2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Reinforcement Learning Control Outperforms Iterative Learning in Exoskeleton-Assisted Gait Training

Andy Li, Haoran Li, Aytac Teker, Mariana H. Rocha, Biruk A. Gebre,
Karen J. Nolan, Kishore Pochiraju and Damiano Zanotto*

***Abstract*— Learning-based controllers are increasingly adopted in lower-extremity powered exoskeletons, yet their advantages over traditional adaptive approaches remain underexplored. We compared two adaptive assist-as-needed (AAN) controllers for gait training with an ankle exoskeleton: a reinforcement learning-based controller (RL-AAN) and a conventional iterative learning controller (ILC-AAN). Both adjusted assistance stride-by-stride, delivering torque as a percentage of the wearer's biological plantarflexion moment—estimated online with a subject-agnostic model—and progressively faded assistance as performance improved. Healthy participants walked on a self-paced treadmill under a perturbed-gait protocol. Performance was assessed as average percent stride-velocity (SV) improvement relative to unassisted perturbed walking ($\Delta\%\varepsilon^{+}_{SV}$) and percent of strides above the SV threshold ($N\%^{+}_{SV}$). During training, RL-AAN and ILC-AAN elicited comparable gains in $\Delta\%\varepsilon^{+}_{SV}$, but RL-AAN yielded greater adherence, as indicated by larger $N\%^{+}_{SV}$. After training, RL-AAN demonstrated superior retention in $\Delta\%\varepsilon^{+}_{SV}$ and $N\%^{+}_{SV}$. These results support RL-AAN as a promising strategy for subject-tailored gait training, motivating future studies in clinical cohorts with neurological or musculoskeletal gait impairments.**

***Index Terms*— Personalized Ankle Exoskeleton, Assist-as-Needed Control, Robot-Assisted Gait Training, Reinforcement Learning.**

I. Introduction

High-Energy Lower Extremity Trauma (HELET) resulting from blast or blunt injuries, motor-vehicle collisions, or falls, is a major source of long-term mobility disability in both military and civilian populations, often involving complex fractures and soft-tissue damage [1]. Advances in reconstructive surgery have increased the feasibility of limb salvage, yet many individuals with salvaged limbs continue to experience pain, weakness, and sensorimotor deficits that equal or even exceed those reported by amputees [2]. Conventional rehabilitation is typically delivered by professional therapists through progressive gait and strengthening programs and intense, task-specific exercise. While these interventions can restore function, their accessibility is often constrained by cost and therapist availability. Device-based interventions use ankle–foot orthoses (AFOs) to augment propulsion and stability. Passive, energy-storing AFOs can improve walking performance after intensive therapy programs but, by resisting plantarflexion, may elicit compensatory strategies and suboptimal mechanics that impact recovery [3].

Powered AFOs and ankle exoskeletons can deliver torque in phase with biological demand, improving propulsion while allowing more physiological kinematics. Powered ankle assistance has proven effective in eliciting immediate walking speed gains in several populations. In children with cerebral palsy, an untethered powered AFO that delivered a fraction of the estimated biological ankle moment increased overground speed by 6.3% [4]. In adults with hemiparesis, a unilateral cable-driven ankle exosuit providing timed push-off and ground-clearance support yielded a 6.7% increase in overground speed [5]. In healthy adults, human-in-the-loop online optimization of a parameterized torque profile on a tethered ankle exoskeleton produced up to a 42% increase in self-selected treadmill speed relative to unassisted walking [6]. However, the use of powered AFOs for HELET individuals has been largely underexplored [7]. Additionally, evidence of post-intervention retention after exercise protocols mediated by powered AFOs remains limited [8].

Active engagement is a key enabler of motor recovery during exercise-based rehabilitation [9]. The assist-as-needed (AAN) control framework implements this principle by reducing assistance to the wearer when performance is adequate and intervening only to reduce excessive task errors, thereby encouraging user effort [10]. Learning-based control strategies have gained popularity in both assistive and rehabilitative applications in recent years [11]. While conventional AAN has traditionally relied on iterative learning control (ILC) [10], [12] or fixed-impedance rules [13], more recent AAN implementations have used a range of learning-based policies, including reinforcement-learning (RL) controllers that adapt assistance over strides/repetitions or phases [14]–[17], data-driven personalization frameworks [18], and robust RL approaches that accommodate uncertainty in patient–device dynamics [19]. Despite these advances, direct head-to-head comparisons of traditional and learning-based AAN control laws under identical sensing, actuation, and dosing conditions have not been conducted, making it difficult to attribute performance gains to the adaptation rule rather than to hardware or protocol differences [20].

As a first step towards addressing this gap, we present a within-subject, crossover comparison of ILC-based and RL-based AAN controllers (ILC-AAN and RL-AAN, respectively), implemented on the same ankle exoskeleton and operating under the same low-level cascaded torque-

*Corresponding author: (dzanotto@stevens.edu).

A. Li, H. Li, and D. Zanotto are with the WRS Lab., Stevens Inst. of Technology, Hoboken, NJ 07030, USA.

M. H. Rocha, A. Teker, B. A. Gebre, and K. Pochiraju are with the PROOF Lab., Stevens Inst. 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.

velocity controller [21]. Both controllers modulate assistance stride-by-stride, by adjusting the percentage of the wearer's biological plantarflexion moment—estimated online using a subject-agnostic model [22]—to ensure personalized and biomechanically relevant support. ILC-AAN adjusts assistance proportionally to recent stride velocity (SV) errors, integrating a discount factor that drives assistance to zero when sustained performance is achieved. RL-AAN employs an actor–critic structure in which the actor's control objective balances predicted long-term SV errors against assistance minimization [23]. We adopted a perturbed-gait protocol on a self-paced treadmill, whereby participants wore ankle weights to disrupt natural gait, and each subject's unperturbed, unassisted SV served as the training target for exoskeleton-assisted walking. We assessed immediate SV changes across four training sessions and short-term retention during three post-tests conducted up to 25 minutes after the final training session. The remainder of the paper is organized as follows: Section II presents an overview of the ankle exoskeleton and describes the ILC-AAN and RL-AAN controllers. Section III details the perturbed gait protocol with healthy individuals. Section IV summarizes the results from the training sessions. Section V discusses implications and concludes the paper.

II. Methods

A. Ankle Exoskeleton

We employed a unilateral, cable-driven, untethered ankle exoskeleton with a subject-tailored ankle unit (*Strider*, Fig. 1), which represents an improved version of a previous prototype [21]. The ankle unit is fabricated in carbon-fiber–reinforced PLA (PLA-CF) via fused-deposition modeling (FDM), using an automated workflow that leverages a 3D scan of the wearer's shank and foot to ensure a close fit and reduce relative body-device motions [21], [24]. The ankle unit (0.8 kg) includes quick-release straps and integrated mounts for an ankle-angle sensor and for the elastic module of a rotary series elastic actuator (0.2 kg, 43 Nm/rad) [25].

The backpack-like actuation unit can be configured to drive either left or right ankle units. Compared with the previous design, it is lighter (4.7 kg vs. 5.4 kg) and includes an updated semi-rigid vest featuring both front and back articulated plates retrofitted with 2-in webbing straps to conform more closely to the user's body, improving fit and load distribution. The actuation unit houses a brushless DC motor (EC-90 flat, 600 W, Maxon Group, Switzerland), a motor drive (FE060-25-EM, ADVANCED Motion Controls, USA), a real-time target machine (Speedgoat GmbH, Switzerland), and an updated stacked PCB–now fully enclosed in the actuation unit's housing–responsible for data acquisition and conditioning (DAQ) as well as power distribution. The motor output shaft drives a threaded spool (effective diameter: 12.4 mm) to wind/unwind a Bowden cable loop without overlap. Mechanical power is transmitted through the cable loop and routed via a tensioner to the bidirectional rotary elastic module mounted on the ankle unit. The module stiffness has an effective outer diameter of 76.6 mm, yielding an

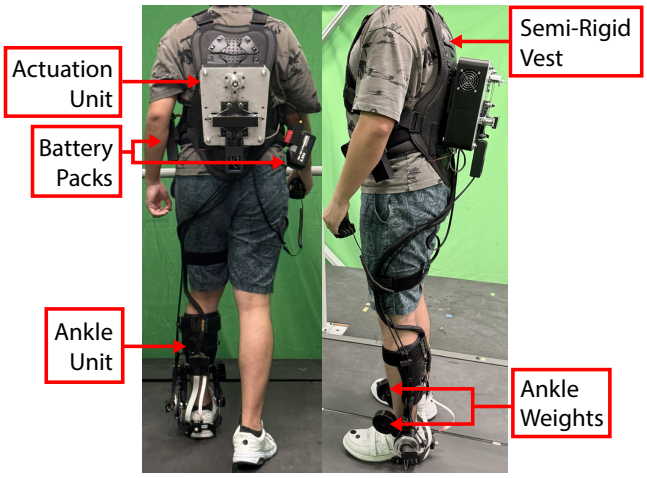

Fig. 1. Strider exoskeleton featuring updated actuation unit with integrated data-acquisition board and lighter battery pack. Two ankle weights (1 kg each) were attached to the distal end of the exoskeleton's shank bracket to generate a perturbation.

overall transmission ratio of 6:1, which results in a maximum assistive torque of 18 Nm. Two battery packs (Li-ion, 36 V, 4 Ah for the power circuit; Li-Po, 18.5 V, 3 Ah for the logic circuit) were secured to a sturdy tool belt using custom mounts, adding a total of 0.5 kg to power the device—representing a 31% weight reduction compared with the previous design. On-board sensing includes a magnetic encoder (AS5048A, ams-OSRAM, Austria) to measure rotary elastic element deflection; a precision potentiometer (NP24HS, P3 America, USA) to measure ankle plantar- and dorsi-flexion angle; and an insole-embedded 8-cell force-sensitive resistor (FSR) array (IEE S.A., Luxembourg) together with a 9-DOF inertial measurement unit (3-Space IMU, Yost Labs, USA). The DAQ board, the motor drive, and the real-time target machine are interfaced over an EtherCAT fieldbus at 1 kHz.

B. Adaptive Controllers for Walking Speed Training

To provide personalized and biomechanically relevant assistance, we employed a subject-agnostic online ankle moment estimator based on ensemble Gaussian Process Regression (GPR) [22], which integrates signals from the insole-embedded FSR array, the IMU, and the ankle potentiometer, together with an estimated stance phase derived from a pool of adaptive frequency oscillators (Fig. 2). Both ILC-AAN and RL-AAN are designed to update the assistance gain $K(i) \in [0, 10]$, representing a percentage of the wearer's estimated biological plantarflexion moment, after each initial contact (heel strike). The commanded ankle torque $\tau_{\text{ref}}(t)$ is mapped directly from the estimator output as

$$\tau_{\text{ref}}(t) = \hat{M}_{\text{PF}}(t)\, K(i)/100, \tag{1}$$

where $\hat{M}_{\text{PF}}(t)$ is the estimated instantaneous plantarflexion moment and $K(i)$ is the assistance gain applied during the i-th gait cycle. The 0–10% range provides an interpretable, subject-normalized scale and keeps assistance within hardware-feasible, comfortable levels. $K(i) = 0$ yields transparent behavior (equivalent to a zero-torque con-

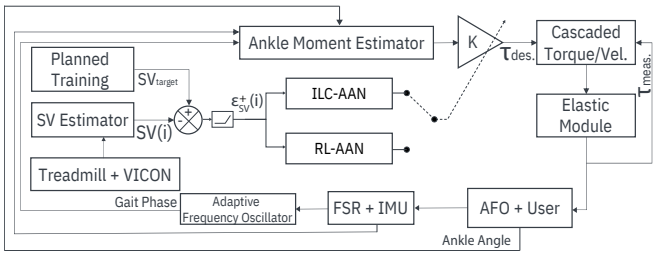

Fig. 2. Control architecture, integrating the proposed ILC-AAN and RL-AAN controller with a subject-agnostic ankle moment estimator [22].

troller), while $K(i) = 10$ corresponds to the maximum applicable assistance.

C. RL-AAN

We implement an actor–critic Action-Dependent Heuristic Dynamic Programming (ADHDP) controller adapted from [23] to elicit gait adaptations matching or exceeding a target SV (SV_{target}) while minimizing assistance. The state is the positive SV error

$$\varepsilon^{+}_{SV}(i) = \max(0, SV_{\text{target}} - SV(i)), \tag{2}$$

where $SV(i)$ is the stride velocity at the i-th gait cycle. We emphasize that (2) represents the positive part of the error, since the controller is designed to elicit improvements in SV, rather than achieve precise tracking of SV_{target}. The critic estimates the long-term cost of the current policy $\tilde{V}(i)$ by minimizing the temporal-difference error

$$t_d(i) = r(i) + \gamma\,\tilde{V}(i) - \tilde{V}(i-1), \tag{3}$$

where $r(i)$ represent the immediate negative reward:

$$r(i) = \tfrac{1}{2}\,\varepsilon^{+}_{SV}(i)^2. \tag{4}$$

Accordingly, we define the critic objective as the quadratic loss

$$E_c(i) = \frac{1}{2}t_d^2(i). \tag{5}$$

The actor updates $K(i)$ after each initial contact, by minimizing the cost

$$E_a(i) = \tfrac{1}{2}\left[\tilde{V}(i)\,(1 - U_c(i)) + K(i)\,U_c(i)\right]^2, \tag{6}$$

where the control objective $U_c(i) \in [0,1]$ is updated stride-by-stride to balance predicted long-term SV errors against assistance minimization through the following law

$$U_c(i{+}1) = \begin{cases} \min\big(1,\, U_c(i) + \beta\big), & \overline{\varepsilon}^{+}_{SV}(i) = 0, \\ \max\big(0,\, U_c(i) - \beta\big), & \overline{\varepsilon}^{+}_{SV}(i) > 0\ , \end{cases} \tag{7}$$

where $\overline{\varepsilon}^{+}_{SV}(i)$ denotes the arithmetic mean of the positive SV errors computed over the last m strides. As SV approaches SV_{target}, U_c progressively increases in increments of β, driving $K(i)$ toward weaker assistance. Conversely, when SV falls behind, U_c decreases at a similar rate to prioritize error reduction.

Actor and critic are single-hidden-layer feedforward neural networks with sigmoid activations, trained online at each stride. Based on preliminary tests, the following hyperparameters were used: $\gamma = 0.5$, $m = 10$ strides, $\beta = 0.1$. More detail about the RL-AAN implementation can be found in [23].

D. ILC-AAN

Iterative Learning Control enhances performance in repetitive tasks by leveraging past errors to adjust a device's response in the subsequent repetition. Given its central role in previous AAN implementations [10], [12], we adopted a simple proportional-type (P-type) ILC as the benchmark for RL-AAN. The ILC-AAN updates the assistance gain once per gait cycle according to the following law

$$K(i+1) = \gamma_{\text{ILC}}\,K(i) + f(\varepsilon^{+}_{SV}(i)), \tag{8}$$

where $\gamma_{\text{ILC}} \in (0,1)$ is the discount factor and $f(\varepsilon^{+}_{SV}(i))$ is the error function evaluated at the i-th gait cycle:

$$f(\varepsilon^{+}_{SV}(i)) = g_{\text{ILC}}\,\varepsilon^{+}_{SV}(i) - b. \tag{9}$$

Constant parameters $g_{\text{ILC}} > 0$ and $b \geq 0$ represent the error gain and the intercept, respectively, while $\varepsilon^{+}_{SV}(i)$ is the positive SV error at the i-th gait cycle, computed using (2). Thus, a persistent $\varepsilon^{+}_{SV}(i)$ causes the assistance gain to increase, while the discount factor reduces $K(i)$ when errors are small. With appropriate choices of γ_{ILC} and g_{ILC}, the update (8) remains bounded and convergent under steady conditions [10]. From preliminary testings, we selected $\gamma_{\text{ILC}} = 0.85$, $g_{\text{ILC}} = 4$, and $b = 0$.

III. Experimental Setup

A. Protocol

Seven healthy male volunteers (weight 83.5 ± 11.5 kg, age 23.0 ± 3.0 yrs, height 180.0 ± 6.0 cm) completed the study. Procedures were approved by the U.S. Army Office of Human Research Oversight (OHRO; protocol E03214.1a-1), and all participants provided written informed consent.

Self-selected walking speed was estimated using a self-paced treadmill realized by configuring a split-belt instrumented system (BERTEC) with a feedback loop that used embedded force plates and a 9-camera motion-capture system (VICON Vero), as described [6]. A single reflective marker attached to the posterior side of the exoskeleton's shoe bracket enabled accurate SV estimation at each heel strike [26], while allowing natural speed adjustments in response to perturbations from ankle weights (2 kg total; Fig. 1) and assistive torques. Stride-by-stride SV estimates $SV(i)$ were sent via Wi-Fi over UDP to the onboard real-time target machine to compute per-stride controller updates.

The protocol comprised three lab visits. Visit 1 included a leg scan for fabricating the subject-tailored ankle unit [21]. In Visits 2 and 3, participants donned their personalized AFO on the non-dominant leg, which was selected for its stabilizing

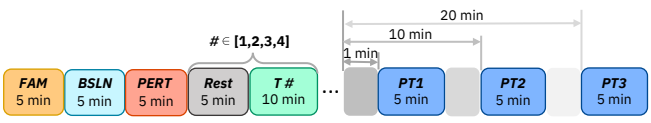

Fig. 3. Experimental protocol.

role during gait [27] and walked under either RL-AAN or ILC-AAN. Controller order was balanced across participants, and each participant's Visits 2 and 3 were separated by at least 48 hours to enable sufficient washout.

As indicated in Fig. 3, each visit involved 10 walking bouts, consistent with our previous work [23]. Familiarization (FAM) involved walking both with and without ankle weights (2.5 min each) under a zero-torque controller ($K(i) \equiv 0$) to help participants acclimate to the setup. Baseline ($BSLN$) consisted of walking without ankle weights under the zero-torque controller. SV_{target} was determined as the mean SV measured during the final minute of $BSLN$. Additionally, for RL-AAN, we used each participant's $BSLN$ data to pretrain the critic network. After establishing this baseline, ankle weights were applied and kept in place for the remainder of the visit. Perturbation ($PERT$) matched the duration of $BSLN$, but was performed with ankle weights, allowing us to quantify the effect of added inertia on participants' SV. Training ($T\#$) comprised four 10-min bouts with the assistive controller active (RL-AAN or ILC-AAN), separated by 5-min rest periods to avoid fatigue. Finally, Post-Training ($PT\#$) included three 5-min bouts performed under the zero-torque controller at 1, 10, and 20 min after the final training bout, to assess short-term retention of SV improvements over a 25-min window.

B. Statistical Analysis

The primary outcome was selected as the average per-session percentage change in ε^{+}_{SV} relative to the unassisted perturbed-walking bout:

$$\Delta\%\varepsilon^{+}_{SV,*} = 100 \cdot \frac{\varepsilon^{+}_{SV,PERT} - \varepsilon^{+}_{SV,*}}{\varepsilon^{+}_{SV,PERT}}, \qquad (10)$$

where $\varepsilon^{+}_{SV,*}$ denotes ε^{+}_{SV} averaged over walking bout $*$. As a secondary outcome, we analyzed the per-session percentage of gait cycles at or above SV_{target}, indicated as $N\%^{+}_{SV}$. Larger $\Delta\%\varepsilon^{+}_{SV}$ values reflected greater SV improvement relative to $PERT$, while higher $N\%^{+}_{SV}$ percentages indicated better adherence to the target walking speed. The level of exoskeleton assistance was quantified as the session-averaged assistance gain $\hat{K}$. The session-averaged stride-to-stride difference in $K(i)$, denoted as ΔK, was also computed to evaluate the controller responsiveness.

Given the small sample size, we did not rely on parametric tests. To begin, the effects of ankle weights on subjects' baseline SV were assessed using one-sample Wilcoxon signed-rank tests applied to BSL–$PERT$ SV differences. Subsequently, dependent variables collected from training and retention bouts were analyzed separately using aligned-rank transform (ART) repeated-measures ANOVA, with Controller (C) and *Session* (S) as factors. When a significant effect of S was observed, focused post-hoc comparisons were conducted on ART-transformed variables using paired t-tests—$T4$ vs. $T1$ for training and $PT3$ vs. $PT1$ for retention—to evaluate start-to-finish changes in SV. All analyses were performed in IBM SPSS Statistics 29.0 with the significance level set at $\alpha = .05$.

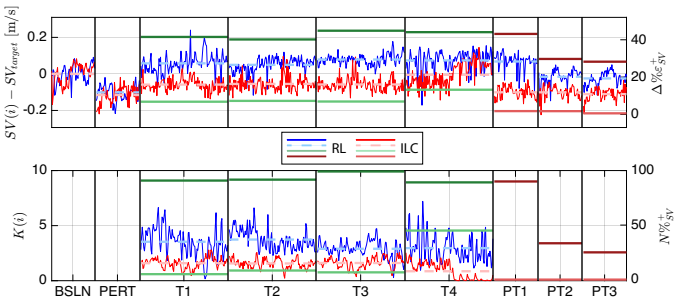

Fig. 4. Trends of $(SV(i) - SV_{\text{target}})$ and $K(i)$ from a representative study participant, under ILC-AAN and RL-AAN. Dashed horizontal lines indicate per-session averages; solid horizontal lines show session-level outcomes ($\Delta\%\varepsilon^{+}_{SV}$ and $N\%^{+}_{SV}$).

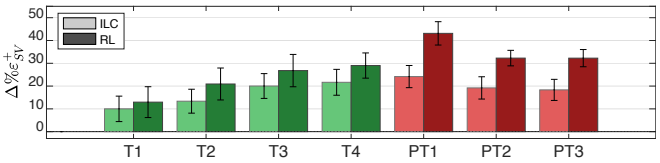

Fig. 5. Group averages of $\Delta\%\varepsilon^{+}_{SV}$ across training and post-training sessions for ILC-AAN and RL-AAN.

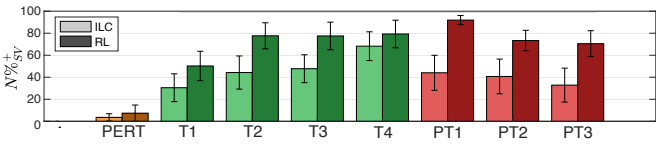

Fig. 6. Group average percentage of $N\%^{+}_{SV}$ across perturbation, training and post-training sessions for ILC-AAN and RL-AAN.

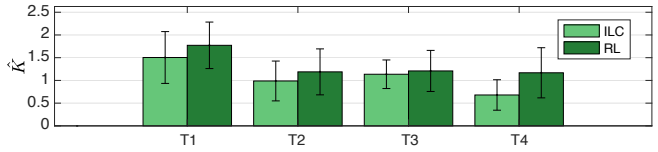

Fig. 7. Group average of $\hat{K}$ across training sessions for ILC-AAN and RL-AAN.

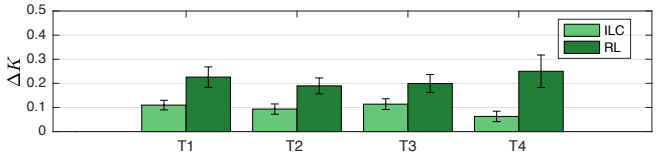

Fig. 8. Group average of ΔK across training sessions for ILC-AAN and RL-AAN.

TABLE I

ART RM-ANOVA p-VALUES DURING T AND PT.

		C	S	C×S
T	$\Delta\%\varepsilon^{+}_{SV}$	ns	$p < .05$	ns
	$N\%^{+}_{SV}$	$p < .01$	$p < .05$	ns
	$\hat{K}$	ns	ns	ns
	ΔK	$p < .05$	ns	ns
PT	$\Delta\%\varepsilon^{+}_{SV}$	$p < .05$	$p < .05$	ns
	$N\%^{+}_{SV}$	$p < .01$	ns	ns

C = Controller; S = Session.

IV. Results

As expected, the ankle weight disrupted participants' self-selected, unperturbed SV. The decrease in SV from $BSLN$ to $PERT$ averaged 18.4 cm/s and 16.1 cm/s for the RL-AAN and ILC-AAN visits, respectively, and was significant ($p < .001$) in both cases. However, no significant differences were found between visits, consistent with identical controller behavior when $K(i) \equiv 0$. Figure 4 shows the trends in $(SV(i) - SV_{\text{target}})$, $K(i)$, $\Delta\%\varepsilon^{+}_{SV}$, and $N\%^{+}_{SV}$ for a representative participant. With ankle weights and no assistance, the SV during $PERT$ fell below the $BSLN$ average, i.e., the assigned SV_{target}. During training, both RL-AAN and ILC-AAN elicited gradual improvements in SV, approaching or even exceeding SV_{target}. These gains were generally retained at $PT1$ for RL-AAN, though there was a noticeable decline in performance across the post tests. Post-test performance decline was more marked for ILC-AAN, indicating worse retention. The assistance gain $K(i)$ demonstrated higher stride-to-stride variability in RL-AAN compared to ILC-AAN. These trends were confirmed by the group averages, as described in Sec. IV-A and Sec. IV-B.

A. Training

$\Delta\%\varepsilon^{+}_{SV}$ improved significantly during training (Fig. 5), irrespective of controller type (main effect of S, $p < .05$, Tab. I). Post-hoc analyses on pooled data confirmed significant improvement from $T1$ to $T4$ ($p < .05$), indicating that participants were able to progressively compensate for the ankle weight. $N\%^{+}_{SV}$ also improved significantly during training (Fig. 6), irrespective of controller type (main effect of S, $p < .05$), again with significant improvement observed from $T1$ to $T4$ ($p < .05$) indicating improved adherence to SV_{target}. Interestingly, participants took more strides at or above SV_{target} when walking with RL-AAN compared to ILC-AAN (main effect of C, $p < .01$). No significant differences were found in the level of assistance provided by the two controllers (Fig. 7), though RL-AAN demonstrated higher ΔK across sessions ($p < .01$), indicating higher stride-to-stride variability (Fig. 8). Inspection of subject-level trajectories confirmed that RL-AAN made larger and more frequent stride-to-stride adjustments than ILC-AAN.

B. Retention

During PT bouts, RL-AAN outperformed ILC-AAN in terms of recovery from the perturbed SV (main effect of C, $p < .05$, Tab. I), though $\Delta\%\varepsilon^{+}_{SV}$ declined for both controllers (main effect of S, $p < .05$). Post-hoc analyses on pooled data confirmed decline from $PT1$ to $PT3$ ($p < .05$), indicating a trend toward the unassisted perturbed SV following training (Fig. 5). RL-AAN also outperformed ILC-AAN in terms of post-training adherence to SV_{target} (main effect of C, $p < .01$), with no significant changes in $N\%^{+}_{SV}$ observed across PT bouts (Fig. 6). Taken together, these results suggest that RL-AAN may promote enhanced SV retention compared to ILC-AAN, which manifest in lower average SV errors ($\Delta\%\varepsilon^{+}_{SV}$) and greater adherence to SV_{target} ($N\%^{+}_{SV}$).

V. Discussion and Conclusion

This study presented a head-to-head comparison of conventional and state-of-the-art AAN control strategies for gait training, both implemented on the same ankle exoskeleton and operating under an identical low-level cascaded torque–velocity controller. This design isolated the control policy as the independent variable of interest, while the experimental protocol—leveraging a self-paced treadmill and ankle-weight perturbations—provided a repeatable and ecologically valid challenge to healthy adults.

During training, both controllers progressively reduced perturbation-related deficits in self-selected SV, with no significant differences between controllers. However, participants showed better adherence to SV_{target} with RL-AAN, as indicated by higher $N\%^{+}_{SV}$. Average assistance magnitude ($\hat{K}$) did not differ between controllers and showed no consistent session-wise trend, whereas stride-to-stride assistance modulation (ΔK) was higher with RL-AAN. In post-training tests, SV recovery relative to the perturbed speed ($\Delta\%\varepsilon^{+}_{SV}$) declined across PT sessions for both controllers, while adherence to SV_{target} ($N\%^{+}_{SV}$) remained relatively stable. Importantly, RL-AAN elicited better retention than ILC-AAN in terms of both SV recovery and adherence to SV_{target}. $\Delta\%\varepsilon^{+}_{SV}$ and $N\%^{+}_{SV}$ capture distinct motor behaviors, as confirmed by secondary analyses of bout-averages values showing no significant correlation. $\Delta\%\varepsilon^{+}_{SV}$ reflects the average recovery toward the target relative to the perturbed baseline, whereas $N\%^{+}_{SV}$ quantifies how often strides meet or exceed the target. A policy that induces more frequent SV threshold crossings without changing the mean SV may raise $N\%^{+}_{SV}$ even in the absence of significant differences in $\Delta\%\varepsilon^{+}_{SV}$. This pattern is consistent with the elevated stride-to-stride assistance modulation (ΔK) observed under RL-AAN and the absence of significant between-controller differences in assistance magnitude ($\hat{K}$).

Two factors may account for the superior retention performance observed with RL-AAN. First, the actor–critic policy balances tracking errors against assistance penalties through the control objective U_c, while also estimating future cumulative SV costs. As a result, RL-AAN delivers assistance aimed at optimizing expected long-term outcomes, rather than merely correcting the most recent SV errors as in ILC-AAN. Second, RL-AAN induced greater stride-to-stride variability in assistance while providing, on average, similar levels of assistance as ILC-AAN. Within each training bout, RL-AAN dynamically scaled assistance based on performance, thereby promoting user engagement without increasing mean assistance. Such stride-to-stride modulation may encourage active error-based corrections and broader motor exploration during gait training, facilitating more refined motor adaptations after training. Future research is warranted to determine whether RL-AAN's post-training benefits can translate into more effective rehabilitation programs for HELET patients.

An important limitation of this study is the absence of physiological measures to directly quantify participant engagement. We did not collect metabolic cost or muscle

activations, thus physical engagement was inferred from motor behaviors. Additionally, the small, homogeneous healthy cohort and treadmill-only protocol limit generalizability to overground training in clinical populations with neurological or musculoskeletal gait impairments. Moreover, the feasible interval for the assistance gain $K(i)$ prioritized comfort and feasibility; however, broader or adaptive bounds may be needed for clinical cohorts with larger deficits. Looking forward, studies on RL-AAN should evaluate overground, examine multi-session exercise programs to explore the feasibility of long-term retention, and expand investigations to clinical populations with gait impairments. Multi-objective RL that incorporates symmetry, push-off power, or metabolic cost while preserving assistance minimization is a promising direction. Taken together, the present findings support RL-AAN as a subject-tailored strategy for gait-speed training that enhances user engagement during training and induces short-term motor adaptations afterward.

ACKNOWLEDGMENT

This work was supported by the U.S. National Science Foundation under Grant CMMI-1944203 and by the US Department of Defense, through the Peer Reviewed Orthopaedic Research Program (PRORP), under Award No. W81XWH-22-1-0193. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Department of Defense.

REFERENCES

[1] W. T. Gordon, K. Kuhn, G. Staeheli, and D. Dromsky, "Challenges in definitive fracture management of blast injuries," *Curr. Rev. Musculoskelet. Med.*, vol. 8, no. 3, pp. 290–297, 2015.

[2] C. K. Black, L. D. Ormiston, K. L. Fan, V. S. Kotha, C. Attinger, and K. K. Evans, "Amputations versus Salvage: Reconciling the Differences." *J. Reconstr. Microsurg.*, vol. 37, no. 1, pp. 32–41, Jan. 2021, place: United States.

[3] M. J. Highsmith, L. M. Nelson, N. T. Carbone, T. D. Klenow, J. T. Kahle, O. T. Hill, J. T. Maikos, M. S. Kartel, and B. J. Randolph, "Outcomes Associated With the Intrepid Dynamic Exoskeletal Orthosis (IDEO): A Systematic Review of the Literature," *Mil. Med.*, vol. 181, no. suppl_4, pp. 69–76, Nov. 2016.

[4] G. Orekhov, Y. Fang, J. Luque, and Z. F. Lerner, "Ankle exoskeleton assistance can improve over-ground walking economy in individuals with cerebral palsy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 461–467, 2020.

[5] L. H. Sloot, L. M. Baker, J. Bae, F. Porciuncula, B. F. Clément, C. Siviy, R. W. Nuckols, T. Baker, R. Sloutsky, D. K. Choe, K. O'Donnell, T. D. Ellis, L. N. Awad, and C. J. Walsh, "Effects of a soft robotic exosuit on the quality and speed of overground walking depends on walking ability after stroke," *J. NeuroEng. Rehabil.*, vol. 20, no. 1, p. 113, 2023.

[6] 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.

[7] E. R. Esposito, K. A. Schmidtbauer, and J. M. Wilken, "Experimental comparisons of passive and powered ankle-foot orthoses in individuals with limb reconstruction," *J. NeuroEng. Rehabil.*, vol. 15, no. 1, pp. 1–10, 2018, publisher: BioMed Central.

[8] L. N. Awad, A. Esquenazi, G. E. Francisco, K. J. Nolan, and A. Jayaraman, "The rewalk restore™ soft robotic exosuit: a multi-site clinical trial of the safety, reliability, and feasibility of exosuit-augmented post-stroke gait rehabilitation," *J. NeuroEng. Rehabil.*, vol. 17, no. 1, p. 80, 2020.

[9] V. S. Huang and J. W. Krakauer, "Robotic neurorehabilitation: a computational motor learning perspective," *J. NeuroEng. Rehabil.*, vol. 6, p. 5, 2009.

[10] 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," *J. NeuroEng. Rehabil.*, vol. 4, pp. 1–16, 2007.

[11] Q. Zhang, D. Zanotto, M. Sharifi, M. Kim, and Z. Li, *Closing the Loop Between Wearable Robots and Machine Learning: A New Paradigm for Steering Assistance Personalization Control*. Cham: Springer Nature Switzerland, 2024, pp. 65–101.

[12] S. Maggioni, N. Reinert, L. Lünenburger, and A. Melendez-Calderon, "An adaptive and hybrid end-point/joint impedance controller for lower limb exoskeletons," *Front. Robot. AI*, vol. 5, p. 104, 2018.

[13] 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 Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 6, pp. 956–963, 2015.

[14] Y. Zhang, S. Li, K. J. Nolan, and D. Zanotto, "Adaptive Assist-as-needed Control Based on Actor-Critic Reinforcement Learning," in *2019 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. Macau, China: IEEE, Nov. 2019, pp. 4066–4071.

[15] ——, "Reinforcement learning assist-as-needed control for robot assisted gait training," in *2020 8th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatron. (BioRob)*, 2020, pp. 785–790.

[16] ——, "Shaping individualized impedance landscapes for gait training via reinforcement learning," *IEEE Trans. Med. Robot. Bionics*, vol. 4, no. 1, pp. 194–205, 2022.

[17] A. Li, R. Minto, M. Dölling, G. Boschetti, and D. Zanotto, "Personalized adaptive assistance with reinforcement learning control enhances engagement, performance, and retention in robot-assisted arm-reaching exercises," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 34, pp. 532–542, 2025.

[18] X. Tu, M. Li, M. Liu, J. Si, and H. H. Huang, "A data-driven reinforcement learning solution framework for optimal and adaptive personalization of a hip exoskeleton," in *2021 Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 10 610–10 616.

[19] S. Luo, G. Androwis, S. Adamovich, E. Nunez, H. Su, and X. Zhou, "Robust walking control of a lower limb rehabilitation exoskeleton coupled with a musculoskeletal model via deep reinforcement learning," *J. NeuroEng. Rehabil.*, vol. 20, no. 1, p. 113, 2023.

[20] J. de Miguel-Fernández, J. Lobo-Prat, E. Prinsen, J. M. Font-Llagunes, and L. Marchal-Crespo, "Control strategies used in lower limb exoskeletons for gait rehabilitation after brain injury: a systematic review and analysis of clinical effectiveness," *J. NeuroEng. Rehabil.*, vol. 20, no. 1, p. 23, Feb. 2023.

[21] M. T. Eraky, A. Li, M. H. Rocha, A. Teker, B. A. Gebre, K. J. Nolan, K. Pochiraju, and D. Zanotto, "A novel personalized ankle exoskeleton with co-located sea for gait training," in *2024 10th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatron. (BioRob)*, 2024, pp. 1715–1720.

[22] Q. Zhao, R. Deepak, B. A. Gebre, K. J. Nolan, K. Pochiraju, and D. Zanotto, "Gaussian process regression models for on-line ankle moment estimation in exoskeleton-assisted walking," in *2024 10th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatron. (BioRob)*. IEEE, 2024, pp. 1171–1176.

[23] A. Li, H. Li, A. Teker, M. H. Rocha, B. A. Gebre, K. J. Nolan, K. Pochiraju, and D. Zanotto, "Reinforcement learning assist-as-needed control promotes recovery of walking speed following ankle weight perturbations," in *2025 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2025, pp. 46–51.

[24] H. M. Schmidt, A. Li, A. Teker, M. H. Rocha, B. A. Gebre, K. J. Nolan, K. Pochiraju, and D. Zanotto, "Digital design workflow for individualized 2-dof ankle exoskeletons," in *2025 Int. Conf. Rehabil. Robot. (ICORR)*, 2025, pp. 1–6.

[25] M. T. Eraky, A. Li, B. A., K. Pochiraju, and D. Zanotto, "A compact rotary series elastic actuator with wide deflection range and linear torque response for phri applications," in *2026 Int. Conf. Robot. Autom. (ICRA)*, 2026.

[26] H. Zhang, Y. Guo, and D. Zanotto, "Accurate ambulatory gait analysis in walking and running using machine learning models," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 191–202, 2020.

[27] H. Sadeghi, P. Allard, F. Prince, and H. Labelle, "Symmetry and limb dominance in able-bodied gait: a review," *Gait & Posture*, vol. 12, no. 1, pp. 34–45, 2000.