2020 8th IEEE International Conference on Biomedical
Robotics and Biomechatronics (BioRob)
New York, USA. Nov 29 - Dec 1, 2020

Reinforcement Learning Assist-as-needed Control for Robot Assisted Gait Training

Yufeng Zhang, *Student Member, IEEE*, Shuai Li, *Student Member, IEEE*, Karen J. Nolan and Damiano Zanotto*, *Member, IEEE*

***Abstract*— The primary goal of an assist-as-needed (AAN) controller is to maximize subjects' active participation during motor training tasks while allowing moderate tracking errors to encourage human learning of a target movement. Impedance control is typically employed by AAN controllers to create a compliant force-field around the desired motion trajectory. To accommodate different individuals with varying motor abilities, most of the existing AAN controllers require extensive manual tuning of the control parameters, resulting in a tedious and time-consuming process. In this paper, we propose a reinforcement learning AAN controller that can autonomously reshape the force-field in real-time based on subjects' training performances. The use of action-dependent heuristic dynamic programming enables a model-free implementation of the proposed controller. To experimentally validate the controller, a group of healthy individuals participated in a gait training session wherein they were asked to learn a modified gait pattern with the help of a powered ankle-foot orthosis. Results indicated the potential of the proposed control strategy for robot-assisted gait training.**

***Index Terms*— Assist-as-needed controller, robot-assisted gait training, reinforcement learning, wearable robotics, rehabilitation robotics**

I. Introduction

The introduction of compliant controllers in robotic trainers was motivated by clinical studies on human motor learning evidencing the importance of individuals' active participation in achieving desirable therapeutic outcomes [1]. Assist-as-needed (AAN) controllers aim to provide assistance only when the subject is unable to complete the training task on his/her own. Conventional AAN controllers modulate the assistive force based on an impedance control (IC) law [2]–[4]. This force-based control strategy is typically more compliant compared to early trajectory-based controllers [5] and has demonstrated its efficacy in improving motor training outcomes in several studies [6]–[9].

The use of IC in rehabilitation robotics was first introduced with the MIT-MANUS [10]. Later adopters such as the Anklebot extended the IC algorithm by adding programmable features that enabled adjustable control gains [11]. To encourage active subject participation when kinematic errors are small, the error-based AAN controller developed in [8], [9] features a width-adjustable virtual tunnel surrounding the desired trajectory, wherein no robotic assistance is provided. However, the "hard-coded" control parameters implemented in the aforementioned AAN controllers do not provide subjects with personalized assistance levels that are tailored to their motor abilities. As patients' adaptability and mobility may vary significantly from one to another [12], learning-based AAN controllers that adapt the control law based on individuals' training performances are desirable.

The adaptive AAN proposed in [7] uses radial basis functions (RBF) to learn an impairment model of the subject's motor capability within a given workspace, and modulates the robotic assistance level accordingly through a force decay term. Pehlivan *et al.* advanced this RBF-based method by introducing a subject-adaptive feedback gain modification algorithm that allows manipulation of the force support and the admissible error bound [13]. Iterative learning control (ILC) excels in reducing recurring errors in repetitive tasks, and hence has been extensively used to develop adaptive AAN controllers for gait training [3], [14], [15] and arm reaching tasks [16].

Reinforcement learning (RL) is a reward-based machine learning paradigm that has been utilized in various robotics applications such as assistive exoskeletons [17] and active prostheses [18]. Unlike conventional supervised and unsupervised machine learning algorithms, the goal-oriented behavior of a RL agent [19] makes it a convenient candidate for rehabilitation applications where patients' good performances and the associated control policies need to be remembered and rewarded. In fact, the underlying principle of RL echos the way human learns to interact with the environment [20], and the reward-based feedback can benefit the retention of learned motor skills [21]. Yet, only a paucity of research has investigated the application of the RL control paradigm in robot-assisted rehabilitation [22]–[24] and none in the context of gait training.

In this paper, we propose a real-time RL-AAN controller that is designed for robot-assisted gait training (RAGT). The actor-critic (AC) method is employed to realize action dependent heuristic dynamic programming (ADHDP), thereby eliminating the need for a system model [25]. Subjects' active participation is ascertained through a periodically updated control objective value that reduces the stiffness of an IC when kinematic errors are maintained sufficiently small for a certain period.

*Corresponding author.

Y. Zhang, S. Li and D. Zanotto (dzanotto@stevens.edu) are with the Wearable Robotics Systems (WRS) Laboratory, Department of Mechanical Engineering, Stevens Institute of Technology, Hoboken, NJ, 07030, USA.

K. J. Nolan is with Human Performance and Engineering Research, *Kessler Foundation*, West Orange, NJ 07052, USA and Rutgers-NJMS, Newark, NJ 07103, USA.

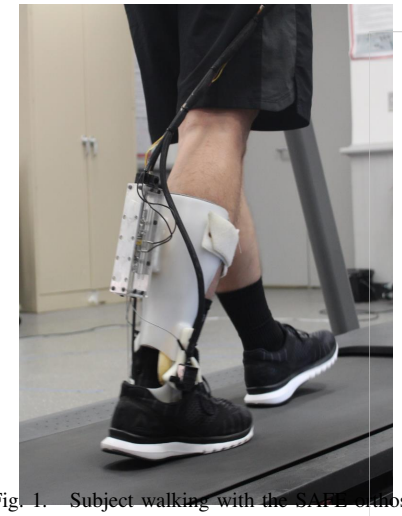
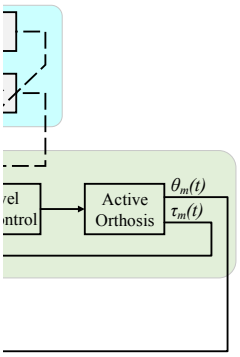
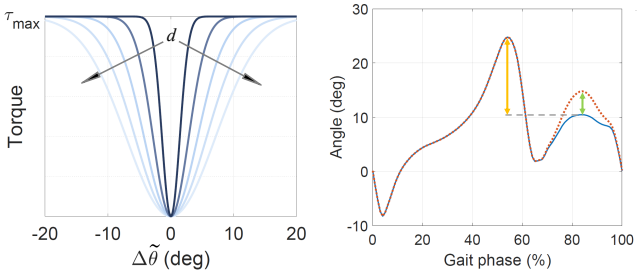


Fig. 1. Subject walking with the SAFE orthosis





(b) (c)

Fig. 2. **(a)** Block diagram of the RL-AAN controller (shaded area in green: IC loop; blue: AC structure; yellow: gait phase estimation) **(b)** for a given τ_{max}, the increase of d smoothens the stiffness profile, thereby resulting in a more compliant force-field **(c)** natural (blue) and target (red) gait pattern recorded from a pilot test. The green arrow indicates the increase in the peak ankle dorsiflexion angle during swing. This was computed as 30% of the difference between peak dorsiflexion angles in stance and swing (yellow arrow) in the natural gait pattern.

Building upon our previous work on RL-AAN control ankle mobilization exercises [24], this paper provides following contributions: 1) the extension of the RL-A controller to treadmill-based gait training exercises; 2 new tier-based update law for the control objective va to achieve enhanced adaptability of the control objectiv subjects' kinematic errors; 3) the integration of an on gait phase estimator, along with the a phase-depen target trajectory, to improve robustness of the RL-A controller to stride-to-stride variability; 4) a proof-of-conc validation of the proposed control strategy through RA tests with healthy individuals.

II. Experimental Setup

The lack of ankle dorsiflexion during swing can subst tially increase the risk of tripping and falling for str survivors [26]. Previous research on RAGT that sough address this problem has proposed Virtual Model Cont which lifts up a subject's foot via a virtual spring during swing phase to increase foot clearance [4]. The gait trair task defined in this work shares a similar goal, namely train a subject to walk with an enlarged foot clearance excercising with the Stevens Ankle-Foot Electromechanical (SAFE) orthosis, a powered orthosis developed by our group [27]. To this end, a subject-specific reference ankle trajectory with exaggerated dorsiflexion angle is generated based on each individual's natural gait pattern (Section III-B). During training, the RL-AAN controller is used to assist subjects to follow the target ankle motion during the swing phase of the gait cycle.

As shown in Fig. 1, the SAFE orthosis is a cable-driven device that can provide active torque control over ankle dorsiflexion and plantarflexion [27]–[29]. Two BLDC motors (EC45, Maxon) placed on an off-board actuation platform are used to actuate the cables. Two load-cells (LSB200, Futek) are connected in-line with the respective cables to monitor the applied forces and close the inner force control loop. A quadrature encoder is mounted on top of the rotating joint to measure the ankle joint angle. Two piezoresistive force sensors are embedded within the insole to detect heel-strike and toe-off events. Data acquisition and high/low level control are carried out by a myRIO board (National Instrument), with the low-level torque control loop running at 400 Hz and the high-level RL-AAN controller at 100 Hz.

III. RL-AAN Framework

As illustrated in Fig. 2(a), the proposed RL-AAN controller is in essence an IC, whose stiffness is modulated by an AC structure. We define the IC law as

$$\tau = \tau_{max}[1 - exp(-(\Delta\tilde{\theta}/d)^2)] \tag{1a}$$

$$\Delta\tilde{\theta} = \begin{cases} \text{sign}\,(\Delta\theta)\,(|\Delta\theta| - \theta_{\text{db}}) & \text{if } |\Delta\theta| \geq \theta_{db} \\ 0 & \text{if } |\Delta\theta| < \theta_{db} \end{cases} \tag{1b}$$

in which τ_{max} bounds the torque output τ and $\Delta\theta$ represents the difference between the reference ankle angle θ_{sr} and the measured angle θ_{sm}. To accommodate natural gait variation, $\Delta\theta$ within a deadband of $\theta_{db} = 2\ deg$ are disregarded when computing τ. Figure 2(b) illustrates the impact of a varying d on the stiffness profiles. The key element that sets the RL-AAN controller apart from conventional ICs [8], [9] is its ability to adapt the stiffness parameter d in real-time based on subject's ability. This is made possible through online evaluation of the subject's performance and periodical modulation of the control objective, both of which are formulated under the RL framework via the AC method [24].

A. Gait Phase Estimation

In order to realize selective subtask control over the swing phase of the gait cycle, the reference swing phase trajectory is generated as a function of the gait phase, starting from a subject's natural gait pattern measured during treadmill walking at comfortable pace, with the orthosis controlled in transparent mode [28]. A pool of adaptive frequency oscillators [30] is adopted as the online gait phase estimator. It takes the measured ankle trajectory θ_m as the input and compares it with the estimated angle $\bar{\theta}$, written as

$$\bar{\theta}(t) = \theta_0 + \sum_{i=1}^{M} \alpha_i sin(\phi_i(t)), \tag{2}$$

to update the offset θ_0, amplitudes α_i and phases ϕ_i. M is the number of oscillators, which was selected as 6 based on previous work [28]. The phase of the dominant harmonic ϕ_1 is regarded as the estimated gait phase. In addition, the phase-alignment strategy first proposed in [31] is implemented to ensure null phase at each heel-strike. The corrected phase ϕ_a is computed as

$$\phi_a(t) = mod\left[(\phi_1(t) - \phi_e(t), 2\pi\right], \tag{3}$$

in which the smooth phase correction term ϕ_e is updated based on the difference between the estimated null phase and the actual timing of the heel-strike detected by the insole-embedded force sensor.

B. Target Gait Pattern

The minimum toe clearance increases with larger ankle dorsiflexion angles [32], [33]. In this work, each participant's natural gait pattern is first recorded during treadmill walking with the orthosis controlled in transparent mode [28]. The average gait phase at which the ankle joint reaches maximum dorsiflexion during swing is defined as ϕ_μ. The target swing phase trajectory $\theta_{sr}(\phi_a)$ is a combination of the subject's natural gait pattern $\theta(\phi_a)$ and a Gaussian function centered at ϕ_μ

$$\theta_{sr}(\phi_a) = \theta(\phi_a) + Ke^{-(\phi_a - \phi_\mu)^2/(2\sigma^2)}, \tag{4}$$

where K defines the maximum angle to be added and σ controls the width of the bell-shaped curve. We set K to 30% of the difference between the peak dorsiflexion angles in stance phase and swing phase, as illustrated in Fig. 2(c). From preliminary tests, we found that this value was sufficiently large to induce measurable changes in gait (i.e., above stride-to-stride variability), while not jeopardizing gait stability. σ was set to 6 so that the added trajectory fades to effectively zero in the stance phase. It should be noted that this reference trajectory was selected as an example of functionally relevant target gait pattern for healthy individuals to learn during RAGT, thereby providing proof-of-concept validation of the RL-AAN controller (described in the following section). However, the applicability of the proposed RL-AAN controller is not limited to this specific trajectory.

C. Formalization of the RL-AAN controller

We consider a Markov decision process consisting of states, actions, reward, and a deterministic policy. A subject's performance is quantitatively evaluated after each gait cycle using spatial and temporal gait metrics that form the state vector S:

$$S(k) = [e_\phi(k), e_p(k), e_{rms}(k)]^T \in \mathbb{R}^{3\times 1} \tag{5}$$

e_ϕ and e_p indicate the phase and angle differences between the actual and the reference gait patterns when reaching the maximum ankle dorsiflexion in swing phase, e_{rms} indicates the root-mean-square (RMS) error in swing phase and k is the index representing the gait cycle. The immediate reward r (i.e., stage cost [34]) incurred at the k-th cycle is computed as a weighted sum of the three errors

$$r(k) = \lambda_1 e_\phi^2(k) + \lambda_2 e_p^2(k) + \lambda_3 e_{rms}^2(k), \tag{6}$$

in which $\lambda_{1,2,3}$ are arbitrary weights. The associated infinite-horizon cost with a discount factor γ is written as:

$$V(k) = \sum_{i=0}^{\infty} \gamma^i r(k+i+1). \tag{7}$$

We set λ_2 and λ_3 to 1.5 and λ_1 to 1 based on preliminary tests. γ was set to 0.2 to leverage recent rewards, thereby allowing the RL-AAN controller to promptly match the varying control objectives.

The AC structure resembles the one described in our previous work [24]. However, an updated actor objective function imposes tier-based constrains on the control objectives, with the goal of better responding to a subject's motor behavior. Nonlinear feedforward neural networks are used as the function approximator for the AC structure to simultaneously learn the action-value function and update its policy. The objective function of the critic neural network is written as

$$E_c(k) = \frac{1}{2}\delta^2(k), \tag{8}$$

in which $\delta(k)$ denotes the temporal-difference (TD) error at cycle k:

$$\delta(k) = r(k) + \gamma\widetilde{V}(k) - \widetilde{V}(k-1). \tag{9}$$

By updating the weight matricies of the neural network using gradient descent along the objective function, the critic network iteratively approximates the infinite-horizon cost $\widetilde{V}(k)$.

The action network seeks to learn an optimal policy and reach the principle optimality of a given problem. In gait rehabilitation, an ideal training protocol should involve repetitive, progressive, and task-oriented training sessions [35]. To integrate the rehabilitation objective within the control objective, the ultimate control goal U_c of the actor network is defined as a varying target that progressively guides subjects towards a desired performance by supplying assistive forces with varying stiffnesses. Since the discount factor γ was chosen to prioritize recent performances, the value of U_c can be thought of as the admissible error allowed during recent movement executions and a bound to the level of the assistive forces [13]. To this end, each new target

$U_c(k+1)$ is computed based on the subject's performance in the previous m strides. A three-tier saturation constraining U_c within [0, $\zeta_{1,2,3}$] is also introduced to ensure that the subject undergoes sufficient amount of practice at each level of force assistance. The update law of U_c is correlated with the subjects' performance via the moving average of the error-based rewards in the previous m strides, $\bar{\varepsilon}(k,m) = \sum_{i=k-m+1}^{k} r(k)/m$, according to the following law

$$U_c(k+1) = \begin{cases} U_c(k) + \beta^u & \text{if } \bar{\varepsilon}(k, m_n^u) \leq \eta_n^u \\ U_c(k) - \beta^d & \text{if } \bar{\varepsilon}(k, m^d) > \eta^d \end{cases} \quad (10)$$

$$\bar{U}_c(k+1) = \max\left[0, \min\left[U_c(k+1), \zeta_n\right]\right] \quad (11)$$

where β^u and β^d are positive and negative rates of change of U_c, respectively, and η_n^* sets the level of the acceptable kinematic errors within the n-th tier.

Based on pilot tests, we set $\beta^u = 1$, $\beta^d = 2$. The first tier evaluates the smallest amount of strides ($m_1^u = 5$) while accepting the largest kinematic errors ($\eta_1^u = 30$) and saturates the stiffness at a relatively high level ($\zeta_1 = 10$), suitable when the subject is struggling to follow the reference trajectory and necessitates large assistance from the robot. As the subject learns the target trajectory and the kinematic errors are reduced for a longer period (i.e., $m_2^u = 10$ strides), the second tier comes into effect and reduces both the acceptable errors ($\eta_2^u = 20$) and the assistance saturation level ($\zeta_2 = 20$). Further, the third tier has the longest evaluation window ($m_3^u = 30$), the smallest acceptable errors ($\eta_3^u = 10$) and the minimum assistance saturation level ($\zeta_3 = 30$). The third tier is expected to engage when the subject has learned the target motion and can therefore perform the task consistently well with little to no assistance. Conversely, when the subject exhibits markedly large kinematic errors ($\eta^d = 30$) in the most recent $m^d = 5$ strides, the control objective U_c is promptly reduced by β^d to yield greater assistance. Since each value of ζ_n represents a tolerance range that an optimized policy needs to ultimately stay within, the tier-based formulation of ζ_n essentially defines a set of tolerable error regions that bound the update of the policy [36].

In order to direct $\widetilde{V}(k)$ towards the desired $\bar{U}_c(k)$, the actor network objective function is formulated as

$$E_a(k) = \frac{1}{2}(\widetilde{V}(k) - \bar{U}_c(k))^2. \quad (12)$$

Since the actor output $\tilde{d}$ is intrinsically bounded within $(0, 1)$ by the sigmoid activation function of the neural network, it is scaled up as

$$d = \rho\tilde{d} \quad (13)$$

before being used in (1). We chose $\tau_{max} = 5\ Nm$ and $\rho = 20$ based on preliminary tests, so that desired robot torque 1(a) can be modulated from effectively zero to a sufficiently large value to guide the subject's ankle joint.

IV. Experiments

A total of $N = 8$ healthy subjects (all males, 26.13±2.10 years) participated in treadmill-walking tests to evaluate the RL-AAN controller. Participants were selected such that their right foot and shank could comfortably fit in the SAFE orthosis. The study was approved by Stevens Institutional Review Board and all participants provided informed consent prior to testing.

Each subject went through three walking conditions, which included one 3-min baseline trial (*B*), followed by three 5-min training trials (*T*) and a 3-min post evaluation trial (*P*). The treadmill was paced at 1 m/s across all trials. A 30-second ramp-up period was added before each trial to allow the treadmill speed, the subject's walking pattern, and the online gait phase estimator to reach steady state.

The SAFE orthosis was controlled in transparent mode in trials *B* and *P* to allow subjects to walk freely [28]. The RL-AAN controller was switched on during the *T* trials to provide subjects with the assistive forces needed for the intended training. The target swing phase trajectory used for the *T* trials was computed using (4), based on each subject's average ankle trajectory measured during the last minute of trial *B*.

To compare the immediate training effects on subjects' ankle motion, we analyzed the ankle trajectories recorded during the last minute of the *B* and *P* trials. Heel-strike events detected by the piezoresistive force sensor were used to split the continuous trajectories into gait cycles. Three metrics, namely e_ϕ, e_p and e_{rms}, were computed at each gait cycle, and average values of the three metrics were determined for each subject. Separate Wilcoxon rank-sum tests were performed on each metric, to check for significant ($\alpha = 0.05$) differences between the *B* and *P* trials. All data analysis was accomplished using custom scripts developed in MATLAB (MathWorks).

V. Results

The changes in S, $\bar{U}_c$, and d recorded from a representative subject over the course of a *T* trial are illustrated in Fig. 3. The overall trend in the subject's performance, as evaluated by spatiotemporal errors, was well-captured by $\bar{U}_c$. Small errors led to progressively larger $\bar{U}_c$ values, which in turn generated a more compliant IC (i.e., larger d) to encourage subject's active participation. Conversely, large errors drove $\bar{U}_c$ to smaller values, resulting in a stiffer IC (i.e., smaller d). The tier-based saturation ζ_n prevented $\bar{U}_c$ from ever-increasing, thereby avoiding abrupt IC stiffness changes and allowing subjects to experience a smooth stiffness modulation across movement repetitions.

Due to the small discount factor selected ($\gamma = 0.2$), the AC networks tend to learn the desired policy in a "myopic" manner, prioritizing the subjects' recent performances over the estimated long-term rewards, thus allowing the actor network to rapidly adapt to $\bar{U}_c$. As a consequence, the resulting trend of d, shown in Fig. 3(b), follows the dynamics of $\bar{U}_c$, yet with a small lag. This lag was likely caused by the additional time the AC networks needed to update the weights in order to compute an appropriate stiffness d that was capable of altering the assistance level to reflect the changing objective $\bar{U}_c$. The varying d is indicative of a continuously adapting control stiffness that tailors the

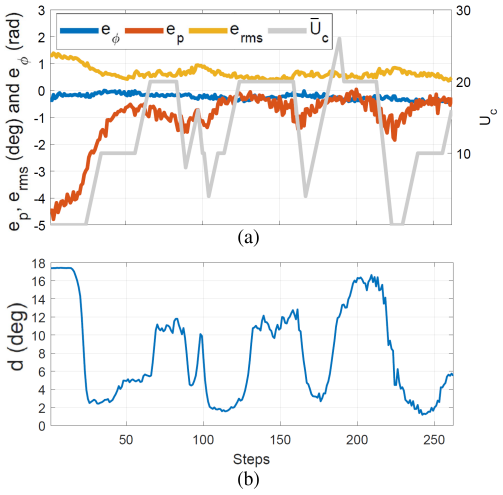


Fig. 3. **(a)** Error metrics, $\bar{U}_c$, and **(b)** d over the course of a T trial (a moving average with a 5-step window size was applied to e_ϕ, e_p and e_{rms})

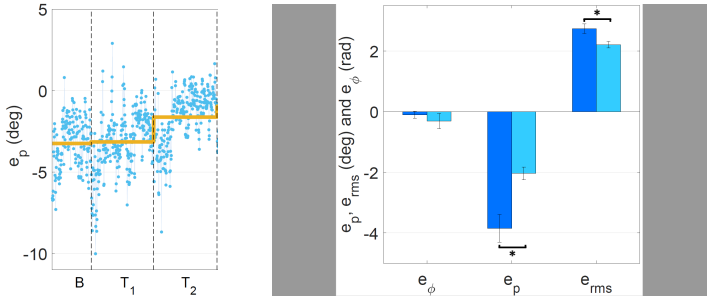


Fig. 4. **(a)** Stride-by-stride (blue dots) and average (yellow lines) e_p for a representative subject across different walking conditions, RL-AAN controller was turned on during $T_{1,2,3}$ **(b)** comparison of training effects with respect to the three error metrics analyzed. Asterisk indicates $p < 0.05$, error bars show ± 1 standard error, darker/lighter shade indicate baseline (B) and post-test (P) performance, respectively.

force-field to subjects' abilities. It is important to note that, since d does not directly determine the amount of assistance provided to the subjects but rather modulates the shape of the underlying force-field, the critical causal relationship between effort and error in the process of motor learning is preserved by the proposed RL-AAN controller.

Figure 4(a) illustrates the stride-by-stride and average values of e_p obtained from a representative subject across all walking conditions. It can be noticed that, despite the variance exhibited in the first training trial (T_1), this subject was able to progressively reduce e_p in T_2 and T_3, suggesting that the subject was able to successfully adapt to the RL-AAN controller across the training sessions. The trend of e_p during the post-training trial P (when the orthosis was controlled in transparent mode), suggests a rather good retention of the target movement. Figure 4(b) shows the group averages of the three performance metrics in the pre- and post-training, along with their standard errors. Significant differences were found in e_p and e_{rms}, indicating that the subjects were able to benefit from the proposed RL-AAN controller and modify their gait patterns towards the target motion immediately after training. No substantial changes were found on e_ϕ, which is an expected outcome given that the target gait pattern was designed to maintain the same peak dorsiflexion timing as in the natural gait. The slight increase in e_ϕ after training was accompanied by larger inter-subject variability, suggesting that some subjects adapted to the target gait pattern by slightly advancing the timing of the peak dorsiflexion angle while others reached that peak with some lag.

VI. Discussion and Conclusion

This paper introduced the first RL-AAN controller for robot-assisted gait training and its proof-of-concept validation. The proposed *model-free* controller is capable of modulating the stiffness of the assistive force field of a powered orthosis on the fly based on the wearer's performance in previous gait cycles. The RL-AAN controller builds upon our previous work on robot-assisted ankle mobilization [24], but embeds a three-tier update law for the control goal and an online phase estimator, which enable improved adaptability of the level of assistance and robustness to stride-to-stride adjustments in cadence.

The proposed RL-AAN controller is promising for the following reasons. First, the use of the ADHDP allowed the RL-AAN to be implemented without prior knowledge of a system model, thereby presenting a distinct advantage over model-based AAN controllers [37]. Second, the vast expressive power of nonlinear neural networks allows the algorithm to be easily scaled up to comprehend complex relationships between between robot control decisions and subjects' motor behaviors, thereby providing an appealing alternative solution to RBF-based algorithms [7], [13] that are prone to the "curse of dimensionality" [19]. ILC-based controllers have been extensively used in RAGT because of their ability to deal with cyclic tasks. Yet, in the context of AAN controllers for robotic trainers, their reliance on simplistic assumptions about the human motor learning dynamics makes them less robust to cope with different levels of subjects' abilities [3]. This issue is alleviated by the proposed RL-AAN controller, since the human performances are not only being continuously evaluated online, but also inherently embedded in the RL algorithm that directly influences the control decisions.

While promising, the results obtained from this study are still preliminary. For instance,the adaptation of the force-filed was limited to the stride level in the current setup. Future study will seek to extend the granularity of the adaptation law to enable phase-dependent modulation of the control stiffness, in order to meet the wearer's potentially varying needs among different gait phases. In addition, the longitudinal effects of the RL-AAN controller relative to more conventional AAN strategies will need to be appropriately analyzed with multi-session studies on healthy individuals and patients with neurological disorders affecting gait.

The motion task studied in this paper involved a well-defined spatial perturbation to the natural walking patterns of healthy individuals. Thus, even though results suggested

that the inclusion of the temporal error in the value function does not benefit training outcomes, this may not be the case for hemiparetic patients, whose gait often shows abnormal spatial and temporal features [38].

ACKNOWLEDGMENTS

This work was supported in part by the US National Science Foundation (Grant Number 1944203) and the New Jersey Health Foundation (Research Grants Program, Grant Number PC 6-18).

REFERENCES

[1] M. Lotze, G. Scheler, H.-R. Tan, C. Braun, and N. Birbaumer, "The musician's brain: functional imaging of amateurs and professionals during performance and imagery," *Neuroimage*, vol. 20, no. 3, pp. 1817–1829, 2003.

[2] R. Riener, L. Lunenburger, S. Jezernik, M. Anderschitz, G. Colombo, and V. Dietz, "Patient-cooperative strategies for robot-aided treadmill training: first experimental results," *Transactions on neural systems and rehabilitation engineering*, vol. 13, no. 3, pp. 380–394, 2005.

[3] J. L. Emken, R. Benitez, and D. J. Reinkensmeyer, "Human-robot cooperative movement training: learning a novel sensory motor transformation during walking with robotic assistance-as-needed," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, p. 8, 2007.

[4] B. Koopman, E. H. Van Asseldonk, and H. Van der Kooij, "Selective control of gait subtasks in robotic gait training: foot clearance support in stroke survivors with a powered exoskeleton," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 3, 2013.

[5] K. P. Westlake and C. Patten, "Pilot study of Lokomat versus manual-assisted treadmill training for locomotor recovery post-stroke," *Journal of neuroengineering and rehabilitation*, vol. 6, no. 1, p. 18, 2009.

[6] J. L. Emken and D. J. Reinkensmeyer, "Robot-enhanced motor learning: accelerating internal model formation during locomotion by transient dynamic amplification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 33–39, 2005.

[7] E. T. Wolbrecht, V. Chan, D. J. Reinkensmeyer, and J. E. Bobrow, "Optimizing compliant, model-based robotic assistance to promote neurorehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 3, pp. 286–297, 2008.

[8] D. Zanotto, P. Stegall, and S. K. Agrawal, "Adaptive assist-as-needed controller to improve gait symmetry in robot-assisted gait training," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 724–729.

[9] S. Srivastava, P.-C. Kao, S. H. Kim, P. Stegall, D. Zanotto, J. S. Higginson, S. K. Agrawal, and J. P. Scholz, "Assist-as-needed robot-aided gait training improves walking function in individuals following stroke," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 956–963, 2015.

[10] N. Hogan, H. I. Krebs, J. Charnnarong, P. Srikrishna, and A. Sharon, "MIT-MANUS: a workstation for manual therapy and training. I," in *1992 IEEE International Workshop on Robot and Human Communication*. IEEE, 1992, pp. 161–165.

[11] A. Roy, H. I. Krebs, D. J. Williams, C. T. Bever, L. W. Forrester, R. M. Macko, and N. Hogan, "Robot-aided neurorehabilitation: a novel robot for ankle rehabilitation," *IEEE Transactions on Robotics*, vol. 25, no. 3, pp. 569–582, 2009.

[12] J. L. Patton, M. E. Stoykov, M. Kovic, and F. A. Mussa-Ivaldi, "Evaluation of robotic training forces that either enhance or reduce error in chronic hemiparetic stroke survivors," *Experimental brain research*, vol. 168, no. 3, pp. 368–383, 2006.

[13] A. U. Pehlivan, F. Sergi, and M. K. O'Malley, "A subject-adaptive controller for wrist robotic rehabilitation," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 3, pp. 1338–1350, 2014.

[14] J. Bae and M. Tomizuka, "A gait rehabilitation strategy inspired by an iterative learning algorithm," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 2857–2864, 2011.

[15] S. Maggioni, N. Reinert, L. Lunenburger, and A. Melendez-Calderon, "An adaptive and hybrid end-point/joint impedance controller for lower limb exoskeletons," *Frontiers in Robotics and AI*, 2018.

[16] C. T. Freeman, E. Rogers, A.-M. Hughes, J. H. Burridge, and K. L. Meadmore, "Iterative learning control in health care: Electrical stimulation and robotic-assisted upper-limb stroke rehabilitation," *IEEE Control Systems Magazine*, vol. 32, no. 1, pp. 18–43, 2012.

[17] R. Huang, H. Cheng, Q. Chen, H.-T. Tran, and X. Lin, "Interactive learning for sensitivity factors of a human-powered augmentation lower exoskeleton," in *2015 IEEE International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 6409–6415.

[18] Y. Wen, J. Si, X. Gao, S. Huang, and H. H. Huang, "A new powered lower limb prosthesis control framework based on adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 2215–2220, 2017.

[19] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT Press, Cambridge, MA, 1998.

[20] W. Schultz, P. Dayan, and P. R. Montague, "A neural substrate of prediction and reward," *Science*, vol. 275, no. 5306, 1997.

[21] D. P. Losey, L. H. Blumenschein, and M. K. O'Malley, "Improving the retention of motor skills after reward-based reinforcement by incorporating haptic guidance and error augmentation," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2016, pp. 857–863.

[22] C. Obayashi, T. Tamei, and T. Shibata, "Assist-as-needed robotic trainer based on reinforcement learning and its application to dart-throwing," *Neural Networks*, vol. 53, pp. 52–60, 2014.

[23] X. Huang, F. Naghdy, H. Du, G. Naghdy, and C. Todd, "Reinforcement learning neural network (RLNN) based adaptive control of fine hand motion rehabilitation robot," in *2015 IEEE Conference on Control Applications (CCA)*. IEEE, 2015, pp. 941–946.

[24] Y. Zhang, S. Li, K. J. Nolan, and D. Zanotto, "Adaptive assist-as-needed control based on actor-critic reinforcement learning," in *IEEE International Conference on Intelligent Robots and Systems*, 2019.

[25] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Transactions on Neural Networks*, vol. 8, no. 5, pp. 997–1007, 1997.

[26] J. L. Burpee and M. D. Lewek, "Biomechanical gait characteristics of naturally occurring unsuccessful foot clearance during swing in individuals with chronic stroke," *Clinical Biomechanics*, vol. 30, no. 10, pp. 1102–1107, 2015.

[27] Y. Zhang, R. J. Kleinmann, K. J. Nolan, and D. Zanotto, "Preliminary validation of a cable-driven powered ankle-foot orthosis with dual actuation mode," *IEEE Transactions on Medical Robotics and Bionics*, vol. 1, no. 1, pp. 1–8, 2019.

[28] Y. Zhang, K. J. Nolan, and D. Zanotto, "Oscillator-based transparent control of an active/semiactive ankle-foot orthosis," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 247–253, Dec. 2018.

[29] Y. Zhang, N. Karen, and D. Zanotto, "Immediate effects of force feedback and plantar somatosensory stimuli on inter-limb coordination during perturbed walking," in *IEEE International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019.

[30] L. Righetti, J. Buchli, and A. J. Ijspeert, "Adaptive frequency oscillators and applications," *The Open Cybernetics and Systemics Journal*, vol. 3, no. ARTICLE, pp. 64–69, 2009.

[31] T. Yan, A. Parri, V. R. Garate, M. Cempini, R. Ronsse, and N. Vitiello, "An oscillator-based smooth real-time estimate of gait phase for wearable robotics," *Autonomous Robots*, vol. 41, no. 3, 2017.

[32] M. A. Moosabhoy and S. A. Gard, "Methodology for determining the sensitivity of swing leg toe clearance and leg length to swing leg joint angles during gait," *Gait & Posture*, vol. 24, no. 4, pp. 493–501, 2006.

[33] C. Brown, "Foot clearance in walking and running in individuals with ankle instability," *The American journal of sports medicine*, vol. 39, no. 8, pp. 1769–1777, 2011.

[34] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems*, vol. 32, no. 6, pp. 76–105, 2012.

[35] B. French, L. H. Thomas, J. Coupe, N. E. McMahon, L. Connell, J. Harrison, C. J. Sutton, S. Tishkovskaya, and C. L. Watkins, "Repetitive task training for improving functional ability after stroke," *Cochrane database of systematic reviews*, no. 11, 2016.

[36] R. Hafner and M. Riedmiller, "Reinforcement learning in feedback control," *Machine learning*, vol. 84, no. 1-2, pp. 137–169, 2011.

[37] S. Hussain, S. Q. Xie, and P. K. Jamwal, "Adaptive impedance control of a robotic orthosis for gait rehabilitation," *IEEE Transactions on Cybernetics*, vol. 43, no. 3, pp. 1025–1034, 2013.

[38] A.-L. Hsu, P.-F. Tang, and M.-H. Jan, "Analysis of impairments influencing gait velocity and asymmetry of hemiplegic patients after mild to moderate stroke," *Archives of physical medicine and rehabilitation*, vol. 84, no. 8, pp. 1185–1193, 2003.