Reinforcement Learning-Based Adaptive Biofeedback Engine for Overground Walking Speed Training

Huanghe Zhang*†, Shuai Li*†, Qingya Zhao†, Ashwini K. Rao‡, Yi Guo§, Damiano Zanotto†

***Abstract*—Wearable biofeedback systems (WBS) have been proposed to aid physical rehabilitation of individuals with motor impairments. Due to significant inter- and intra-individual differences, the effectiveness of a given biofeedback strategy may vary for different users and across therapeutic sessions, as a patient's functional recovery progresses. To date, only a paucity of research has investigated the use of biofeedback strategies that can self-adapt based on the user's response. This letter introduces a novel reinforcement learning with fuzzy logic biofeedback engine (RLFLE) for personalized overground walking speed training. The method leverages reinforcement learning and a fuzzy inference strategy to continuously modulate underfoot vibrotactile stimuli that encourage users to achieve a target walking speed. This stimulation strategy also enables the determination of a user's maximum steady-state walking speed during a gait training session overground. The RLFLE was implemented in a custom-engineered WBS and validated against two simpler biofeedback strategies during walking tests with healthy adults. Participants showed lower walking speed errors when training with the RLFLE. Additionally, results indicate that the new method is more effective in determining an individual's maximum steady-state walking speed. Given the importance of walking speed as an indicator of health status and as an essential outcome of exercise-based interventions, these results show promise for implementation in future technology-enhanced gait rehabilitation protocols.**

***Index Terms*—Wearable biofeedback system, instrumented footwear, human-in-the-loop, reinforcement learning, fuzzy logic, gait training.**

I. Introduction

WALKING speed is a valid predictor of health status in a wide range of clinical populations [1]. Low walking speed is associated with increased risk of falls, cognitive impairment, and institutionalization in older adults [2], [3]. Conversely, improvements in walking speed have been correlated with better quality of life in patients with Parkinson's disease [4], multiple sclerosis [5], stroke [6] and cerebral palsy [7]. Moreover, self-selected gait speed represents a major outcome in many exercise-based rehabilitation programs [8].

The use of external cueing in gait exercises has been shown to elicit significant improvements in walking speed [9]. Past research has focused on open-loop strategies using various modalities, such as visual, auditory, and tactile cues [4], [10], [11]. While open-loop stimulation methods are relatively easy to implement, these strategies have several limitations. First, because gait speed is proportional to the product of cadence and stride length, it is nearly impossible to accurately predict a target improvement in walking speed by eliciting a fixed increment in either variable [12]. Second, due to intra- and inter-individual variability, the stimulation parameters might only suit a given individual at a given time [13]. For these reasons, tuning the stimulation parameters for open-loop strategies often requires multiple readjustments [14]. Third, open-loop stimulation methods such as rhythmic auditory stimulation are typically applied to straight-line walking tasks and assume that individuals walk at a consistent velocity [11]. Hence, these methods may not generalize to real-life walking tasks, which involve changes in speed and direction.

Recent advances in wearable technologies have paved the way for self-administered gait training protocols in patients' living environments [15]. Wearable biofeedback systems (WBS) represent a promising approach in gait rehabilitation, since their gait analysis capability may inform feedback engines that provide closed-loop visual, auditory, or tactile stimulation to encourage changes in a person's gait patterns, and may also help to assess a patient's progress longitudinally [16]. Visual cues delivered by augmented-reality glasses can be adjusted in real-time according to the user's current walking speed [17]; auditory [5] or vibrotactile [18] stimuli can be triggered at specific gait events to reduce gait variability. More recently, Yasuda et al. proposed a vibratory cueing system to improve the walking speed in stroke survivors. Their system progressively increases the frequency of the rhythmic cues during a walking session, following a predefined, empirically-tuned law [19]. Wu et al. applied Gaussian process regression to generate a personalized model of the user's response to rhythmic auditory stimuli offline. Then, they used the model within an on-line optimization algorithm to elicit desired changes in the user's cadence [13]. Despite the growing body of research on WBS for gait training, how to best design a biofeedback engine to elicit a target walking speed during overground exercises is still an open problem.

Manuscript received: January, 27, 2022; Revised: May, 14, 2022; Accepted: June, 14, 2022.

This paper was recommended for publication by Editor Jee-Hwan Ryu upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the US National Science Foundation under Grants IIS-1838799, IIS-1838725, and CMMI-1944203.

*These authors contributed equally to this work.

†HZ, SL, QZ, and DZ (dzanotto@stevens.edu) are with the Dept. of Mechanical Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA.

‡AKR is with the Dept. of Rehabilitation & Regenerative Medicine, Columbia University, New York, NY 10032, USA.

§YG is with the Dept. of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA.

Digital Object Identifier (DOI): see top of this page.

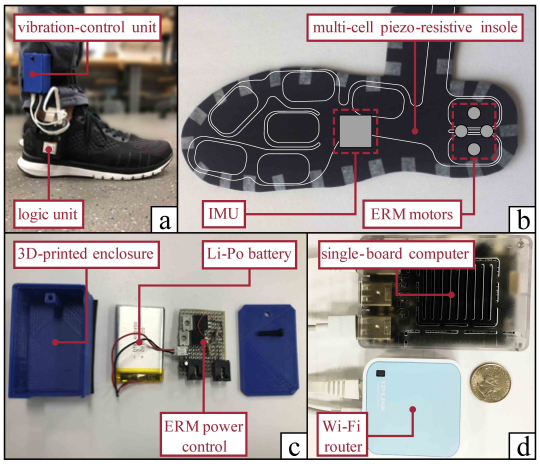

Fig. 1. The WBS **(a)** consists of two insole modules **(b)**, a vibration-control unit **(c)**, and a single-board computer **(d)**.

Our group has recently developed a minimally obtrusive WBS that measures stride velocity and gait phase in real-time, Fig. 1. These two variables are fed to a closed-loop biofeedback engine to elicit desired changes in the wearer's walking speed [12]. Similar to open-loop stimulation methods, the closed-loop biofeedback engine induces changes in the wearer's walking speed by adjusting their cadence via rhythmic stimuli. However, unlike open-loop strategies, the method proposed in [12] adjusts the gait phase at which the stimuli are delivered based on the measured velocity errors, using anticipatory and delayed feedback to elicit automatic modulations in the wearer's cadence that lead to systematic adjustments in walking speed. This approach is motivated by the theory of action-perception coupling [20], which states that human movement is controlled by comparing expected feedback (generated by internal models) and actual feedback (i.e., reafference) resulting from a motor command, and this comparison leads to movement calibration. Consequently, the execution of a motor task may be affected in a predictable fashion by artificially manipulating sensory reafferences [21]. While this method has shown promising results, both with constant and with time-varying target walking speeds, it relies on the simplistic assumption that users are always able to adapt to the prescribed target speed. However, how to automatically determine an appropriate overground target walking speed for gait exercises is a challenging task, given considerable inter- and intra-personal differences [13]. Moreover, when an individual can no longer follow the stimuli (e.g., due to fatigue), the target speed should be adapted accordingly, to maintain user engagement in the training task.

To address these two limitations, herein we propose a novel reinforcement learning with fuzzy logic bioefeedback engine (RLFLE) that enables adaptive and personalized walking speed training. The objective of the reinforcement learning (RL) framework is to learn an optimal policy for an autonomous agent through repeated interactions with its environment [22]. The RL architecture is particularly suitable for complex systems that are difficult to model, such as the human motor control system. Although RL has proven to be an effective control strategy for lower-extremity powered prostheses [23] and orthoses [24], [25], none of the past studies has investigated the application of RL to biofeedback systems for gait training. In the proposed implementation, a fuzzy inference strategy is introduced to embed expert knowledge into the RL biofeedback engine. Compared with our previous approach described in [12], the RLFLE enables the decoupling of the *target gait speed* (i.e., the goal speed for the current training bout) from the *guided gait speed* (i.e., the speed on which the vibrotactile stimuli are computed). Specifically, the RLFLE learns how to best adapt the guided speed on-line, to keep the wearer engaged in the walking exercise, while progressively directing him/her towards the target speed.

The contributions of this work can be summarized as follows: (i) a novel RL-based biofeedback engine (RLFLE) for personalized walking speed training; (ii) a new method to determine an individual's maximum steady-state overground walking speed, based on the proposed RLFLE; (iii) a validation of the RLFLE in relation to simpler stimulation strategies, namely the *constant speed* (CS) method and the *constant increment* (CI) method. The CS method is equivalent to our previous work [12], which assumes that the target and the guided speed are coincident. The CI method is based on hard-coded update rules to adjust the guided speed. These methods were included to evaluate whether the increased complexity of the RLFLE is well justified by improved adaptability and user performance compared to non-adaptive methods. The remainder of this letter is organized as follows. Section II describes the WBS; Section III introduces the RLFLE. The experimental protocol and data analysis are illustrated in Section IV. Results are presented in Section V and discussed in Section VI. Finally, the letter is concluded in Section VII.

II. Wearable Biofeedback System

The WBS (Fig. 1) builds upon our previous work on instrumented footwear [26]–[32]. It features two insole modules, a vibration-control unit, and a single-board computer (SBC). Each insole module consists of a eight-cell piezoresistive sensor (IEE Inc., Luxemburg), a nine-DOF (degree-of-freedom) inertial measurement unit (IMU, Yost Labs Inc., Portsmouth, OH, USA), four eccentric rotating mass (ERM) motors, a logic unit, and a Li-Po battery. The piezoresistive sensor array is used to estimate normal ground reaction forces (GRF). The logic unit (32-bit ARM Cortex-M4, PJRC, Sherwood, OR, USA) runs the low-level closed-loop vibrotactile control (Sec. III-A) and activates the ERM motors through the vibration-control unit. The SBC fits inside a running belt that can be worn by the user or can be optionally located off-board within a 30-meter range from the user. A small Wi-Fi router connected to the SBC serves as an access point for the WBS. The SBC runs the RLFLE (Sec. III-B) and the data logger software. The WBS is lightweight (120 g) and fits in the user's shoes. It can be donned in less than 5 minutes. More details about the design of the WBS can be found in [12].

III. Adaptive Biofeedback Control with RLFLE

The overall control architecture of the WBS is shown in Fig. 2. The high-level control consists of the RLFLE, which

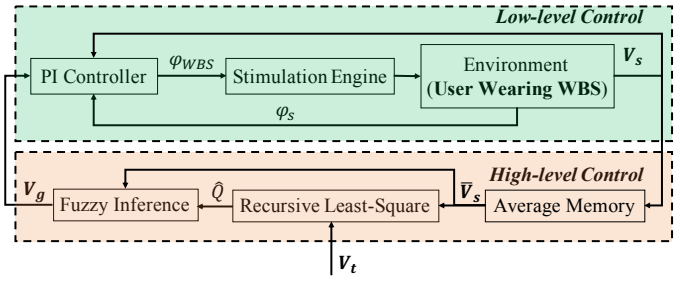

Fig. 2. Flowchart of the RLFLE. V_s is the subject's current gait speed. The average memory block uses the most recent 6 strides to compute $\bar{V}_s$, the average gait speed. V_g, V_t, φ_s and φ_{WBS} are the guided speed, the target speed, the user's current gait phase and the target phase, respectively. $\hat{Q}$ is the long-term cost of the RLFLE.

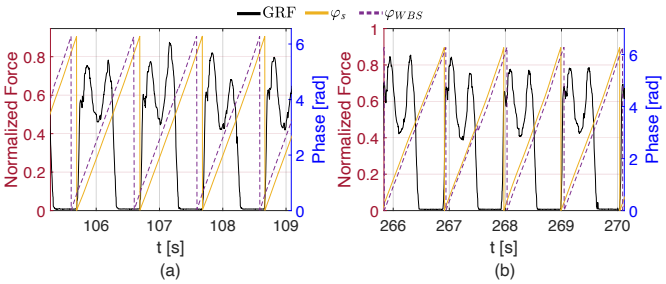

Fig. 3. Effects of closed-loop vibrotactile stimuli on the gait of a representative study participant. The black line represents the normalized GRF extracted from the WBS insole. φ_s and φ_{WBS} are the subject's current gait phase and the WBS phase, respectively. The stimuli are triggered every time φ_{WBS} crosses zero. If the subject's current walking speed is slower than the guided speed ($V_s < V_g$), the stimuli lead the initial contacts (IC) to encourage a faster pace **(a)**; Conversely, if $V_s > V_g$, the stimuli lag the IC, to elicit a slower pace **(b)**.

includes an average memory block, a recursive least square block, and a fuzzy inference block. The RLFLE generates the new guided speed V_g by evaluating the subject's average walking speed ($\bar{V}_s$) over the most recent 6 strides in relation to the target speed (V_t) and the previous guided speed. The new V_g is then fed to the low-level controller as a feed-forward term. Through the output φ_{WBS}, the low-level controller modulates the gait phase at which discrete plantar vibrotactile stimuli are delivered to the wearer, with the goal of motivating him/her to follow V_g. In the following subsections, the low-level and the high-level controllers are described in detail.

A. Low-level Closed-loop Vibrotactile Control

The low-level controller consists of three modules: WBS, PI controller, and stimulation engine (Fig. 2). The WBS measures stride velocity (V_s) and phase of the gait cycle (φ_s) in real-time. The PI controller takes the measured V_s and the V_g generated by the RLFLE as the inputs, and computes the phase shift $\Delta\varphi$, which is added to φ_s to obtain the trigger phase φ_{WBS}. The latter determines the instant at which the stimulation engine will generate the plantar stimuli [12]. The estimation of V_s starts from the determination of heel strike, foot-flat, and toe-off events, based on the underfoot multi-cell piezo-resistive sensor. After compensating for gravity, V_s is computed by double integration of accelerometric signals with Zero Velocity Update and Velocity Drift Compensation techniques, as detailed in [27], [28], [30], [31]. The current gait phase φ_s is determined by a pool of adaptive frequency oscillators [33], which take the measured foot pitch angle (i.e., the sagittal-plane angle between the foot sole and the ground) as the input [12]. The stimulation engine delivers a plantar vibrotactile stimulus with pulse duration of 150 ms each time φ_{WBS} crosses zero. The PI controller modulates the phase shift $\Delta\varphi$ of the vibrotactile stimuli relative to the estimated IC. If $V_s < V_g$, the stimuli lead the IC to encourage a faster pace. Conversely, if $V_s > V_g$, the stimuli lag the IC to elicit a slower pace (Fig. 3).

B. Formalization of the RLFLE

We consider a Markov decision process described by the tuple $\langle S_k,\ A_k,\ T_k,\ R_k\rangle$, where k is the index of the RLFLE cycles, each corresponding to 6 consecutive strides[1]. We define the state and the action at the k-th cycle as $S_k := \bar{V}_{s,k}$ and $A_k := V_{g,k}$, where $\bar{V}_{s,k}$ and $V_{g,k}$ indicate the subject's average walking speed over the last 6 strides and the guided speed determined by the RLFLE, respectively. We assume a deterministic policy (i.e., transition probability $T_k \equiv 1$) and compute the stage cost $R_k := r_k$ from the velocity error vector $\boldsymbol{e}_k$ as follows:

$$\begin{aligned} r_k &= \frac{1}{2}\boldsymbol{e}_k^T \boldsymbol{\Lambda}_1 \boldsymbol{e}_k \\ \boldsymbol{e}_k &= \left[\bar{V}_{s,k} - V_{g,k},\ V_{g,k} - V_{t,k}\right]^T \end{aligned} \tag{1}$$

In the previous expression, $V_{t,k}$ indicates the target speed at the k-th cycle and $\boldsymbol{\Lambda}_1 \in \mathbb{R}^{2\times 2}$ is a diagonal weight matrix. The associated infinite horizon cost Q_k is given by

$$Q_k = \sum_{i=k}^{\infty} \gamma^{i-k}\, r_i, \tag{2}$$

where $\gamma = 0.8$ is a user-defined discount factor. Q_k can be approximated iteratively by minimizing the temporal difference error δ_k at each cycle [22]:

$$\delta_k = r_k + \gamma\, Q_{k+1} - Q_k \tag{3}$$

To do so, we first parameterize Q_k as

$$Q_k = \boldsymbol{W}^T \boldsymbol{\Phi}_k + \epsilon_k, \tag{4}$$

where $\boldsymbol{W} \in \mathbb{R}^3$ is the weight vector, $\boldsymbol{\Phi}_k = \left[\bar{V}_{s,k}^2,\ \bar{V}_{s,k}V_{g,k},\ V_{g,k}V_{t,k}\right]^T$ is the activation function, and $\epsilon_k \in \mathbb{R}$ indicates the approximation error. Substituting (4) into (3) yields:

$$\delta_k - \gamma\epsilon_{k+1} + \epsilon_k = r_k + \gamma \boldsymbol{W}^T \boldsymbol{\Phi}_{k+1} - \boldsymbol{W}^T \boldsymbol{\Phi}_k \tag{5}$$

The optimal $\boldsymbol{W}$ (in the least-square sense) can be determined by minimizing the squared sum of past residuals:

$$\begin{aligned} &\min \sum_{i=1}^{k-1} \left(\delta_i - \gamma\epsilon_{i+1} + \epsilon_i\right)^2 \\ &= \min \sum_{i=1}^{k-1} \left((\gamma\boldsymbol{\Phi}_{i+1} - \boldsymbol{\Phi}_i)^T \boldsymbol{W} - r_i\right)^2 \end{aligned} \tag{6}$$

[1]The number of consecutive strides affects the rate of adaptation of V_g and was empirically tuned during preliminary tests. If the value is too small, the user may find it hard to follow the varying V_g. If it is too large, it may take long for the user to adapt to V_g.

The resulting approximated long-term cost $\hat{Q}_k$ is given by

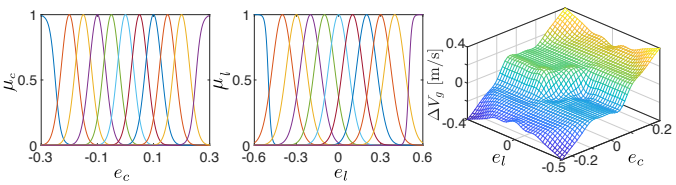

Fig. 4. Sigmoidal fuzzy membership functions mapping the crisp inputs e_c **(a)** and e_l **(b)** to fuzzified inputs. **(c)** Rule surface plot illustrating the mapping between e_c, e_l and crisp output ΔV_g.

$$\hat{Q}_k = \hat{\boldsymbol{W}}_k^T \boldsymbol{\Phi}_k, \tag{7}$$

where the recursive least-square (RLS) method is adopted to update the weight vector $\hat{\boldsymbol{W}}_k$ in real-time, using the following update rules:

$$\begin{aligned} \hat{\boldsymbol{W}}_k &= \hat{\boldsymbol{W}}_{k-1} + \boldsymbol{K}_k \left(r_k - \boldsymbol{\Phi}_k^T \hat{\boldsymbol{W}}_{k-1} \right) \\ \boldsymbol{K}_k &= \frac{\boldsymbol{P}_{k-1}\boldsymbol{\Phi}_k}{\alpha + \boldsymbol{\Phi}_k^T \boldsymbol{P}_{k-1}\boldsymbol{\Phi}_k} \\ \boldsymbol{P}_k &= \frac{1}{\alpha}\left(\boldsymbol{I}_m - \boldsymbol{K}_k \boldsymbol{\Phi}_k^T \right) \boldsymbol{\Phi}_{k-1} \end{aligned} \tag{8}$$

In the previous equations, $\alpha = 0.998$ is the forgetting factor[2], $\boldsymbol{P}_k, \boldsymbol{I}_m \in \mathbb{R}^{m \times m}$ are the inverse covariance matrix and the identity matrix, and $\boldsymbol{K}_k \in \mathbb{R}^m$ is the Kalman gain vector. The RLS method was selected for its low computational complexity. However, to improve the convergence rate of RLS, the initial values $\hat{\boldsymbol{W}}_0$ and $\boldsymbol{P}_0$ are determined by solving a conventional least square (LS) problem on the first batch of 10 RLFLE cycles (i.e., 60 strides) of the current walking bout. The convergence of this approach was demonstrated in [34].

The action V_g is generated at each RLFLE cycle by a fuzzy inference strategy that embeds expert rules into the RL architecture without significantly increasing the computational load of the high-level controller. The crisp inputs at the k-th cycle include the tracking error

$$e_{c,k} = \bar{V}_{s,k} - V_{g,k} \tag{9}$$

and the long-term feedback term

$$e_{l,k} = c \tanh\left(-\lambda \frac{\partial J_{l,k}}{\partial V_g} \right), \tag{10}$$

where c and λ are positive constant parameters and $J_{l,k}$ is the performance index defined as follows:

$$J_{l,k} = \frac{1}{2} \begin{bmatrix} \hat{Q}_k \\ r_k \end{bmatrix}^T \boldsymbol{\Lambda}_2 \begin{bmatrix} \hat{Q}_k \\ r_k \end{bmatrix} \tag{11}$$

$\boldsymbol{\Lambda}_2$ is a positive definite diagonal weight matrix controlling the relative importance of immediate and long-term costs. $e_{c,k}$ and $e_{l,k}$ are converted to fuzzified inputs based on two sets of $N = 11$ sigmoidal membership functions shown in Fig. 4(a-b), resulting in the degrees of membership $\mu_c^i(e_{c,k})$ and $\mu_l^j(e_{l,k})$, $i, j = 1, .., N$. The corresponding N^2 fuzzy outputs $y_{i,j}$ are computed through the fuzzy rule base summarized in Tab. A (Appendix). These inference rules were determined empirically through preliminary tests. The fuzzy outputs $y_{i,j}$ are then combined to determine the increment to the guided speed $\Delta V_{g,k}$ according to the center-of-gravity defuzzification method [35]:

$$\Delta V_{g,k} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} y_{i,j}\mu_c(e_{c,k})\,\mu_l(e_{l,k})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \mu_c(e_{c,k})\,\mu_l(e_{l,k})} \tag{12}$$

Finally, the new guided speed is computed as:

$$V_{g,k+1} = V_{g,k} + \Delta V_{g,k}. \tag{13}$$

The smooth relationship between inputs e_c, e_l and the crisp output ΔV_g resulting from (12) is illustrated in Fig. 4(c), and the complete RLFLE algorithm is summarized in Algorithm 1.

Algorithm 1 RLFLE algorithm

Initialize:
– collect 10 tuples $\{\bar{V}_{s,k}, V_{g,k}, r_k, \bar{V}_{s,k+1}\}$ to form batch $\mathcal{B}$
– get initial $\hat{\mathbf{W}}_0$ and $\hat{Q}_0$ using LS and (7).
while walking in progress **do**
  collect next 6 strides and update $\boldsymbol{\Phi}$, r
  update $\hat{Q}$ using (8), (7)
  get ΔV_g using (12)
  update the guided speed $V_g \leftarrow V_g + \Delta V_g$
end while

IV. Performance Evaluation

To highlight the advantages of the proposed RL-based stimulation method we compared the immediate effects of three stimulation methods on the gait of healthy individuals:

i) *Constant speed method (CS)*: This method only relies on the low-level closed-loop control described in Sec. III-A and hence it is equivalent to our previous work [12]. By following this method, the guided speed V_g is set to a constant value $V_g = V_t$ instead of being adapted on-the-fly according to the user's response. As a consequence, this method cannot determine an individual's maximum steady-state overground walking speed.

ii) *Constant increment method (CI)*: This method updates ΔV_g based on a myopic policy that only considers the current tracking error according to the following rule:

$$\Delta V_{g,k} = \begin{cases} 0.1 & |e_{c,k}| < 0.05\,m/s \\ -0.1 & e_{c,k} < -0.1\,m/s \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

The rationale behind this stimulation method is that the new guided speed $V_{g,k+1}$ computed from (13) and (14) will only be increased when the user is able to approximately follow $V_{g,k}$. Conversely, if the user cannot reach $V_{g,k}$, then $V_{g,k+1}$ is decreased by the same amount to help the user follow the vibrotactile stimuli and maintain user engagement. In all other cases, the method assumes that the user is capable of eventually following $V_{g,k}$ but needs more adaptation time, hence $\Delta V_{g,k}$ is not updated. Additionally, an upper bound V_t is set on $V_{g,k+1}$, to ensure V_g never exceeds V_t.

iii) *RL with fuzzy logic biofeedback engine (RLFLE)*: This method is described in Sec. III-A and Sec. III-B and represents the main contribution of this work.

[2] α controls the importance of recent observations relative to older observations. $\alpha = 1$ is equivalent to the conventional least squares algorithm, whereas $\alpha < 1$ gives more importance to recent observations.

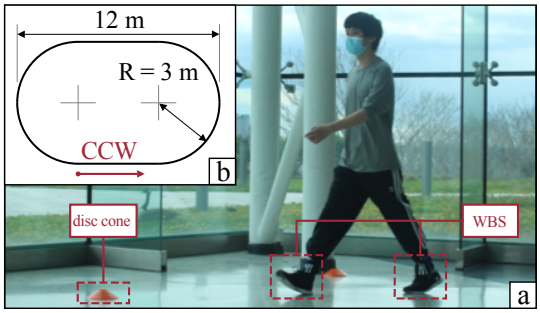

Fig. 5. Participants wearing the WBS were instructed to walk counter-clockwise **(a)** along an oval path delimited with disc cones **(b)**.

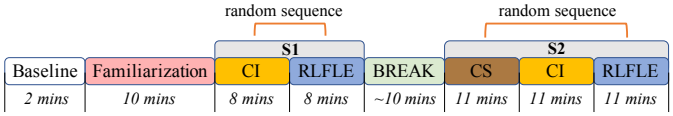

Fig. 6. Experimental protocol (CS = constant speed method, CI = constant increment method, RLFLE = RL with fuzzy logic biofeedback engine).

A. Experimental Protocol

Twelve healthy participants (9M, age 25.7±1.5 yr, height 1.72±0.07 m, weight 73.2±12.1 kg) volunteered for this study. All participants were healthy adults with no musculoskeletal or neurological problems that would affect their ability to walk for 12 minutes. The protocol was approved by the Institutional Review Board of Stevens Institute of Technology and all participants provided written consent.

After the system setup, participants were instructed to walk counter-clockwise (CCW) following an oval path marked on the floor with disc cones, Fig. 5. The oval path was selected instead of a straight-line path to test the RLFLE's performance in a task that includes both straight and curve walking. First, participants completed a 2-minute baseline test without stimuli, followed by a 10-minute familiarization session, Fig. 6. The last minute of the baseline test was used to determine the participant's average baseline stride velocity (V_b). During the familiarization session, participants experienced all three stimulation methods (CS, CI, and RLFLE) in random order. The goal of the familiarization session was to help participants get accustomed to the vibrotactile stimuli; therefore, for each stimulation method the value of V_t was set to V_b for about one minute, after which it was arbitrarily increased to a value within the range 1.0 to 1.5 m/s. After the familiarization session, participants were instructed to complete two 8-minute walking bouts under CI and RLFLE (Session 1, S1). The goal of S1 was to validate and characterize the ability of the WBS to autonomously determine each subject's maximum steady-state training speed ($V_{\max}$). For this reason, the target walking speed in S1 was set to a very high value ($V_t = 2.5$ m/s) which none of the participants could possibly reach. This enabled a broad range of admissible values for V_g. Based on the collected data, $V_{\max}$ was computed as

$$V_{\max} = \underset{v}{\operatorname{argmax}}\, f_{num}(v),\ v \in [V_b, V_t], \tag{15}$$

where f_{num} is a user-defined function that returns the number of strides within a walking bout for which the subject's stride velocity V_s was 'sufficiently close' to a given value v while the subject approximately followed the current guided velocity $V_{g,k}$, as determined by the following two conditions:

$$|V_s - v| < 0.1\, m/s \ \text{ and } \ |V_s - V_{g,k}| < 0.05\, m/s \tag{16}$$

It is worth noting that V_s in (16) indicates the subject's current walking speed, which in general differs from the average $\bar{V}_{s,k}$ discussed in Sec. III-B.

After S1, participants rested for approximately ten minutes. Then, they were instructed to complete three 11-minute walking bouts, each under a different stimulation method (Session 2, S2). During S2, the target walking speed V_t was updated as a function of the elapsed session time t:

$$V_t(t) = \begin{cases} V_b & 0 < t \leq 100s \text{ (STARTUP)} \\ 0.5\,(V_{\max} + V_b) & 100s < t \leq 280s \text{ (LOW1)} \\ V_{\max} & 280s < t \leq 460s \text{ (HIGH)} \\ 0.5\,(V_{\max} + V_b) & t > 460s \text{ (LOW2)} \end{cases} \tag{17}$$

For each subject, the value of $V_{\max}$ used in (17) was set to the largest of the two estimates (CI vs. RLFLE) computed after S1 using (15). S2 was included in the protocol to determine the performance of the RLFLE method in relation to CS and CI during a low-high-low intensity gait training session.

The sequence of the stimulation methods in all sessions (familiarization session, S1, S2) was randomized using block randomization [36]. Participants were instructed to adjust their gait to the stimuli but were blinded to the purpose of the stimuli and to the type of stimulation method.

B. Data Analysis

For S1, the group averages of $V_{\max}$ and $f_{num}(V_{\max})$ were selected as the outcome metrics. For S2, data analysis was separated into three sub-sessions as indicated in (17): the first low-intensity training session (LOW1), the high-intensity training session (HIGH), and the second low-intensity training session (LOW2). The maximum percentage overshoot (OS), and the rising time (RT) were selected as the transient-response outcome metrics for each sub-session. The mean absolute velocity error (MAE) computed with respect to the target speed V_t and the coefficients of variation (CV) of V_s were regarded as the steady-state outcome metrics. Because it took participants less than two minutes to adapt to the guiding speed V_g, we only included the last minute of each sub-session when computing MAE and CV from S2 data.

Paired Wilcoxon signed-rank tests were carried out to assess significant ($\alpha < 0.05$) differences between CI and RLFLE in determining participants' $V_{\max}$. One-sample Wilcoxon signed-rank tests were used to check whether increments in $V_{\max}$ were significant. We applied two-way repeated-measures ANOVA to identify significant effects of training intensity (LOW1, HIGH, and LOW2) and stimulation method (CS, CI, and RLFLE), as well as potential interactions among the two factors. When significant effects were identified, post-hoc comparisons using the Bonferroni-Holm correction were applied as appropriate. All statistical analysis was carried out in SPSS v28 (IBM Corporation, Armonk, NY, USA).

TABLE I
p-VALUE OF THE REPEATED-MEASURES ANOVA AND POST-HOC ANALYSES.

Metrics	Repeated-Measures ANOVA			Post-Hoc Analysis for SM		
	SM	TI	SM*TI	CS/CI	CS/RLFLE	CI/RLFLE
MAE	<0.001	ns	ns	<0.05	<0.001	<0.001
CV	<0.05	ns	ns	ns	<0.01	ns
OS	<0.05	ns	ns	ns	<0.05	ns
RT	<0.001	<0.01	ns	<0.001	<0.01	ns

SM: stimulation method; TI: training intensity; MAE: mean absolute velocity error; CV: coefficient of variation for velocity; OS: maximum percentage overshoot; RT: rising time; ns: not significant.

V. RESULTS

Participant's V_b ranged from 1.12 to 1.60 m/s (1.30±0.15 m/s, mean and standard deviation). Fig. 7(a-b) shows the trends of V_s, V_g, and V_t for a representative participant during S1, under the stimulation methods CI and RLFLE. Fig. 7(c-e) reports the trends of V_s, V_g, and V_t for the same study participant during S2, under the stimulation methods CS, CI, and RLFLE. Fig. 8 reports the group averages of MAE, CV, OS, and RT induced by the three stimulation methods during S2. Results of the two-way repeated-measures ANOVA and post-hoc analyses for stimulation method are reported in Table I.

During S1, the RLFLE induced a significantly larger $V_{\max}$ in the study participants, compared to CI. The group averages of $V_{\max}$ for CI and RLFLE were 1.73±0.23 m/s and 1.79±0.23 m/s, respectively, with an average increment of 6.0 cm/s ($p < 0.05$). Moreover, participants were able to maintain $V_{\max}$ significantly longer when training with the RLFLE. The group averages of $f_{num}(V_{\max})$ for CI and RLFLE were 59±15 strides and 71±18 strides, respectively, reflecting an average increment of 12 strides ($p < 0.05$). Thus, the estimate of $V_{\max}$ obtained with RLFLE reflected more closely the wearer's steady-state maximum overground walking speed.

During S2, the RLFLE resulted in significantly smaller MAE than both CS and CI methods. RLFLE also induced smaller CV and OS than the other methods, however differences were significant only between RLFLE and CS. The CS method yielded a significantly smaller MAE than the CI method, however no significant differences were found between CS and CI in terms of CV or OS. RT was significantly shorter for CS compared to both CI and RFLE, however differences between CI a RFLE were not significant. The analysis did not evidence any interactions between training intensity and stimulation method. Interestingly, RT was the only outcome metric affected by the training intensity. Post-hoc analyses revealed significant differences between LOW1 and LOW2 ($p < 0.05$), and between HIGH and LOW2 ($p < 0.05$), but no significant difference was found between LOW1 and HIGH. This suggests that participants responded more abruptly when V_g decreased compared to when V_g increased.

VI. DISCUSSION

This letter proposed a novel vibrotactile stimulation method for personalized walking speed training. To the best of the authors' knowledge, this is the first study applying the RL framework to personalize the biofeedback strategy of a WBS for gait exercises. Compared to recent work on feedback personalization for WBS [13], the proposed RLFLE modulates the stimuli based on direct measures of an individual's gait speed, which is a widely used, clinically meaningful outcome for rehabilitation programs. An interesting feature of the RLFLE that sets it apart from previous research relies on its ability to self-determine a subject's $V_{\max}$ by progressively increasing V_g in a subject-specific and adaptive fashion, as indicated by the S1 data. Past studies on external stimulation for gait training have set the target exercise velocity V_t to an arbitrarily defined percentage of an individual's self-selected gait speed, ranging from 100% V_b to 125% V_b [13], [37], [38]. However, these values may not reflect a person's true $V_{\max}$. An automated way to establish $V_{\max}$ such as the one proposed here may aid the design of more targeted gait training protocols, as well as more direct assessments of key predictors of health status and gait/balance disturbances (e.g., the walking speed reserve [39]). Although the CI stimulation method can, in principle, capture a person's $V_{\max}$, it requires the definition of strict update rules, as exemplified in (14). Such hard-coded rules might be sub-optimal for certain users, thereby affecting the validity of the $V_{\max}$ estimates, as suggested by the worse results obtained in S1. Indeed, when comparing the estimates of $V_{\max}$ obtained with CI and RLFLE, the average difference was 6.0 cm/s, which is above the minimal clinically important difference (MCID) for older adults [40] and persons with Parkinson's disease [41]. Further research is warranted to determine whether the proposed RLFLE can be applied to assess $V_{\max}$ in older adults and clinical populations.

In past research, high- and low-intensity walking bouts have been traditionally analyzed as separate sessions. The self-adaptation of V_g allows the RLFLE to smoothly direct the wearer toward a time-varying target velocity V_t, making it possible to design more versatile, personalized training protocols. Compared with CS, not only did the RLFLE allow users to follow V_t more accurately and with less variability, but it also smoothed their transient response to the stimuli (i.e., smaller OS, larger RT). While we cannot rule out that the poorer performances of the CS method might have been affected, in part, by the participants being exposed to CS to a lesser extent than they were to CI and RLFLE (since CS was not used for S1), we believe this is highly unlikely. Indeed, all 3 stimulation methods share the same lower-level controller triggering the stimuli, and those stimuli are rather intuitive to follow, regardless of the stimulation modality. For this reason, participants were likely well accustomed to the plantar stimuli by the end of the familiarization session. Compared with CI, the RLFLE significantly improved the accuracy in the users' velocity adaptations, although changes in transient responses were non-significant. This result suggests that the rules governing how to increase V_g towards a higher V_t should be different from those associated with decreasing V_g to a lower V_t. Overall, the higher accuracy yielded by RLFLE compared to both CI and CS indicate that these non-adaptive stimulation methods may be less effective in eliciting a desired V_t in the wearer.

The life-long learning structure is a distinctive feature

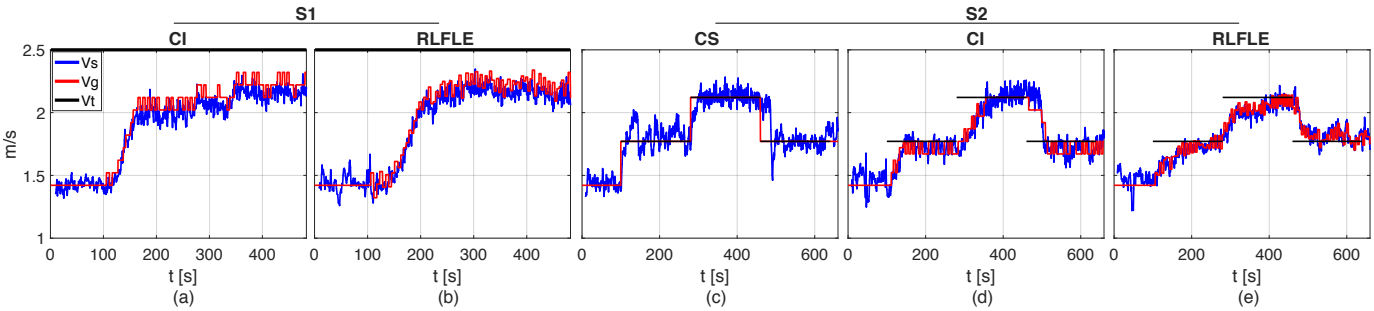

Fig. 7. Trends of V_s, V_g, and V_t for a representative participant under the three stimulation methods (CS, CI, and RLFLE) during S1 **(a-b)** and S2 **(c-e)**.

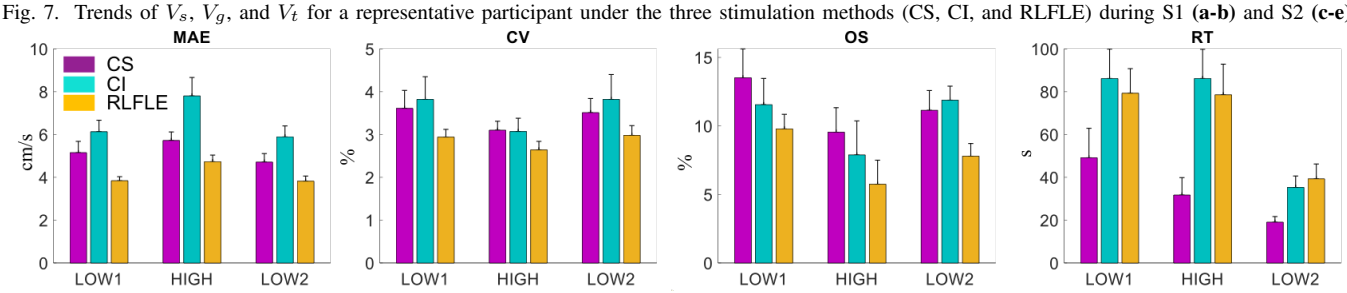

Fig. 8. Group averages of the mean absolute velocity error (MAE), coefficient of variation (CV), percentage overshoot (OS), and rising time (RT) induced by the three stimulation methods (CS, CI, and RLFLE) during S2. Error bars indicate +/-1SE.

differentiating RL from off-line learning methods such as Gaussian process regression [13]. RL methods based on the exact solution of Bellman's equation, such as policy iteration and value iteration methods [22], [42], traditionally require large amounts of training data, making them impractical for real-time applications. Instead, we approximated the long-term cost $\hat{Q}$ through parameterization, and used RLS to update these estimates. This approach takes full advantage of prior data and achieves reasonable performance without intensive computation. To simplify the RL modeling and training, only the average walking speed $\bar{V}_s$ was fed as input to the high-level controller. This leaves a rich set of WBS-measured kinematic and kinetic features unused. Furthermore, in the fuzzy inference strategy, the exploration was conducted by only relying on the acquired knowledge, without adding exploration noise. Hence, future work will include the investigation of additional input features for the high-level controller and the addition of exploration noise to excite the learning process. Another limitation of this study is the relatively small and homogeneous sample size. Future work will test the RLFLE with larger and more heterogeneous samples (e.g., older adults with gait and balance impairments). Additionally, this study only investigated the immediate benefits that occurred through gait training. A future study will need to assess the carry-over effects of training with the proposed stimulation method.

VII. Conclusion

This study provides proof-of-concept validation of a new RL-based strategy for wearable biofeedback devices used in walking exercises within an indoor environment. The proposed RLFLE proved more effective than two non-adaptive biofeedback strategies in modulating closed-loop underfoot vibrotactile stimuli to induce a desired gait speed in the wearer, resulting in lower walking speed errors. Additionally, results indicated that the RLFLE can autonomously determine an individual's maximum steady-state overground gait speed. Further research is warranted to determine the extent to which these findings translate to clinical populations with gait and balance impairments.

References

[1] A. Middleton, S. L. Fritz, and M. Lusardi, "Walking speed: The functional vital sign," *Journal of Aging and Physical Activity*, vol. 23, no. 2, pp. 314 – 322, 2015.

[2] I. L. Kyrdalen, P. Thingstad, L. Sandvik, and H. Ormstad, "Associations between gait speed and well-known fall risk factors among community-dwelling older adults," *Physiotherapy research international*, vol. 24, no. 1, p. e1743, 2019.

[3] A. Marengoni *et al.*, "Combining gait speed and recall memory to predict survival in late life: Population-based study," *Journal of the American Geriatrics Society*, vol. 65, no. 3, pp. 614–618, 2017.

[4] A. Nieuwboer *et al.*, "Cueing training in the home improves gait-related mobility in Parkinson's disease: the rescue trial," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 78, no. 2, pp. 134–140, 2007.

[5] Y. Baram and A. Miller, "Auditory feedback control for improvement of gait in patients with multiple sclerosis," *Journal of the neurological sciences*, vol. 254, no. 1-2, pp. 90–94, 2007.

[6] A. Schmid *et al.*, "Improvements in speed-based gait classifications are meaningful," *Stroke*, vol. 38, no. 7, pp. 2096–2100, 2007.

[7] Y. Baram and R. Lenger, "Gait improvement in patients with cerebral palsy by visual and auditory feedback," *Neuromodulation: Technology at the Neural Interface*, vol. 15, no. 1, pp. 48–52, 2012.

[8] C. de Labra *et al.*, "Effects of physical exercise interventions in frail older adults: a systematic review of randomized controlled trials," *BMC geriatrics*, vol. 15, no. 1, pp. 1–16, 2015.

[9] L. R. Nascimento *et al.*, "Walking training with cueing of cadence improves walking speed and stride length after stroke more than walking training alone: a systematic review," *Journal of physiotherapy*, vol. 61, no. 1, pp. 10–15, 2015.

[10] J. R. Vaz *et al.*, "Auditory and visual external cues have different effects on spatial but similar effects on temporal measures of gait variability," *Frontiers in physiology*, vol. 11, p. 67, 2020.

[11] M. H. Thaut and M. Abiru, "Rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research," *Music perception*, vol. 27, no. 4, pp. 263–269, 2010.

[12] H. Zhang *et al.*, "Wearable biofeedback system to induce desired walking speed in overground gait training," *Sensors*, vol. 20, no. 14, p. 4002, 2020.

[13] T. L. Wu, A. Murphy, C. Chen, and D. Kulić, "Human-in-the-loop auditory cueing strategy for gait modification," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3521–3528, 2021.

[14] P. Arias and J. Cudeiro, "Effects of rhythmic sensory stimulation (auditory, visual) on gait in Parkinson's disease patients," *Experimental brain research*, vol. 186, no. 4, pp. 589–601, 2008.

[15] J. Spencer, S. L. Wolf, and T. M. Kesar, "Biofeedback for post-stroke gait retraining: A review of current evidence and future research directions in the context of emerging technologies," *Frontiers in Neurology*, vol. 12, p. 419, 2021.

[16] L. Lünenburger, G. Colombo, and R. Riener, "Biofeedback for robotic gait rehabilitation," *Journal of neuroengineering and rehabilitation*, vol. 4, no. 1, pp. 1–11, 2007.

[17] S. Janssen *et al.*, "Usability of three-dimensional augmented visual cues delivered by smart glasses on (freezing of) gait in Parkinson's disease," *Frontiers in neurology*, vol. 8, p. 279, 2017.

[18] K. N. Winfree *et al.*, "The effect of step-synchronized vibration on patients with Parkinson's disease: Case studies on subjects with freezing of gait or an implanted deep brain stimulator," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 5, pp. 806–811, 2013.

[19] K. Yasuda, Y. Hayashi, A. Tawara, and H. Iwata, "Development of a vibratory cueing system using an implicit method to increase walking speed in patients with stroke: A proof-of-concept study," *ROBOMECH Journal*, vol. 7, no. 1, pp. 1–8, 2020.

[20] S. Schütz-Bosbach and W. Prinz, "Perceptual resonance: action-induced modulation of perception," *Trends in cognitive sciences*, vol. 11, no. 8, pp. 349–355, 2007.

[21] C. Kennel *et al.*, "Auditory reafferences: the influence of real-time feedback on movement control," *Frontiers in Psychology*, vol. 6, p. 69, 2015.

[22] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[23] Y. Wen *et al.*, "Online reinforcement learning control for the personalization of a robotic knee prosthesis," *IEEE transactions on cybernetics*, vol. 50, no. 6, pp. 2346–2356, 2019.

[24] Y. Zhang, S. Li, K. J. Nolan, and D. Zanotto, "Reinforcement learning assist-as-needed control for robot assisted gait training," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, 2020, pp. 785–790.

[25] Y. Zhang *et al.*, "Shaping individualized impedance landscapes for gait training via reinforcement learning," *IEEE Transactions on Medical Robotics and Bionics*, 2021.

[26] D. Zanotto, L. Turchet, E. M. Boggs, and S. K. Agrawal, "SoleSound: Towards a novel portable system for audio-tactile underfoot feedback," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 2014, pp. 193–198.

[27] S. Minto *et al.*, "Validation of a footwear-based gait analysis system with action-related feedback," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 9, pp. 971–980, 2015.

[28] J. Montes *et al.*, "Gait assessment with SoleSound instrumented footwear in spinal muscular atrophy," *Muscle & nerve*, vol. 56, no. 2, pp. 230–236, 2017.

[29] H. Zhang, D. Zanotto, and S. K. Agrawal, "Estimating CoP trajectories and kinematic gait parameters in walking and running using instrumented insoles," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2159–2165, 2017.

[30] H. Zhang *et al.*, "Regression models for estimating kinematic gait parameters with instrumented footwear," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 2018, pp. 1169–1174.

[31] H. Zhang, Y. Guo, and D. Zanotto, "Accurate ambulatory gait analysis in walking and running using machine learning models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 191–202, 2019.

[32] H. Zhang *et al.*, "Transductive learning models for accurate ambulatory gait analysis in elderly residents of assisted living facilities," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, In press.

[33] T. Yan *et al.*, "An oscillator-based smooth real-time estimate of gait phase for wearable robotics," *Autonomous Robots*, vol. 41, no. 3, pp. 759–774, 2017.

[34] S. J. Bradtke and A. G. Barto, "Linear least-squares algorithms for temporal difference learning," *Machine learning*, vol. 22, no. 1, pp. 33–57, 1996.

[35] T. J. Ross, *Fuzzy logic with engineering applications*. John Wiley Sons, 2005.

[36] K. Suresh, "An overview of randomization techniques: an unbiased assessment of outcome in clinical research," *Journal of human reproductive sciences*, vol. 4, no. 1, p. 8, 2011.

[37] W. O. C. Lopez *et al.*, "Listenmee® and listenmee® smartphone application: synchronizing walking to rhythmic auditory cues to improve gait in Parkinson's disease," *Human movement science*, vol. 37, pp. 147–156, 2014.

[38] J. M. Hausdorff *et al.*, "Rhythmic auditory stimulation modulates gait variability in Parkinson's disease," *European Journal of Neuroscience*, vol. 26, no. 8, pp. 2369–2375, 2007.

[39] M. L. Callisaya *et al.*, "Cognitive status, fast walking speed and walking speed reserve—the Gait and Alzheimer Interactions Tracking (GAIT) study," *GeroScience*, vol. 39, no. 2, pp. 231–239, Apr. 2017.

[40] G. Pulignano *et al.*, "Incremental value of gait speed in predicting prognosis of older adults with heart failure: insights from the image-hf study," *JACC: Heart Failure*, vol. 4, no. 4, pp. 289–298, 2016.

[41] C. J. Hass *et al.*, "Defining the clinically meaningful difference in gait speed in persons with parkinson disease," *Journal of Neurologic Physical Therapy*, vol. 38, no. 4, pp. 233–238, 2014.

[42] G. Bingjing, H. Jianhai, L. Xiangpan, and Y. Lin, "Human–robot interactive control based on reinforcement learning for gait rehabilitation training robot," *International Journal of Advanced Robotic Systems*, vol. 16, no. 2, p. 1729881419839584, 2019.

APPENDIX

The table below summarizes the fuzzy rule base used to combine the current tracking error $e_{c,k}$ (columns) and long-term feedback $e_{l,k}$ (rows) to determine appropriate fuzzy outputs $y_{i,j}$ (table entries) from which the update of the guided speed $\Delta V_{g,k}$ is computed according to (12). To obtain fuzzified inputs, the ranges of e_c and e_l are first split into $N = 11$ segments labelled as follows: N and P represent 'Negative' and 'Positive'; LL, L, M, and S indicate 'Very Large', 'Large', 'Medium', and 'Small'; ZE represent the 'Zero-Error' region. Each segment is then assigned a membership function as shown in Fig. 4(a-b). The maximum and minimum values of $y_{i,j}$ are set to ± 0.4m/s based on preliminary tests. When e_c and e_l have the same sign, V_g must be incremented in the same direction. Conversely, if e_c and e_l have opposite signs, y_{ij} is determined based on the following rules:

i) If $e_c > 0$ and $e_l < 0$, y_{ij} must be positive to reduce the tracking error.

ii) If $e_c < 0$ and $e_l > 0$, y_{ij} must be negative to reduce the tracking error.

iii) If $e_c \approx 0$, y_{ij} is small, and its sign is determined from e_l.

ΔV_g / e_l \ e_c		***LLN***	***LN***	***MN***	***SN***	***ZEN***	***ZE***	***ZEP***	***SP***	***MP***	***LP***	***LLP***
		-0.25	**-0.2**	**-0.15**	**-0.1**	**-0.05**	**0**	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**
LLN	**-0.5**	-0.4	-0.3	-0.2	-0.1	-0.1	-0.1	-0.1	0.1	0.2	0.3	0.4
LN	**-0.4**	-0.35	-0.25	-0.15	-0.1	-0.1	-0.1	-0.1	0.1	0.15	0.25	0.35
MN	**-0.3**	-0.3	-0.2	-0.15	-0.1	-0.1	-0.1	-0.1	0.1	0.15	0.2	0.3
SN	**-0.2**	-0.25	-0.2	-0.15	-0.1	-0.1	-0.1	-0.1	0.1	0.15	0.2	0.25
ZEN	**-0.1**	-0.25	-0.2	-0.15	-0.1	-0.1	-0.1	-0.1	0.1	0.15	0.2	0.25
ZE	**0**	-0.2	-0.2	-0.15	-0.1	-0.1	0	0.1	0.1	0.15	0.2	0.2
ZEP	**0.1**	-0.25	-0.2	-0.15	-0.1	0.1	0.1	0.1	0.1	0.15	0.2	0.25
SP	**0.2**	-0.25	-0.2	-0.15	-0.1	0.1	0.1	0.1	0.1	0.15	0.2	0.25
MP	**0.3**	-0.3	-0.2	-0.15	-0.1	0.1	0.1	0.1	0.1	0.15	0.2	0.3
LP	**0.4**	-0.35	-0.25	-0.15	-0.1	0.1	0.1	0.1	0.1	0.15	0.25	0.35
LLP	**0.5**	-0.4	-0.3	-0.2	-0.1	0.1	0.1	0.1	0.1	0.2	0.3	0.4