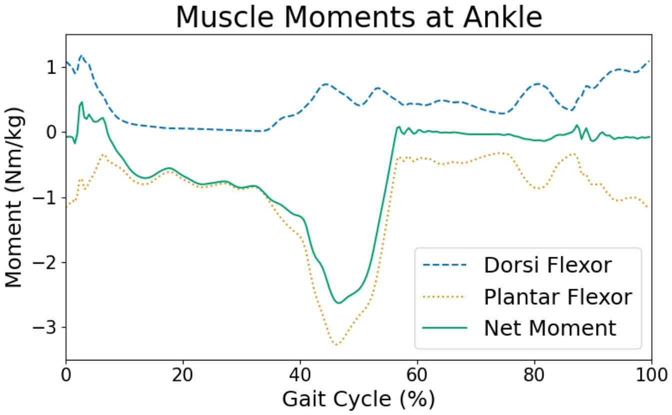

Fig. 7. **Moments generated by antagonistic muscle pairs spanning the ankle joint.** Moments caused by antagonistic muscle pairs in the right ankle of the model, i.e., tibialis (dorsi flexor), soleus, and gastrocnemius (plantar flexors) are shown. The net moment during the swing phase due to the muscles located in the ankle joint is almost zero. The moments in both plantar and dorsiflexion directions are similar in magnitude and opposite in direction at 60–100% of the gait cycle.

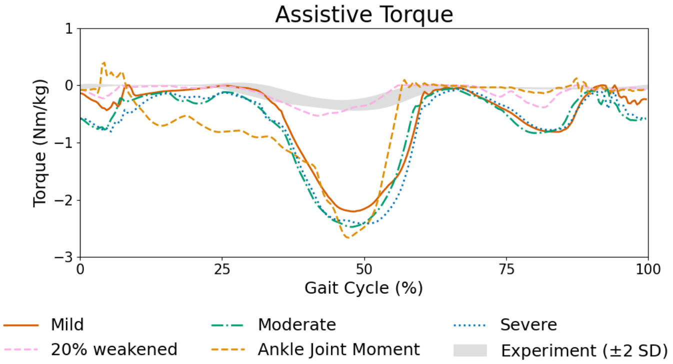

Fig. 8. **Assistive torques of the orthosis.** The assistive torque inputs applied to the human body by π^{orth_ω} in Stage 2 simulation are shown. For comparison, the ankle joint moment occurring in a healthy subject is also shown. The shaded area is the torque curve obtained from the ankle orthosis experiment on healthy subjects using the orthosis policy. The pink dashed line is the simulation result for the 20% Soleus weakened model.

joint moment of the ankle joint in the dorsiflexion direction, which can explain the same directional moment occurring in the swing phase of the affected side. The dorsiflexion moment should have caused a change in ankle joint angle. However, in the swing phase, the ankle angle is almost constant, close to 0 (Ankle R in Fig. 6a). From this, it can be inferred that an external moment is acting on the ankle to make the joint angle constant, which is observed in the assistive torque of the orthosis (Fig. 8).

In the GRF of Stage 2 simulation, fluctuations are observed in the 25%–60% gait cycle of the right GRF (Vertical R in Fig. 6c), which is the section where the assistive torque of orthosis is transmitted to the human body. Other than that, the overall trend is consistent with the reference data and experimental data (Fig. 5c). A high peak value is observed in the 0%–10% gait cycle of the right horizontal GRF as in Stage 1 simulation.

The orthosis controller required for normal walking of the soleus weakened musculoskeletal model is learned in π^{orth_ω} as shown in Fig. 8. The peak torque in the plantarflexion direction to assist the push-off movement in the stance phase was the smallest in the mild deficit model, and the moderate and severe cases were similar. For comparison, the ankle joint moment during walking of a healthy subject obtained through Stage 1 simulation is also shown (orange dashed line in Fig. 8). Assistive torques in plantar flexion direction are observed in approximately 75%–95% of the gait cycle, which corresponds to the swing phase. These torques contribute to making the ankle angle constant by compensating the ankle moment generated by muscles.

The result of a gait experiment performed by applying the orthosis policy to actual ankle orthosis was shown in Fig. 8 (shaded area in Fig. 8). For comparison, simulation result for the 20% soleus weakened model was added using the same orthosis policy as the experiment (pink dashed line in Fig. 8). An orthosis policy that outputs a value between 0 and 1 was used in simulations and experiments. However, the difference in the maximum force of the actuator used in the simulation and experiment causes a difference in torque. In the experiment, the timing of the orthosis policy's assistive torque was similar to the results of other simulations.

Classical optimization algorithms could have solved the problem. However, such an optimization method is effective only in specific settings, and if a small initial posture change is applied, we must solve the problem again. On the other hand, if deep RL is used, we can obtain a policy and a robust solution for conditions changes such as slight initial attitude. In the process of the policy synthesis, the policy can be fine-tuned via manipulation of the reward function in a systematic way. In addition, it can be applied to control the actual system through additional tuning to the learned policy.

We used a time step of 0.005s when learning policies. When performing forward simulation, using the time step in policy learning is unnecessary. We confirmed forward simulation is possible when using time step in the range of about 0.001–0.015s.

C. Analysis of pHRI

The pHRI at the contact area between the orthosis and the human body tended to increase as the deficit of the soleus became more severe (Fig. 9). This was because, the more severe the muscle weakness, the more assistive torque with greater magnitude is transmitted to the human model. In the swing phase, an interaction force of approximately 0.2 body weight (BW) in the −X direction and approximately 0.1 BW in the +Y direction was generated. The forces appear to be due to the moment in the plantar flexion direction from the orthosis to compensate for the moment deficiency. The peak value of the interaction force goes up to 0.5 BW in the −X direction and up to 0.25 BW in the +Y direction for severe deficit model. Such a large pHRI would have occurred because most of the torque for propulsion was assisted by the orthosis, which would not be suitable for actual assistance.

We show the results of pHRI in the X and Y directions obtained using the F/T sensor that measures the force between the human body and the robot in the gait experiment

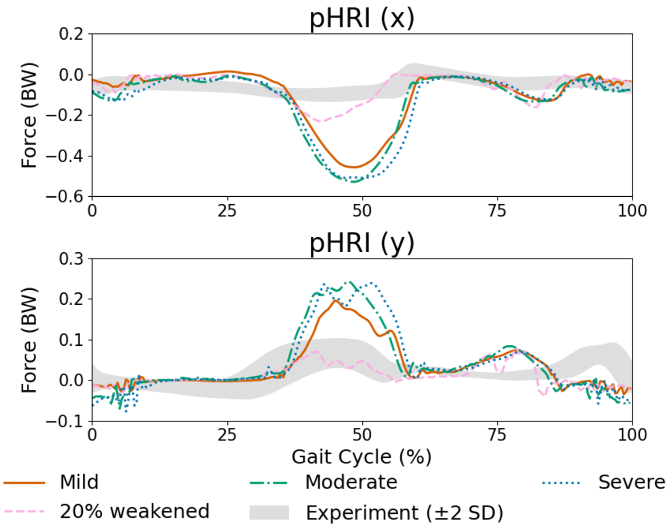

Fig. 9. **pHRI results according to the severity of muscle weakness in the model.** The x and y directions of pHRI are vertical and horizontal to the contact surface, respectively. The x and y coordinate system are the local coordinates attached to the orthosis, as shown in Fig. 2. The shaded area is the pHRI result obtained from the ankle orthosis experiment on healthy subjects using the learned orthosis policy. PHRI tends to be proportional to the amount of assistive torque. (Fig. 8).

(shaded area in Fig. 9). For comparison, the pHRI simulation results of the 20% Soleus weakened model are shown together (pink dashed lines in Fig. 9). In the experiment, the X and Y directional pHRI increased at about 25–60% gait cycle, the timing at which the assistance was applied. In the 20% Soleus weakened model, the magnitude of the X-directional pHRI was twice as large in that section. However, the magnitude of the Y-directional pHRI was located within one sigma of the experimental data. In our gait experiment, the pHRI due to the assistive torque applied in the swing phase (60–100% gait cycle) was almost nonexistent. The cause of pHRI in the Y-direction in 80–100% of the gait cycle was not due to the assistance, but it seems that the dorsiflexion of the late swing phase was measured in the load cell of the orthosis due to the structure of the orthosis.

When pHRI occurs, pressure, which is the force divided by the contact area, acts on the wearer's skin. Accordingly, the pressure applied to the wearer may be reduced by increasing the contact area between the robot and the wearer. In order to minimize the pHRI in the unwanted direction, one can design the orthosis to minimize the misalignment of the rotation axis of the human joint. To predict the wearability of a designed robot using simulated pHRI, it may be possible to calculate the pressure considering the human-robot contact area. In addition, attention should be paid to adjust the contact area and contact parameters to resemble the actual system.

When a human wears an orthosis, a human and a robot form a closed kinematic chain. In this case, if a rigid contact is used for the joint, there may be cases where the joint cannot be moved if there is a mismatch in link length and rotation axis between the human and the robot. Therefore, it is necessary to model a joint or contact part so that it can move relatively. The EFM we used is a contact model that can measure the force at the site where such relative motion occurs. In a future study, a study to optimize the contact parameters of EFM to be realistic will be investigated.

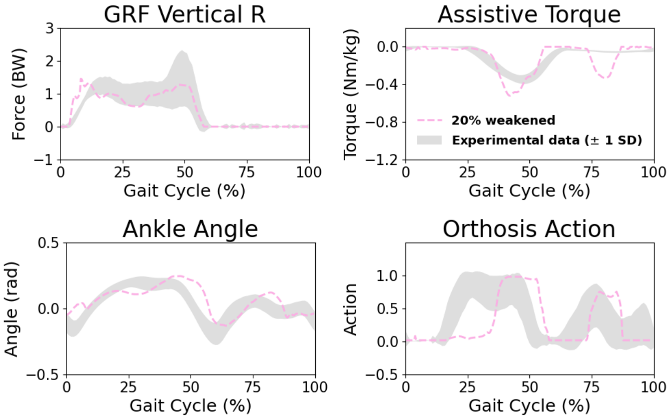

Fig. 10. **Results of gait experiment** The right vertical GRF (GRF Vertical R) and ankle angle (Ankle Angle), observations of the orthosis policy in the gait experiment, are shown. Also, the orthosis policy's action (Orthosis Action) and the applied torque (Assistive Torque) are shown. For comparison, simulation results of a 20% soleus weakened model using the same orthosis policy as the experiment are shown together. Experimental results are shaded areas, and simulations are pink dashed lines.

D. Analysis of Gait Experiment

We implemented the orthosis policy learned in Stage 2 on the fabricated ankle orthosis and performed a gait experiment on healthy subjects. For the experiment, it is necessary to scale the sensor data to the extent of the observation of the simulation. The pneumatic actuator we used in this experiment has an operating delay compared to the simulation due to the contraction time. The action had to be output earlier for the assistive torque to be input at the push-off timing. The orthosis policy outputs high action when the ankle angle is 0.13 rad or more under the condition of vertical GRF of about 0.6 BW or more. Also, in the swing phase with little GRF, high action was generated when the ankle angle was 0 rad or more. Therefore, we set the scale factors of angle and GRF to output the action earlier. In the gait experiment, we confirmed that the right vertical GRF (GRF Vertical R in Fig. 10) and the right ankle angle (Ankle Angle in Fig. 10) were similar to the simulation results. High output of orthosis action occurred earlier than simulation from the scale factor we set considering the delay of the pneumatic actuator (Orthosis Action in Fig. 10). We can see that assistive torque was input at the push-off (Assistive Torque in Fig. 10). In addition, the integrated EMG (iEMG) of the gastrocnemius, a plantar flexor, was reduced by 9.12% compared to the case without assistance. This result shows that the learned orthosis policy can be applied and utilized in the actual orthosis.

E. Limitations

The models and simulation methods we used made it possible to design human and robot controllers, including pHRI. However, there exists some limitations. First, we used a simplified musculoskeletal model. The planar model, which can only move on the sagittal plane, limits the joint degrees of freedom in different directions. This would have caused the kinematic/kinetic characteristics somewhat different from the actual human movement to be simulated. The use of a planar model also makes the use of muscles for balancing movements in the medial-lateral direction untrainable in the π^{human}. Second, the parameters used in the contact model are different from the real world. In the simulation, foot-ground contact and pHRI contact were included. The contact forces are calculated using predefined parameters such as stiffness, damping, and friction coefficient. We empirically determined the parameters, which may have caused the contact motion to differ from reality. For more accurate contact simulation, research on optimizing contact parameters based on the physical properties of the real environment is required. Third, while training the orthosis policy in the human-robot simulation of Stage 2, we kept the human policy from Stage 1 fixed. This assumption was due to the fact that human adpatation is much slower than deep RL's learning speed. Humans will tend to re-update their policy to better leverage the orthosis. For example, a person's self-selected walking speed is affected by robot assistance because of motor learning [61]. However, we kept the human policy fixed for a learning period of deep RL to minimize the complexity of the problem in this paper. A multi-agent RL study can be investigated to include the effect of human policy adaption into our problem that might result in competing or cooperating policies. Fourth, we performed a gait experiment on healthy subjects, not patients. Testing our method to patient and control groups requires careful composition of a new experimental protocol by clinicians and very expensive recruitment of many patients to meet the power analysis, which is beyond the scope of this study. Therefore, we focused on showing the feasibility of the trained policy and conducted a gait experiment on a healthy subject. In a future study, an experiment on the patient may be conducted after sufficient verification of the safety of the orthosis and controller. Fifth, in the gait experiment, we did not use all eight observations of the orthosis policy, but only two observations that dominated the action. Ideally, the measurable sensor data in orthosis should match the observation of the orthosis policy. In future studies, such an orthosis with all sensors will be used for the experiment.

V. Conclusion

In this paper, we showed how to design a policy π^{human} capable of generating human gait motions from sensory feedback and a policy π^{orth_ω} to assist a muscle weakened model through our proposed two-stage policy training using deep RL. As a result, the trained human policy was able to generate the gait of the human model successfully, and the orthosis policy was able to generate appropriate assistance so that the model with weakened soleus could generate a healthy gait. In addition, the pHRI predicted using EFM was verified through a gait assistance experiment using the learned orthosis policy. To the author's knowledge, this is the first study to propose an ankle orthosis policy synthesis that performs human-robot simulation through two-stage policy training without manual effort, including system modeling to control the human body with an orthosis. Our results can be used, for example, to design orthosis to reduce pHRI and predict performance indicators through simulation. In this study, we used the soleus weakened model, which can be similarly applied to other muscles, for example, the dorsiflexor muscle of the ankle, such as the tibialis anterior.

References

[1] S. A. Murray, K. H. Ha, C. Hartigan, and M. Goldfarb, "An assistive control approach for a lower-limb exoskeleton to facilitate recovery of walking following stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 3, pp. 441–449, Mar. 2015.

[2] E. A. Rogers, M. E. Carney, S. H. Yeon, T. R. Clites, D. Solav, and H. M. Herr, "An ankle-foot prosthesis for rock climbing augmentation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 41–51, 2020.

[3] E. P. Grabke, K. Masani, and J. Andrysek, "Lower limb assistive device design optimization using musculoskeletal modeling: A review," *J. Med. Devices*, vol. 13, no. 4, Dec. 2019, Art. no. 040801.

[4] A. Esquenazi, M. Talaty, A. Packel, and M. Saulino, "The ReWalk powered exoskeleton to restore ambulatory function to individuals with thoracic-level motor-complete spinal cord injury," *Amer. J. Phys. Med. Rehabil.*, vol. 91, no. 11, pp. 911–921, 2012.

[5] H. Kawamoto and Y. Sankai, "Power assist method based on phase sequence and muscle force condition for HAL," *Adv. Robot.*, vol. 19, no. 7, pp. 717–734, 2005.

[6] J. F. Veneman, R. Kruidhof, E. E. G. Hekman, R. Ekkelenkamp, E. H. F. V. Asseldonk, and H. V. D. Kooij, "Design and evaluation of the LOPES exoskeleton robot for interactive gait rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 3, pp. 379–386, Sep. 2007.

[7] J.-H. Kim *et al.*, "Design of a knee exoskeleton using foot pressure and knee torque sensors," *Int. J. Adv. Robotic Syst.*, vol. 12, no. 8, p. 112, Aug. 2015.

[8] J.-H. Kim, J. W. Han, D. Y. Kim, and Y. S. Baek, "Design of a walking assistance lower limb exoskeleton for paraplegic patients and hardware validation using CoP," *Int. J. Adv. Robot. Syst.*, vol. 10, no. 2, p. 113, Feb. 2013.

[9] J.-Y. Kuan, K. A. Pasch, and H. M. Herr, "A high-performance cable-drive module for the development of wearable devices," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 3, pp. 1238–1248, Jun. 2018.

[10] M. S. C.-H. Chien, A. Erdemir, A. J. van den Bogert, and W. A. Smith, "Development of dynamic models of the mauch prosthetic knee for prospective gait simulation," *J. Biomech.*, vol. 47, no. 12, pp. 3178–3184, Sep. 2014.

[11] S. Song *et al.*, "Deep reinforcement learning for modeling human locomotion control in neuromechanical simulation," *J. NeuroEng. Rehabil.*, vol. 18, no. 1, pp. 1–17, Dec. 2021.

[12] S. Harkema *et al.*, "Effect of epidural stimulation of the lumbosacral spinal cord on voluntary movement, standing, and assisted stepping after motor complete paraplegia: A case study," *Lancet*, vol. 377, pp. 1938–1947, Jun. 2011.

[13] S. Song and H. Geyer, "A neural circuitry that emphasizes spinal feedback generates diverse behaviours of human locomotion," *J. Physiol.*, vol. 593, no. 16, pp. 3493–3511, Aug. 2015.

[14] S. Song and H. Geyer, "Evaluation of a neuromechanical walking control model using disturbance experiments," *Frontiers Comput. Neurosci.*, vol. 11, p. 15, Mar. 2017.

[15] A. Ramadan, J. Cholewicki, C. J. Radcliffe, J. M. Popovich, Jr., N. P. Reeves, and J. Choi, "Reliability of assessing postural control during seated balancing using a physical human-robot interaction," *J. Biomech.*, vol. 64, pp. 198–205, Nov. 2017.

[16] Z. Yang, Y. Zhu, X. Yang, and Y. Zhang, "Impedance control of exoskeleton suit based on adaptive RBF neural network," in *Proc. Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, Aug. 2009, pp. 182–187.

[17] S. Jezernik and M. Morari, "Controlling the human-robot interaction for robotic rehabilitation of locomotion," in *Proc. 7th Int. Workshop Adv. Motion Control.*, Jul. 2002, pp. 133–135.

[18] G. A. Pratt and M. M. Williamson, "Series elastic actuators," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 1, Aug. 1995, pp. 399–406.

[19] H. Kazerooni, J.-L. Racine, L. Huang, and R. Steger, "On the control of the Berkeley lower extremity exoskeleton (BLEEX)," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2005, pp. 4353–4360.

[20] L. Zhang, Y. Liu, R. Wang, C. Smith, and E. M. Gutierrez-Farewik, "Modeling and simulation of a human knee Exoskeleton's assistive strategies and interaction," *Frontiers Neurorobot.*, vol. 15, p. 13, Mar. 2021.

[21] L. De Vree and R. Carloni, "Deep reinforcement learning for physics-based musculoskeletal simulations of healthy subjects and transfemoral Prostheses' users during normal walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 607–618, 2021.

[22] T. K. Uchida, A. Seth, S. Pouya, C. L. Dembia, J. L. Hicks, and S. L. Delp, "Simulating ideal assistive devices to reduce the metabolic cost of running," *PLoS ONE*, vol. 11, no. 9, Sep. 2016, Art. no. e0163417.

[23] C. L. Dembia, A. Silder, T. K. Uchida, J. L. Hicks, and S. L. Delp, "Simulating ideal assistive devices to reduce the metabolic cost of walking with heavy loads," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180320.

[24] D. G. Thelen, F. C. Anderson, and S. L. Delp, "Generating dynamic simulations of movement using computed muscle control," *J. Biomech.*, vol. 36, no. 3, pp. 321–328, 2003.

[25] A. Agrawal *et al.*, "First steps towards translating HZD control of bipedal robots to decentralized control of exoskeletons," *IEEE Access*, vol. 5, pp. 9919–9934, 2017.

[26] M. Shushtari, R. Nasiri, and A. Arami, "Online reference trajectory adaptation: A personalized control strategy for lower limb exoskeletons," *IEEE Robot. Autom. Lett.*, vol. 7, no. 1, pp. 128–134, Jan. 2022.

[27] Y. Xu, J. Choi, N. P. Reeves, and J. Cholewicki, "Optimal control of the spine system," *J. Biomech. Eng.*, vol. 132, no. 5, May 2010, Art. no. 051004.

[28] S. Arber, "Motor circuits in action: Specification, connectivity, and function," *Neuron*, vol. 74, no. 6, pp. 975–989, Jun. 2012.

[29] L. Kidzinski *et al.*, "Artificial intelligence for prosthetics: Challenge solutions," in *Proc. Competition, From Mach. Learn. Intell. Conversations (NeurIPS)*, 2019, p. 69.

[30] B. A. Richards *et al.*, "A deep learning framework for neuroscience," *Nature Neurosci.*, vol. 22, no. 11, pp. 1761–1770, Nov. 2019.

[31] G. Serrancoli *et al.*, "Subject-exoskeleton contact model calibration leads to accurate interaction force predictions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 8, pp. 1597–1605, Aug. 2019.

[32] M. A. Sherman, A. Seth, and S. L. Delp, "Simbody: Multibody dynamics for biomedical research," *Proc. Iutam*, vol. 2, pp. 241–261, Aug. 2011.

[33] M. W. Hast, B. G. Hanson, and J. R. Baxter, "Simulating contact using the elastic foundation algorithm in OpenSim," *J. Biomech.*, vol. 82, pp. 392–396, Jan. 2019.

[34] D. A. McCrea, "Spinal circuitry of sensorimotor control of locomotion," *J. Physiol.*, vol. 533, no. 1, pp. 41–50, 2001.

[35] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 3803–3810.

[36] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9–44, 1988.

[37] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1861–1870.

[38] S. Reddy, A. D. Dragan, and S. Levine, "SQIL: Imitation learning via reinforcement learning with sparse rewards," 2019, *arXiv:1905.11108*.

[39] S. Ross and D. Bagnell, "Efficient reductions for imitation learning," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 661–668.

[40] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 627–635.

[41] F. Torabi, G. Warnell, and P. Stone, "Behavioral cloning from observation," 2018, *arXiv:1805.01954*.

[42] S. Niekum, S. Osentoski, G. Konidaris, S. Chitta, B. Marthi, and A. G. Barto, "Learning grounded finite-state representations from unstructured demonstrations," *Int. J. Robot. Res.*, vol. 34, no. 2, pp. 131–157, 2015.

[43] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proc. 21st Int. Conf. Mach. Learn. (ICML)*, 2004, p. 1.

[44] N. D. Ratliff, D. Silver, and J. A. Bagnell, "Learning to search: Functional gradient techniques for imitation learning," *Auto. Robots*, vol. 27, no. 1, pp. 25–53, Jul. 2009.

[45] J. Lim, S. Ha, and J. Choi, "Prediction of reward functions for deep reinforcement learning via Gaussian process regression," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 4, pp. 1739–1746, Aug. 2020.

[46] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, "DeepMimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–14, 2018.

[47] S. L. Delp *et al.*, "OpenSim: Open-source software to create and analyze dynamic simulations of movement," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 11, pp. 1940–1950, Nov. 2007.

[48] G. Brockman *et al.*, "OpenAI gym," 2016, *arXiv:1606.01540*.

[49] D. G. Thelen, "Adjustment of muscle mechanics model parameters to simulate dynamic contractions in older adults," *J. Biomech. Eng.*, vol. 125, no. 1, pp. 70–77, Feb. 2003.

[50] M. Millard, T. Uchida, A. Seth, and S. L. Delp, "Flexing computational muscle: Modeling and simulation of musculotendon dynamics," *J. Biomech. Eng.*, vol. 135, no. 2, Feb. 2013, Art. no. 021013.

[51] A. Falisse, G. Serrancolí, C. L. Dembia, J. Gillis, and F. De Groote, "Algorithmic differentiation improves the computational efficiency of OpenSim-based trajectory optimization of human movement," *PLoS ONE*, vol. 14, no. 10, Oct. 2019, Art. no. e0217730.

[52] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: An open-source mesh processing tool," in *Proc. Eurographics Italian Chapter Conf.*, V. Scarano, R. D. Chiara, and U. Erra, Eds. Lisbon, Portugal: Eurographics Association, 2008, pp. 129–136.

[53] C. T. John, F. C. Anderson, J. S. Higginson, and S. L. Delp, "Stabilisation of walking by intrinsic muscle properties revealed in a three-dimensional muscle-driven simulation," *Comput. Methods Biomech. Biomed. Eng.*, vol. 16, no. 4, pp. 451–462, Apr. 2013.

[54] C. F. Ong, T. Geijtenbeek, J. L. Hicks, and S. L. Delp, "Predicting gait adaptations due to ankle plantarflexor muscle weakness and contracture using physics-based musculoskeletal simulations," *PLOS Comput. Biol.*, vol. 15, no. 10, Oct. 2019, Art. no. e1006993.

[55] X. B. Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," 2020, *arXiv:2004.00784*.

[56] H. S. Choi and Y. S. Baek, "Effects of the degree of freedom and assistance characteristics of powered ankle-foot orthoses on gait stability," *PLoS ONE*, vol. 15, no. 11, Nov. 2020, Art. no. e0242000.

[57] M. H. Schwartz, A. Rozumalski, and J. P. Trost, "The effect of walking speed on the gait of typically developing children," *J. Biomech.*, vol. 41, no. 8, pp. 1639–1650, 2008.

[58] X. B. Peng, G. Berseth, and M. van de Panne, "Terrain-adaptive locomotion skills using deep reinforcement learning," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–12, Jul. 2016.

[59] R. Wang and E. M. Gutierrez-Farewik, "Compensatory strategies during walking in response to excessive muscle co-contraction at the ankle joint," *Gait Posture*, vol. 39, no. 3, pp. 926–932, Mar. 2014.

[60] K. Falconer and D. Winter, "Quantitative assessment of co-contraction at the ankle joint in walking," *Electromyogr. Clin. Neurophysiol.*, vol. 25, nos. 2–3, pp. 135–149, 1985.

[61] S. Song and S. H. Collins, "Optimizing exoskeleton assistance for faster self-selected walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 786–795, 2021.