TYPE Original Research
PUBLISHED 23 January 2023
DOI 10.3389/fnbot.2022.1054239

Multimodal bipedal locomotion generation with passive dynamics *via* deep reinforcement learning

Shunsuke Koseki*, Kyo Kutsuzawa, Dai Owaki and Mitsuhiro Hayashibe

Neuro-Robotics Lab, Department of Robotics, Graduate School of Engineering, Tohoku University, Sendai, Japan

OPEN ACCESS

EDITED BY
Yaguang Zhu,
Chang'an University, China

REVIEWED BY
Ravi Shekhar Tiwari,
Chadura Tech, India
Raghu Sesha Iyengar,
Indian Institute of Technology Hyderabad, India
Azhar Aulia Saputra,
Tokyo Metropolitan University, Japan

*CORRESPONDENCE
Shunsuke Koseki
✉ shunsuke.koseki.q4@dc.tohoku.ac.jp

RECEIVED 26 September 2022
ACCEPTED 31 December 2022
PUBLISHED 23 January 2023

CITATION
Koseki S, Kutsuzawa K, Owaki D and Hayashibe M (2023) Multimodal bipedal locomotion generation with passive dynamics *via* deep reinforcement learning.
Front. Neurorobot. 16:1054239.
doi: 10.3389/fnbot.2022.1054239

Generating multimodal locomotion in underactuated bipedal robots requires control solutions that can facilitate motion patterns for drastically different dynamical modes, which is an extremely challenging problem in locomotion-learning tasks. Also, in such multimodal locomotion, utilizing body morphology is important because it leads to energy-efficient locomotion. This study provides a framework that reproduces multimodal bipedal locomotion using passive dynamics through deep reinforcement learning (DRL). An underactuated bipedal model was developed based on a passive walker, and a controller was designed using DRL. By carefully planning the weight parameter settings of the DRL reward function during the learning process based on a curriculum learning method, the bipedal model successfully learned to walk, run, and perform gait transitions by adjusting only one command input. These results indicate that DRL can be applied to generate various gaits with the effective use of passive dynamics.

KEYWORDS

bipedal walking and running, gait transition, deep reinforcement learning, underactuated robot, embodiment

1. Introduction

Humans exhibit multimodal gait patterns such as walking, running, skipping, and jumping (Diedrich and Warren, 1995; Alexander, 1996; Minetti and Alexander, 1997). Moreover, the gait transition between walking and running is observed at the speed boundaries, known as the preferred transition speed (PTS) (Sharbafi and Seyfarth, 2017). With the PTS as the boundary, walking at low speeds and running at high speeds are gaits with optimal energy efficiency in each speed domain (Diedrich and Warren, 1995; Alexander, 1996; Minetti and Alexander, 1997; Srinivasan and Ruina, 2006; Sharbafi and Seyfarth, 2017). To achieve such situation-dependent multimodal bipedal locomotion, not only neural control systems but also body morphology plays a crucial role based on the concept of *embodiment* (Pfeifer and Scheier, 2001; Owaki et al., 2008). Reproducing such human multimodal behaviors in a robot can broaden its locomotion ability as well as improve the understanding of the underlying mechanisms of human gaits and their transitions.

Despite the overwhelming complexity of their inherent dynamics, human locomotion can be represented as a simple conceptual model (Sharbafi and Seyfarth, 2017). Specifically, human walking can be represented as an inverted pendulum (IP) model (Kuo, 2007). During walking, the stance leg behaves as an inverted pendulum that rotates around the ankle joint. Additionally, the changes in the kinematic and potential energies are out-of-phase (i.e., when one is at its positive peak, the other is at its negative peak), whereas the mechanical energy remains almost constant (Cavagna et al., 1976, 2000; Cavagna and Legramandi, 2020). In contrast, the body dynamics of human running are different in that the changes in the kinematic and potential energies are in-phase (i.e., they reach their corresponding phases at the same time)

(Cavagna et al., 1976; Cavagna, 2006). When both the kinematic and potential energies decrease, some amount of energy is stored as elastic energy in the spring-like elements of the body, such as muscles, tendons, and ligaments (Farley and Gonzalez, 1996). Inspired by such biomechanical processes, the spring-loaded inverted pendulum (SLIP) model (Blickhan, 1989; Dickinson et al., 2000) has been adopted to explain and analyze running.

Passive dynamic walking is also based on the inverted pendulum mechanism (McGeer, 1990). In passive walking, a bipedal machine can walk down a gentle slope stably by using only its body dynamics without any actuators. The behavior is purely generated through the interaction between its body and the environment, and the control system is not involved in it. By generating motions exploiting its body morphology, a certain amount of computation for generating the behaviors can be *offloaded* to the body (Owaki et al., 2008; Pfeifer and Gómez, 2009). This reduces the computational cost to the control system and leads to energy-efficient locomotion (Collins and Ruina, 2005; Pfeifer and Gómez, 2009; Bhounsule et al., 2012). Thus, to realize these benefits, it is necessary to use passive dynamics and exploit the body morphology to generate movements.

However, despite the importance of passive dynamics in bipedal locomotion control, it has not been sufficiently investigated in previous bipedal walking and running robots (Hodgins, 1991; Kwon and Park, 2003; Nagasaka et al., 2004; Tajima et al., 2009; Sreenath et al., 2013; Kobayashi et al., 2016; Siekmann et al., 2021). There are two possible reasons for this: (1) Robots with passive joints are more difficult to control than fully actuated robots because there is less scope for the control system to intervene in motion generation. (2) As previously mentioned, walking and running are dynamically different locomotion modes; hence, it is difficult to reproduce multimodal locomotion from a single controller (Smit-Anseeuw et al., 2017; Okajima et al., 2018). Because body dynamics is partially determined by passive joints, generating dynamically different locomotion modes using passive dynamics is a challenging problem.

To address this issue, this study utilized deep reinforcement learning (DRL). In recent years, DRL has attracted attention as a promising technique for generating gaits in robotic systems. The advantage of DRL is that it learns locomotion skills with minimal craftsmanship and does not require careful modeling of the robot dynamics (Haarnoja et al., 2018a; Hwangbo et al., 2019). Previous studies have demonstrated that DRL can acquire controllers for multimodal gait in legged robots. Siekmann et al. (2021) presented a reward specification framework and demonstrated multimodal gaits in bipedal robots, including walking, running, hopping, and their transitions without prior knowledge. Fu et al. (2021) proposed a method for generating the walking, trotting, and bouncing gaits, and achieved smooth gait transitions in a quadrupedal robot by using a single controller *via* a stage-wise distillation approach. Shao et al. (2021) used imitation learning with the guided phase generated by the central pattern generator on a quadruped robot, and demonstrated multiple gaits and smooth transitions. Moreover, DRL can facilitate the realization of controllers for challenging locomotion skills in bipedal robots. Xie et al. (2020) presented a general learning scheme for navigating stepping stones. The bipedal robots in the simulation environment succeeded in walking on terrains consisting of discrete foot placements without falling or stopping. Yu et al. (2018) demonstrated realistic and smooth walking and running through simulation even though they did not make use of prior knowledge for training. In addition, the controllers generated through DRL are robust to variations in system dynamics, such as sensory delays, uneven terrain, and blind conditions (Xie et al., 2018; Castillo et al., 2021; Kang and Lee, 2021; Li et al., 2021).

The purpose of this study is to generate multimodal gaits such as walking, running, and their transitions using passive dynamics through DRL. For this, a bipedal model based on a passive walker was developed using numerical simulation. Subsequently, a reward function and learning scheme was designed for the DRL. The trained controller could achieve walking, running, and their transitions by adjusting only one input command. This study makes two significant contributions to the state of the art. The first contribution is that it established DRL as a promising technique for generating multimodal gaits using passive dynamics. The second contribution is that it presents a learning framework for training a simple control policy for a bipedal robot to switch between walking and running based on only a speed parameter without reference motion.

2. Methods

2.1. Bipedal model

The bipedal model employed in this study is shown in Figure 1. The model parameters are presented in Table 1. The basic mechanical structure is based on the passive dynamic bipedal model from our previous studies (Owaki et al., 2008, 2011). Motion is constrained in the sagittal plane. Notably, the hip joints are passive, i.e., they do not have actuators. Thus, the model needs to indirectly control hip joints such that the controller effectively exploits its body dynamics to swing its legs and move forward. Each leg has a linear actuator with a maximum magnitude of force F that moves the thigh segment up and down along the leg axis (red dotted line in Figure 1B). These actuators can push off the ground to generate a propulsion force in the stance phase and lift the corresponding leg to generate ground clearance of the foot in the swing phase. The hip joints (blue dotted line in Figure 1B) are passive rotational joints with torsional springs (k_{hip}) and dampers (c_{hip}). The spring generates rotational forces to swing the leg forward and backward, and prevents it from opening too widely (Owaki et al., 2011). In addition, a wobbling mass mechanism (Yue and Mester, 2002; Nikooyan and Zadpoor, 2011) consisting of three linear springs (k_1, k_2, k_3), two dampers (c_1, c_2), and a mass (m_{tibia}) was employed, as shown in Figure 1B. This mechanism contributes to reducing the impact between the foot and ground during foot–ground contact. The model has arc-shaped rigid feet with radius $r = l/3$ (Hansen et al., 2004), where l is the leg length. Each foot in the model consists of 20 small spheres. The state variables of the model are the hip segment positions x and z, hip segment orientation θ, leg angles φ_j, displacements of the thigh segments h_j, contractions of legs d_j, and their time derivatives. Here, suffix j denotes the leg ($j = r$: right and $j = l$: left).

2.2. Deep reinforcement learning

In this study, we designed a controller that outputs actuator signals for the input model states and a speed command, as shown in Figure 2. This system diagram can be seen in a previous study (Saputra et al., 2020). The controller was trained through deep reinforcement learning (DRL). DRL learns an action that maximizes

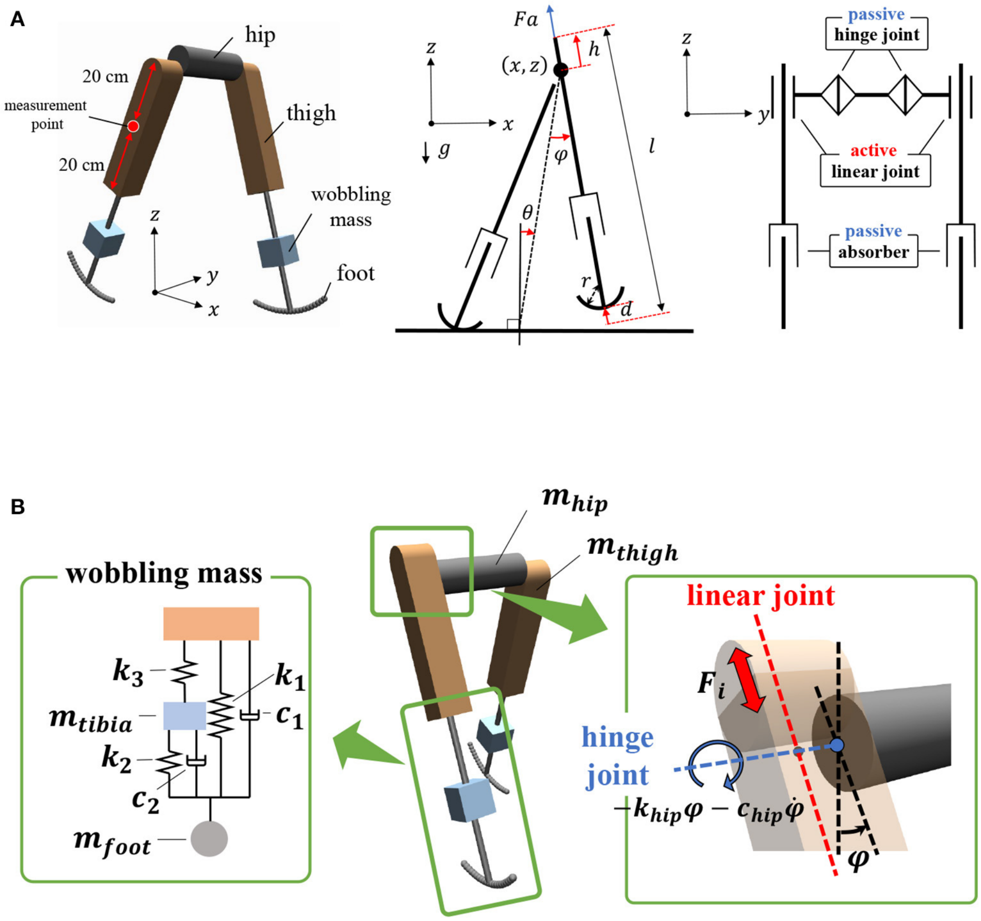


FIGURE 1
Structure of the bipedal model. The motions are constrained in the x–z plane, i.e., the sagittal plane. **(A)** Left: Oblique view of the bipedal model in MuJoCo simulator. Center: Side view of mechanical structure of the model. Right: Front view of mechanical structure. **(B)** Left: Schematic of wobbling mass mechanism. Right: Active linear actuators and passive rotational springs were implemented to thigh segments and hip joints, respectively.

TABLE 1 Parameters of the bipedal model.

Parameter	Unit	Value	Parameter	Unit	Value
l	[m]	0.8	k_{hip}	[Nm/rad]	25
r	[m]	0.27	c_{hip}	[Nms/rad]	2
m_{hip}	[kg]	20	k_1	[N/m]	6,000
m_{thigh}	[kg]	6	k_2	[N/m]	6,000
m_{tibia}	[kg]	3	k_3	[N/m]	10,000
m_{foot}	[kg]	1	c_1	[Ns/m]	300
			c_2	[Ns/m]	650
			F	[N]	600

the expected cumulative reward for the observed state through numerous trial-and-error iterations. This study adopted a soft actor-critic (SAC) (Haarnoja et al., 2018b), which is a model-free DRL algorithm for continuous control tasks, because it is the state-of-the-art technique and is better in terms of exploration. In this algorithm, a bonus reward $\alpha H(\pi)$ is added, where $H(\pi)$ is the entropy of the policy π. This term improves the exploration and provides robustness to policies (Haarnoja et al., 2018b). A stochastic policy π is obtained to maximize the objective function $J(\pi)$:

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}[\gamma^t(r(s_t,a_t) + \alpha H(\pi(\cdot|s_t)))], \tag{1}$$

where γ is the discount rate, r denotes the reward function (described in detail later), and α is the temperature parameter, which determines the emphasis of the entropy term. s_t and a_t denote the states and actions, respectively. In this study, $s_t \in \mathbb{R}^{14}$ comprises model states and a command parameter ω_v (details of which are provided in Section 2.3.2), and $a_t \in \mathbb{R}^2$ comprises the linear actuator signals $a_r \in [-1, 1]$ and $a_l \in [-1, 1]$. Each actuator produces a force a_jF (Figure 1A).

2.3. Learning methods

The objective is to realize a single controller that can achieve walking, running, and their transitions. In this section, we propose a learning framework for realizing a controller that generates multimodal gaits according to an input speed command ω_v, as shown in Figure 2, without any reference motions. We explain the design of the reward function and the learning scheme for DRL.

2.3.1. Design of reward function

In this study, the following reward function was used:

$$r(s_t, a_t) = \frac{1}{1+\omega_v}(-\omega_E|E_t - E_{t-1}| + \omega_v\dot{x} + f_{forward} + f_{alive} + f_{support}), \tag{2}$$

$$f_{forward} = \begin{cases} 0 & (\dot{x} \geq 0) \\ -C_1 & (\dot{x} < 0) \end{cases}, \tag{3}$$

$$f_{alive} = C_2, \tag{4}$$

Where ω_i represents the weight coefficients used to determine the relative significance of each term, E_t is the total energy, i.e., the sum of the potential, kinematic, and elastic energies of the model in t steps, $\dot{x}$ is the velocity of the hip segment in the horizontal direction, and C_1 and C_2 are constant values. The first term $-\omega_E|E_t - E_{t-1}|$ represents the penalty for the total energy variation. When the actuator injects excessive energy into the model, or when the energy lost at touchdown is significant, this term assigns a negative reward. The second term $\omega_v\dot{x}$ is the reward for the forward velocity of a bipedal robot. The term $f_{forward}$ represents a reward for maintaining the model in forward motion, where a large negative constant C_1 is added if the model does not move forward after a step. As presented later, the weight coefficient of the velocity term varied as $\omega_v \in [0, 2.5]$ during the training. We set $f_{forward}$ because when $\omega_v = 0$, there is no factor to determine the movement direction in the reward function. The term f_{alive} prevents the model from falling over. The episode is terminated if $|\theta| > 1.4$ rad, i.e., the model tilts by more than 80°, even if the maximum length of the step for an episode is not reached. Therefore, if the model does not fall over, this term will always provide a positive reward, and the agent can obtain a larger cumulative reward. The term $f_{support}$ enables more efficient learning by adding other rewards during the learning process as follows:

$$f_{support} = \omega_l f_{leg} + \omega_s f_{sym}, \tag{5}$$

$$f_{leg} = \min(|\dot{\varphi}_r - \dot{\varphi}_l|, C_3), \tag{6}$$

$$f_{sym} = -|\pi(s_t) - \Psi_a(\pi(\Psi_o(s_t)))|^2, \tag{7}$$

where the term f_{leg} enables the legs to swing effectively, and $\dot{\varphi}_r - \dot{\varphi}_l$ is the time derivative of the angle between the right and left legs (see in Figure 1A). C_3 is a constant with an upper limit to prevent an excessive outward swing of the legs. Because the hip joints are passive and cannot be driven directly, the agent had difficulty in finding the movement to swing the legs. f_{leg} encourages the policy to swing the legs and leads to efficient exploration. The term f_{sym} (Yu et al., 2018) introduces symmetry to the policy for movement generation to reproduce human-like symmetric movements during normal walking and running. Ψ_a and Ψ_o are functions that map the actions a_t and s_t, respectively, into their mirrored versions. As we will see later, the input for the policy s_t is described as $s_t = \{\dot{x}, \theta, \dot{\theta}, h_r, \dot{h}_r, h_l, \dot{h}_l, \varphi_r, \dot{\varphi}_r, \varphi_l, \dot{\varphi}_l, \dot{d}_r, \dot{d}_l, \omega_v\}$. Ψ_o maps s_t to the state s_t^{mirror}, where the right and left legs are interchanged. $s_t^{mirror} = \Psi_o(s_t) = \{\dot{x}, \theta, \dot{\theta}, h_l, \dot{h}_l, h_r, \dot{h}_r, \varphi_l, \dot{\varphi}_l, \varphi_r, \dot{\varphi}_r, \dot{d}_l, \dot{d}_r, \omega_v\}$. Ψ_a swaps the values of the right and left leg actuator outputs. $\Psi_a(a_t) = \Psi_a(\{a_{right}, a_{left}\}) = \{a_{left}, a_{right}\}$. f_{sym} penalizes the square deviation between the current state and mirrored state (for details, see Yu et al., 2018). $1/(1 + \omega_v)$ was set to reduce the inter-reward variability. The reward function calculates widely different values depending on the given variable $\omega_v \in [0, 2.5]$. Assuming that the bipedal model moved to 2.5 m/s, if $1/(1 + \omega_v)$ is not included, the agent is rewarded up to 1.6 for $\omega_v = 0$ and 7.85 for $\omega_v = 2.5$. If $1/(1 + \omega_v)$ is included, the agent is rewarded up to 1.6 for $\omega_v = 0$ and 2.25 for $\omega_v = 2.5$. Thus, we can reduce the variance among the rewards. The ranges of ω_v and constant values $\omega_E, \omega_l, \omega_s, C_1, C_2$, and C_3 were set as shown in Table 2 through trial and error.

TABLE 2 Parameters for the reward function.

	ω_v	ω_E	ω_l	ω_s	C_1	C_2	C_3
LP1	[0, 1.0]	0.06	0.2	0	1.0	1.0	0.5
LP2	[0, 1.0]	0.2	0.2	0.15	1.0	1.0	0.5
LP3	[0, 2.5]	0.2	0.2	0.15	1.0	1.0	0.5

2.3.2. Learning scheme

This study aimed to generate multimodal gait patterns by learning appropriate actuator outputs according to the input command ω_v values, which are the weight coefficients of the velocity term in the reward function in Equation (2). ω_v determines the relative significance of the forward velocity in the reward. For instance, assuming that the model moves in $\dot{x} = 2.0$ m/s again, the velocity term in the reward function, i.e., $\omega_v\dot{x}$, adds 5.0 to the reward when $\omega_v = 2.5$ and 0 when $\omega_v = 0$. Thus, a higher ω_v helps an agent learn high-speed locomotion. An agent was trained by changing the ω_v value per epoch, which was 1,000 time steps in this study. The bipedal model was set back to the initial state when an epoch ended or when the model fell, i.e., the model tilted forward by more than 80 degrees. We add ω_v to the state variables s_t: $s_t = \{\dot{x}, \theta, \dot{\theta}, h_r, \dot{h}_r, h_l, \dot{h}_l, \varphi_r, \dot{\varphi}_r, \varphi_l, \dot{\varphi}_l, \dot{d}_r, \dot{d}_l, \omega_v\} \in \mathbb{R}^{14}$.

Some hyperparameters for the reward function, including the range of ω_v, were changed depending on the learning phases inspired by curriculum learning (Brendan et al., 2020). Learning was divided into three phases: "Learning Phase 1 (LP1)," "Learning Phase 2 (LP2)," and "Learning Phase 3 (LP3)," where the agent was trained using different parameters, as shown in Table 2. In LP1, we set $\omega_E = 0.06$ and $\omega_s = 0$, and a relatively small range of $\omega_v \in [0, 1.0]$. In this phase, the agent was encouraged to learn a forward-motion movement by swinging its legs. In LP2, ω_E was increased to 0.2 and ω_s to 0.15 to penalize large variations in the total energy and nonsymmetric movement. In LP3, the range of $\omega_v \in [0, 2.5]$ was expanded, such that an agent can learn a wide range of velocities.

2.4. Simulation environment

MuJoCo (Todorov et al., 2012) was used as the physics simulation engine. MuJoCo provides a fast and accurate simulation environment. Thus, they are widely used in the fields of robotics

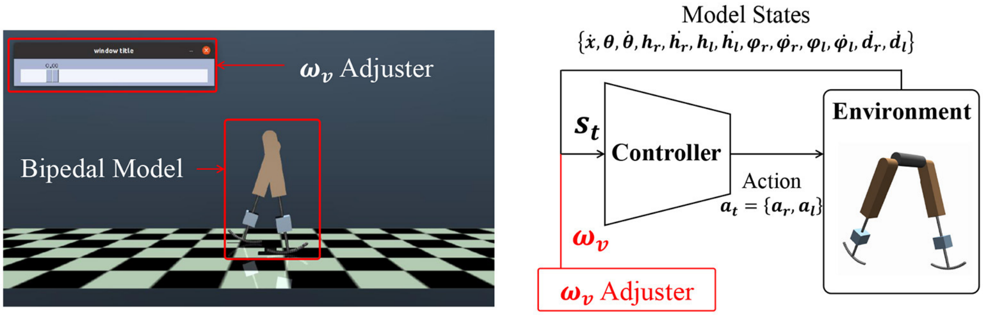
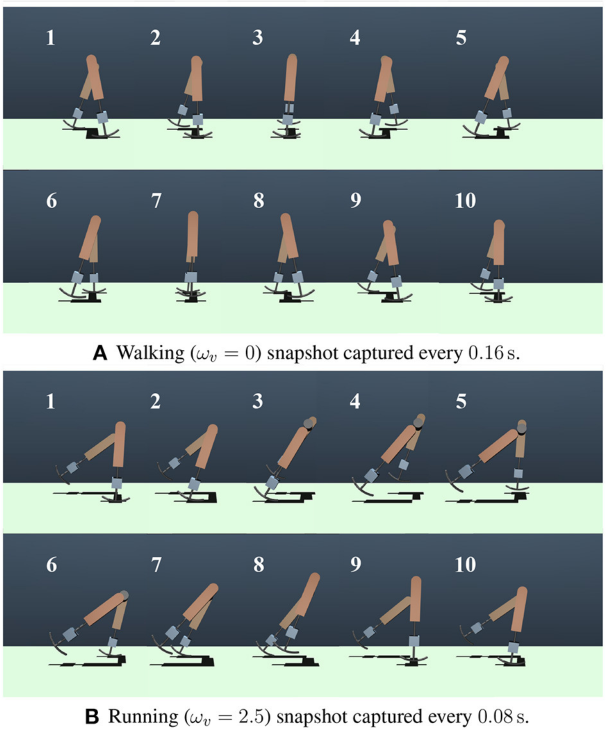

window title
ω_v Adjuster
Bipedal Model
Model States
$\{\dot{x}, \theta, \dot{\theta}, h_r, \dot{h}_r, h_l, \dot{h}_l, \varphi_r, \dot{\varphi}_r, \varphi_l, \dot{\varphi}_l, \dot{d}_r, \dot{d}_l\}$
s_t
Controller
Environment
Action
$a_t = \{a_r, a_l\}$
ω_v
ω_v Adjuster

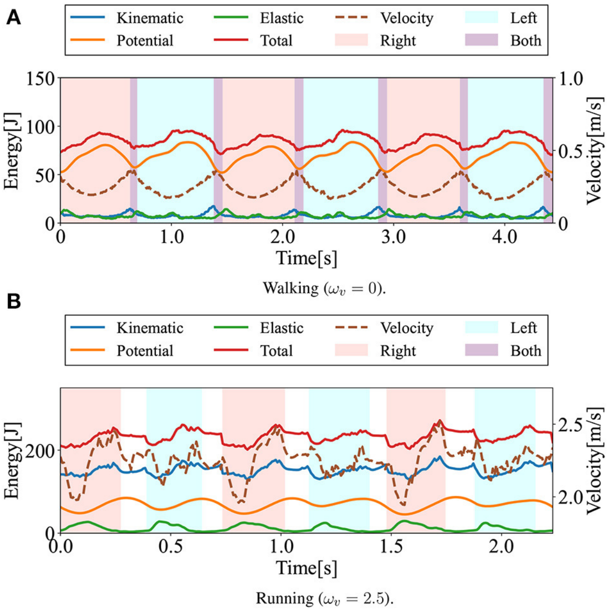
FIGURE 2
Bipedal model control in a simulation environment and the basic diagram of the control system. The inputs for the controller include a command ω_v. The value of ω_v can be changed through the ω_v adjuster. s_t consists of the model states $(\dot{x}, \theta, \dot{\theta}, h_r, \dot{h}_r, h_l, \dot{h}_l, \varphi_r, \dot{\varphi}_r, \varphi_l, \dot{\varphi}_l, \dot{d}_r, \dot{d}_l)$ and ω_v.



A Walking ($\omega_v = 0$) snapshot captured every 0.16 s.

B Running ($\omega_v = 2.5$) snapshot captured every 0.08 s.

FIGURE 3
Snapshots of the generated gaits. **(A)** Walking ($\omega_v = 0$) snapshot captured every 0.16 s. **(B)** Running ($\omega_v = 2.5$) snapshot captured every 0.08 s.



Walking ($\omega_v = 0$).

Running ($\omega_v = 2.5$).

FIGURE 4
Time evolution of energies during walking and running: kinetic (blue), potential (orange), elastic (green), and total (red) energies. The dotted lines represent locomotion velocity. The pink and sky blue colored areas represent the right and left single stance phase, respectively. The purple and white colored areas represent the double stance (DS) phase and the flight (F) phase. **(A)** Walking ($\omega_v = 0$). **(B)** Running ($\omega_v = 2.5$).

and biomechanics. The MuJoCo simulator can reproduce complex dynamic systems with many contact points.

3. Simulation results

The hyperparameters for the SAC were set to $\alpha = 0.2$ and $\gamma = 0.99$. The neural networks used in the actor and critic had two hidden layers of 100 nodes. We empirically selected a small number of hidden layers and nodes for efficient learning. The actor accepts the current state $s_t \in \mathbb{R}^{14}$ and outputs the actuator signals $a_t \in \mathbb{R}^2$. The critic accepts the current state $s_t \in \mathbb{R}^{14}$ and action $a_t \in \mathbb{R}^2$. Reinforcement learning was run for ten million steps; LP1 was performed in the initial 500 thousand time steps, LP2 in the next 2.5 million time steps, and LP3 in the remaining 7 million time steps (Table 2). At each time step, the policy was updated by using a replay buffer with the recent one million samples, with a mini-batch size of 256. The maximum length of each episode was set to 1,000. The time required for training was approximately 2 days on a Lenovo ThinkPad E470 20H2S04L00. The generated gait can be seen in the video available as Supplementary material.

3.1. Generated steady gaits

In this section, the gaits generated in the numerical simulations are presented. Figure 3 shows the snapshots of the generated steady

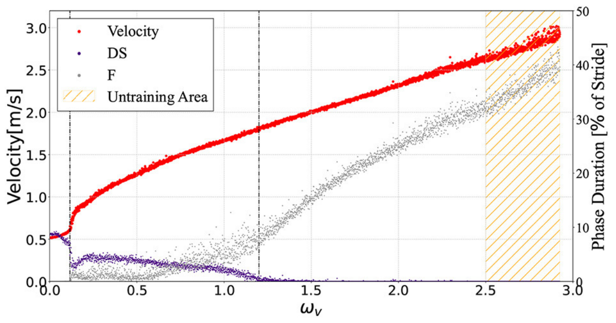


FIGURE 5
The average model velocities for the input ω_v. The purple and gray points denote the ratio of the DS and F phase in the gait cycle respectively. The controller generated steady gaits in the bipedal model for $0 \leq \omega_v \leq 2.93$, even though we did not train with $2.5 \leq \omega_v \leq 2.93$ (orange hatched area). This graph shows the plots when the input command ω_v was given in the range of 0.0–2.93 in 0.01 intervals. For each ω_v, the biped model locomoted for a distance of 30 m from the initial state.

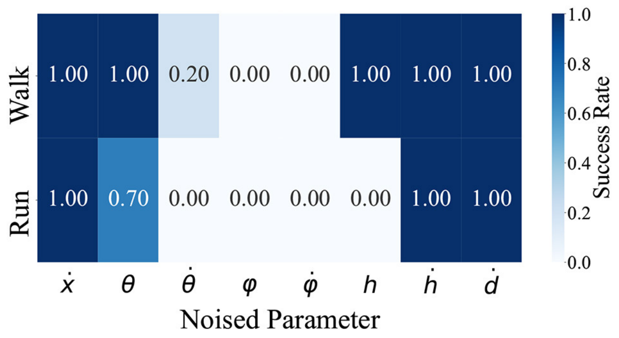


FIGURE 6
The rate of steady gait was maintained when noise was applied. $\dot{x}, \theta, \dot{\theta}, \varphi, \dot{\varphi}, h, \dot{h}, \dot{d}$ are the input parameters for the controller. For each of these seven states, we added Gaussian noise during steady walking and running and checked whether the bipedal model maintains the gaits. We conducted 10 locomotion trials for each condition and judged the success case when the bipedal model was able to move 10 m in the walking gait and 20 m in the running gait after noise was applied.

gaits for Figure 3A $\omega_v = 0$ and Figure 3B $\omega_v = 2.5$. Figure 4 shows the time evolution of the kinematic, potential, and elastic energies for both cases; the upper and lower graphs correspond to the cases in Figures 3A, B, respectively. Walking and running are defined as gaits that progress through a periodic double stance (DS) phase when both legs are on the ground, and through periodic flight (F) phases when both legs are in the air, respectively, Diedrich and Warren (1995) and Alexander (1996). The upper graph of Figure 4 indicates the periodic DS phases (purple areas), whereas the lower graph indicates the periodic F phases (white areas). The results indicate that the gait generated for $\omega_v = 0$ exhibited an out-of-phase relationship between the kinematic and potential energies, whereas the gait generated for $\omega_v = 2.5$ exhibited an in-phase relationship in the single stance phase. In human walking, the kinetic and potential energies exhibit an out-of-phase relationship, with potential energy being the maximum in the mid-stance phase and kinetic energy being the maximum in the DS phase (Cavagna et al., 1976, 2000; Cavagna and Legramandi, 2020). In human running, the kinetic and potential energies change in-phase during the single stance phase, both decreasing from initial contact to the mid-stance phase and increasing from the mid- to late-stance phase (Cavagna et al., 1976; Cavagna, 2006). Therefore, the phase relationship between the kinematic and potential energies in the generated gaits reflects the features of the human gaits. Furthermore, considering that the bipedal model weight (40 kg) is roughly half the weight of an adult, the magnitude of kinetic energy in both walking and running is in good agreement with the human measurement data (Cavagna, 2006; Cavagna and Legramandi, 2020).

Figure 5 illustrates the average velocities and the DS and F phase ratios of a single gait cycle in response to the value $\omega_v \in [0, 2.93]$. The controller generated a running gait for parameters $\omega_v \in [2.5, 2.93]$ outside the training range $\omega_v \in [0, 2.5]$. The speed range observed in this bipedal model was $0.51 \leq \dot{x} \leq 3.02$ m/s. A periodic DS phase was observed for $0 \leq \omega_v \leq 0.1$, while an F phase was not observed. In the range $\omega_v \simeq 0.1$, the ratio of the DS phase suddenly decreased; then, for $0.1 \leq \omega_v \leq 1.2$, transient gait patterns with nonperiodic DS and F phases were found. For $\omega_v \geq 1.2$, neither DS nor periodic F phases were found.

TABLE 3 Standard deviation of gaussian noise.

$\sigma_{walk\,\dot{x}}$	0.3476 [m/s]	$\sigma_{run\,\dot{x}}$	0.6743 [m/s]
$\sigma_{walk\,\theta}$	0.2757 [rad]	$\sigma_{run\,\theta}$	0.3053 [rad]
$\sigma_{walk\,\dot{\theta}}$	0.7986 [rad/s]	$\sigma_{run\,\dot{\theta}}$	1.0773 [rad/s]
$\sigma_{walk\,h}$	0.1942 [m]	$\sigma_{run\,h}$	0.3007 [m]
$\sigma_{walk\,\dot{h}}$	0.5596 [m/s]	$\sigma_{run\,\dot{h}}$	0.9818 [m/s]
$\sigma_{walk\,\varphi}$	0.4433 [rad]	$\sigma_{run\,\varphi}$	0.5628 [rad]
$\sigma_{walk\,\dot{\varphi}}$	0.9192 [rad/s]	$\sigma_{run\,\dot{\varphi}}$	1.664 [rad/s]
$\sigma_{walk\,\dot{d}}$	0.3110 [m/s]	$\sigma_{run\,\dot{d}}$	0.4349 [m/s]

Because the scales of the observation states were not consistent, the standard deviation values of the noise depended on the states. We measured the time evolution of each observation state during steady walking ($\omega_v = 0$) and running ($\omega_v = 2.5$) and calculated standard deviations. Then we set these values as the standard deviations of the Gaussian noise.

We identified the parameters essential for the steady walking and/or running gaits by adding noise to the model state inputs for the controller. As mentioned, the observed model states for the controller comprises the model velocity $\dot{x}$, hip segment orientation θ and its change rate $\dot{\theta}$, leg angles φ and their change rate $\dot{\varphi}$, displacements of the thigh segments h and their change rate $\dot{h}$, and leg contraction rates $\dot{d}$. For each of these seven observation states, we examined whether steady walking ($\omega_v = 0$) and running ($\omega_v = 2.5$) could be maintained when Gaussian noise with the mean 0 and the standard deviation σ was added. Since each observation state has a different scale, calculating the noise with the same standard deviation is not reasonable. Therefore, we measured the time evolution of each observation state during steady walking ($\omega_v = 0$) and running ($\omega_v = 2.5$) and calculated standard deviations, then set these values as the standard deviation of the noise; the detailed values are shown in Table 3. Figure 6 shows the noised parameters; the rates of gait were maintained (i.e., the model did not all down). Both generated walking and running are susceptible to noise at $\dot{\theta}$, φ, and $\dot{\varphi}$. Moreover, in the running gait, the bipedal model was vulnerable to noise at h and θ.

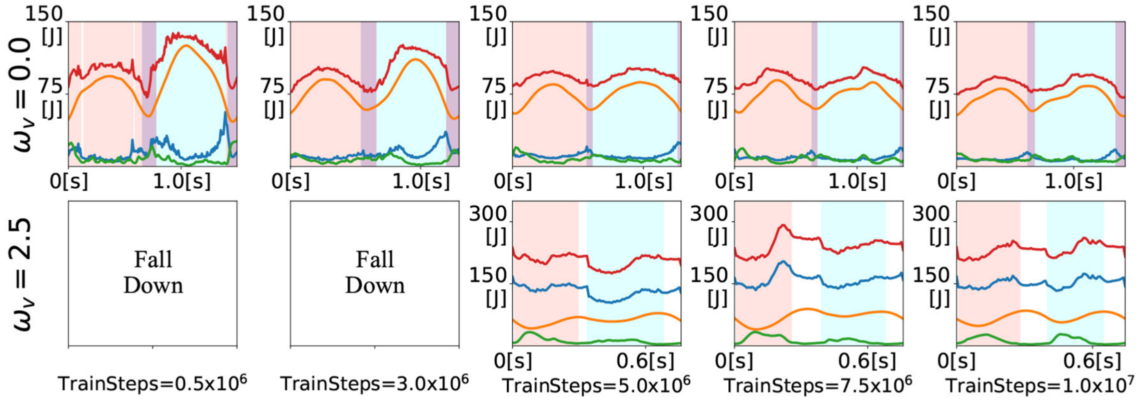


FIGURE 7
Time evolution of energies for the inputs $\omega_v = 0$ and $\omega_v = 2.5$ during training. The colored lines denote the kinematic (blue), potential (orange), elastic (green), and total (red) energies. The shaded areas denote the right single stance phase (pink), left single stance phase (sky blue), DS phase (purple), and F phase.

3.2. Generated gaits during training

In this section, we show the gaits generated during training for the inputs $\omega = 0$ and $\omega = 2.5$. Figure 7 shows the time evolution of the energies at 500 thousand (at the end of LP1), 3 million (at the end of LP2), 5 million, 7.5 million, and 10 million (at the end of training) training time steps. The upper figures show the generated gaits for $\omega = 0$ and the lower figures show the generated gaits for $\omega = 2.5$. For the input $\omega = 0$, the walking gait was already generated at the 500 thousand training time steps. Thereafter, the total energy variation reduced as the training progressed, and at the middle of the training (i.e., 5 million training time steps), the time evolution of the energies was almost the same as that at the end of the training. For the input $\omega = 2.5$, the bipedal model kept falling down until 3 million training time steps. The running gait appeared at the middle of the training. However, it was asymmetric, as evidenced by the relatively short DS phase (white area) after the right single stance phase (pink area). As training progressed, the duration of the DS phase became almost the same. In addition, elastic energy (green) began to be used during the left single stance phase (sky blue area). The gaits generated during training can be seen in the video available as Supplementary material.

3.3. Energetics

The speed range obtained in this bipedal model was $0.51 \le \dot{x} \le 3.02\,\mathrm{m/s}$, and the Froude number (Fr) was $0.18 \le Fr \le 1.08$, using $Fr = \frac{\dot{x}}{\sqrt{gl}}$. In this study, we investigated the energetics of the generated gaits, which can be divided into three gait patterns according to the measured locomotion velocity shown in Figure 5: walking gait for $0.51 \le \dot{x} \le 0.63\,\mathrm{m/s}$, transient gait for $0.63 \le \dot{x} \le 1.81\,\mathrm{m/s}$, and running gait for $\dot{x} \ge 1.81\,\mathrm{m/s}$. As mentioned before, walking and running gaits are gait patterns with periodic DS and F phases, respectively. Transient gait is defined as a gait pattern with both DS and F phases; however, these phases are not periodic. Here, we numerically evaluated the energy efficiency using the cost of transport (CoT) (Ruina et al., 2005), which is described by the following equation:

$$\mathrm{CoT} = \frac{\Delta W}{mg\Delta x} = \frac{1}{mg\Delta x}\sum_{j\in r,l}\int_{t_0}^{t_{end}} \max(F_{a,j}(t)\dot{h}_j(t), 0)dt, \qquad (8)$$

Where ΔW denotes the total energy consumption, m is the model mass, g is the acceleration due to gravity, Δx is the distance traveled, $F_{a,j}(t)$ is the force of each actuator, and $\dot{h}_j(t)$ denotes the displacement velocity of each thigh segment. Figure 8A shows the CoT profile as a function of the measured velocity. A lower CoT value indicates energy-efficient locomotion.

Moreover, we defined the variation of energy (VoE), which is the time integral of the total energy variation, as follows:

$$VoE = \frac{\Delta E}{mg\Delta x} = \frac{1}{mg\Delta x}\int_{t_0}^{t_{end}} |E_t - E_{t-1}|dt, \qquad (9)$$

Where ΔE is the sum of the total energy changes from t_0 to t_{end}. For a lower CoT (i.e., energy-efficient) gait, the VoE also tends to be lower because the energy injected from the actuators and the energy lost at touchdown are small and the total energy remains constant. Hence, there is a high correlation between the VoE and CoT. Figure 8B illustrates the relationship between the measured locomotion velocities and the VoE. A comparison of Figures 8A, B shows that the change in CoT against the measured velocity demonstrates a trend similar to that of VoE except for the range of $1.7 \le \dot{x} \le 1.9\,\mathrm{m/s}$, for which the CoT shows a downward convex trajectory. We can see that the trajectory of the VoE can be clearly divided for each of the three gait patterns. The VoE for the walking gait and transient gait can be approximated by different quadratic curves (red lines), as shown in Figure 8B. The VoE for the running gait is a linear line proportional to the velocity.

3.4. Adaptability to environmental changes

We investigated whether the trained controller could adapt to environmental changes. For this purpose, we set up a stepped environment, where steps of a constant height h appeared every 50 cm, as shown in Figure 9. We verified the steady walking ($\omega_v = 0$)

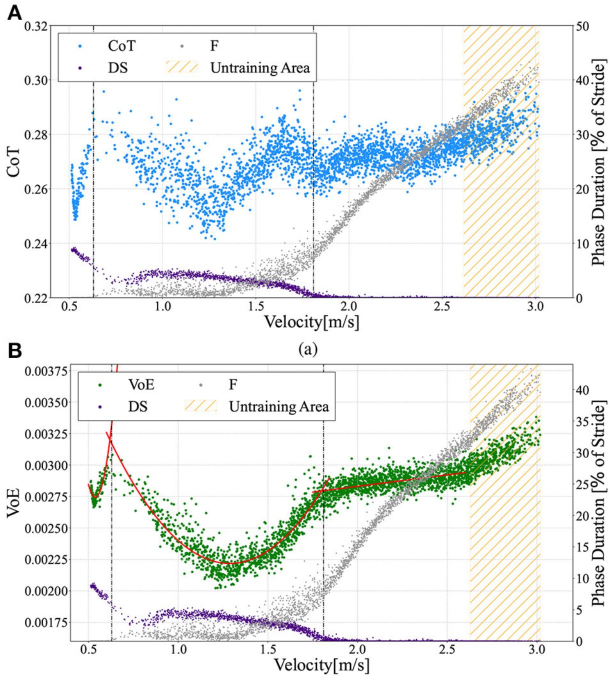


FIGURE 8
(A) Cost of Transport (CoT) vs. measured velocities $\dot{x}$. The scatter graphs here include data from a 10 to 30 m distance locomotion in the ω_v range from 0.0 to 2.93 in 0.01 intervals, similar to Figure 5. **(B)** Variation of Energy (VoE) vs. measured velocities $\dot{x}$. Gaits can be divided into three patterns: (1) walking $0.51 \leq \dot{x} \leq 0.63$ m/s, (2) transient gait $0.63 \leq \dot{x} \leq 1.81$ m/s, and (3) running gait $\dot{x} \geq 1.81$ m/s. The vertical dotted lines describe the borders at $\dot{x} = 0.63$ m/s (left) and $\dot{x} = 1.81$ m/s (right). The VoE for walking gait and transient gait can be approximated by quadratic curves (red lines) of $VoE_w = 0.0694\dot{x}^2 - 0.075\dot{x} + 0.0229$ and $VoE_t = 0.0022\dot{x}^2 - 0.0058\dot{x} + 0.0059$, respectively. For the running gait mode, a straight line represents the function $VoE_r = 0.00018\dot{x} + 0.00246$.

and steady running ($\omega_v = 2.5$) performance in this environment. We judged the success case as the condition in which the bipedal model successfully covered a distance of 30 m. The bipedal model can move without falling down up to $h = 1.0$ cm in the walking gait and $h = 0.8$ cm in the running gait. The attached movie shows the bipedal model walking and running in a changing environment.

3.5. Gait transitions

To evaluate the gait transition ability of the acquired policy, the gait pattern was observed under a change in the command value ω_v during locomotion. In this simulation, the input ω_v was linearly increased from 0.0 to 2.5 at 0.6 s in Figure 10A, whereas ω_v was linearly decreased from 2.5 to 0.0 at 0.6 s, as shown in Figure 10B. For both changes in ω_v, the acquired policy could successfully achieve gait transition. Note that the transition from walking to running was successfully completed in less than two steps, whereas the transition from running to walking required approximately four steps. Moreover, not only single transitions such as walk-to-run and run-to-walk, but also multiple transitions such as walk-to-run-to-walk were achieved.

Figure 11 illustrates the x–z CoM trajectories of the right thigh segment with respect to the hip joint position (Figure 1) for 12 s (before and after the ω_v change). As shown in Figure 11, ω_v changed linearly from 0 to 2.5 in Figure 11A and from 2.5 to 0 in Figure 11B in 0.6 s. In the steady-state motion, it can be observed that walking and running converge to different limit cycles, smaller ones to the right and larger ones to the left, respectively. This limit cycle analysis also confirmed that the transition from walking to running was achieved smoothly and stably, while that from running to walking required more time steps.

Figure 12 shows the time evolution of the hip segment orientation θ (see Figure 1) during gait transitions. The gait patterns generated by the learned policy show that the bipedal model maintained a relatively perpendicular posture to the ground ($\theta \approx 0$) when walking, whereas the model moved with a forward leaning posture ($\theta > 0$) when running. This can also be observed in the snapshots of the generated gaits (Figure 3). The difference between Figures 3A, B in the gait transition duration may be due to the difference in body posture during walking and running: the transition from upright walking to leaning forward running is physically easy, whereas the transition process from leaning forward to upright walking requires the posture to return to upright, which requires more energy; thus, the physical state transition takes more time steps.

3.6. Comparison with partially modified reward function and learning scheme

To investigate the factors contributing to effective multimodal locomotion learning, the agents were trained with partially modified reward functions. Figure 13 shows the representative learning curves for Figure 13A our proposed method; Figure 13B excluding $1(1+\omega_v)$ from Equation (2) not to reduce inter-reward variability; Figure 13C the range of ω_v in LP1 and LP2 was set to [0,2.5]; Figure 13D the range of ω_v in LP1 and LP2 was set to [1.5,2.5]; Figure 13E trained without LP2; and Figure 13F trained without LP1. In these figures, cumulative rewards close to 0 imply that the bipedal model did not move forward or immediately fell. Note that there was a significant amount of variance in the cumulative rewards even after the completion of training. This was because the reward function calculated widely different values depending on the given $\omega_v \in [0, 2.5]$, which changed randomly during the training.

Figure 13B shows that the cumulative rewards gradually dropped to 0 after training for 6 million time steps. It was observed that for relatively small ω_v, the bipedal model exhibited steady locomotion by the 6-million-time-step training; however, it became unstable around 7 million time steps and fell immediately after 8 million time steps. Note that the results in Figure 13B were obtained by excluding $1/(1+\omega_v)$ from the reward function; hence the range of cumulative rewards is different from that in the other subfigures.

In Figure 13C, ω_v was fixed to $[0, 2.5]$ through training, and the agent acquired only walking gaits and fell after a few steps for a relatively large ω_v. In Figure 13D, the value of ω_v in LP1 and LP2 was increased; the agent obtained only running gait; they could not find walking gait and fell immediately for a relatively small ω_v. In Figure 13E, it can be observed that training the agent without LP2 took more time steps to learn the appropriate outputs for ω_v

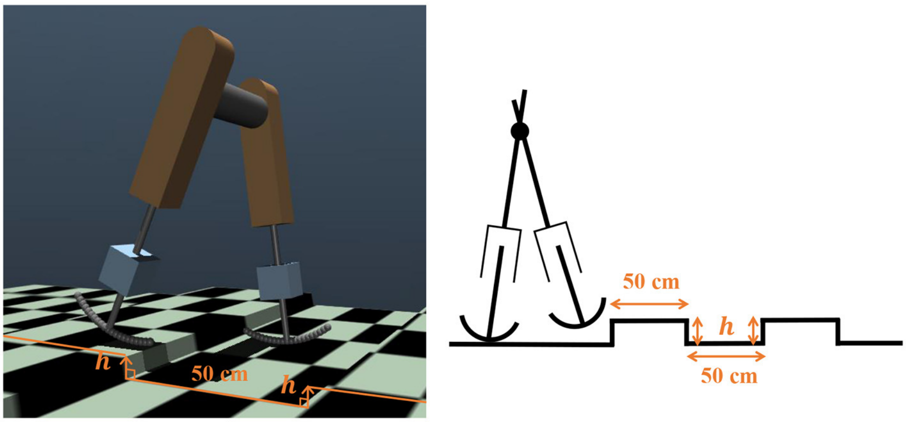


FIGURE 9
Environmental changes introduced using steps in a simulation environment and its diagram.

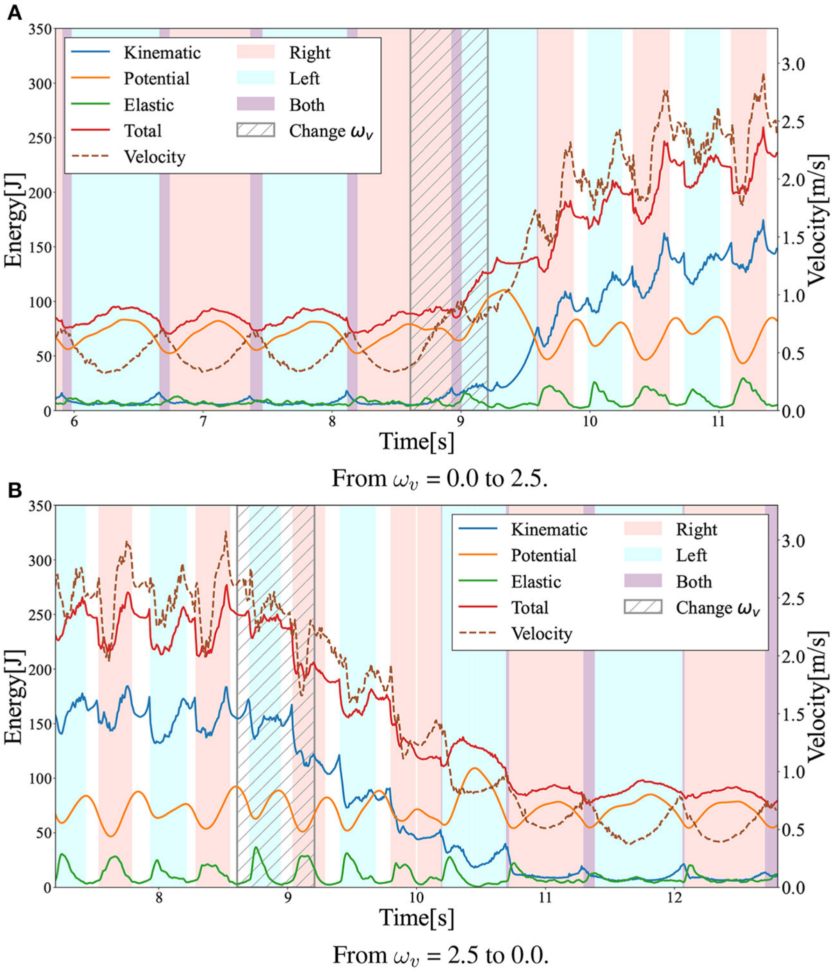


FIGURE 10
Gait transition between walking and running with changes in the ω_v value. The graph shows the time evolution of kinetic (blue), potential (orange), elastic (green), and total (red) energies during the gait transition. The dotted lines represent locomotion velocity. The pink and sky blue areas represent the right and left single stance phase, respectively. The purple and white areas represent the double stance (DS) and flight (F) phases. We linearly changed the ω_v in the shaded period. **(A)** From ω_v = 0.0 to 2.5. **(B)** From ω_v = 2.5 to 0.0.

compared to Figure 13A. From Figure 13F, the agent trained without LP1 failed to find control solutions for steady locomotion; it fell from the start position. A comparison of these results suggests that LP1 is an essential phase for learning the basic gait pattern from an unlearned state, while LP2 is necessary for learning a more stable gait pattern from the initially acquired gait pattern.

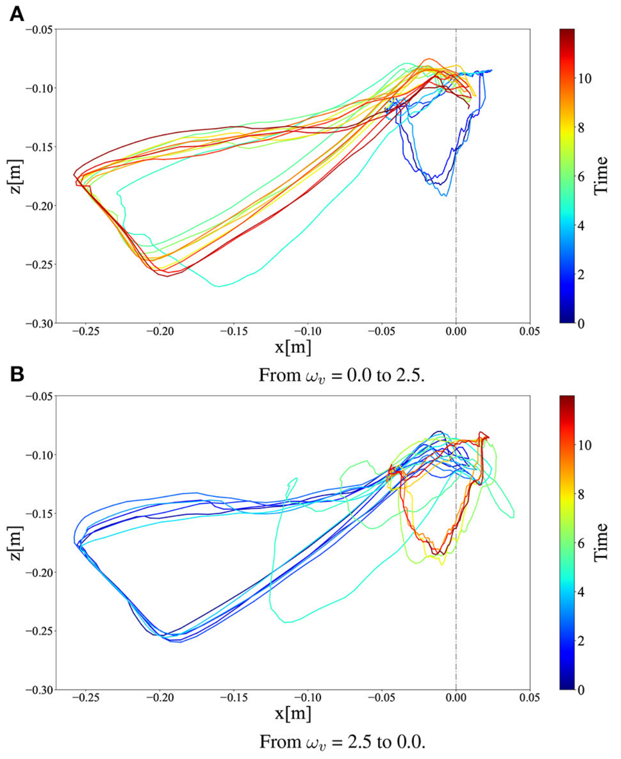


FIGURE 11
x–z CoM trajectories of the right thigh segment in **(A)** walk-to-run transition and **(B)** run-to-walk transition for 12 s (before and after the ω_v change). In the figures, we linearly changed ω_v at 4 s.

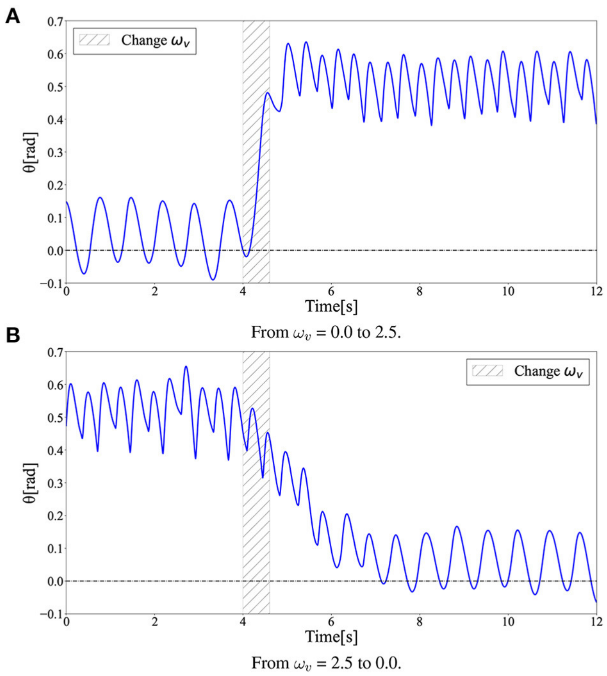


FIGURE 12
Hip segment orientation (θ) during gait transition. The gait patterns generated by the learned policy show that the bipedal model maintained a relatively perpendicular posture to the ground ($\theta \approx 0$) when walking, while the model moved with a forward leaning posture ($\theta > 0$) when running. **(A)** From ω_v = 0.0 to 2.5. **(B)** From ω_v = 2.5 to 0.0.

4. Discussion

This study demonstrated that a trained controller can generate walking, running, and transient gaits in a bipedal model using passive dynamics. We also found that the trained controller was adaptable to environmental changes during steady walking and running. Moreover, the bipedal model with the trained controller exhibited gait transitions by changing a single parameter ω_v. These results suggest that DRL can be applied to generate multimodal bipedal locomotion using passive dynamics. Additionally, the energy efficiency of the locomotion generated by the policy acquired through DRL was verified, including the reliability of learning using curriculum learning (Brendan et al., 2020), i.e., the parameter planning for the settings of the weights for the rewards during the learning process.

Here, Equation (2) was used as the reward function. The first term $-\omega_1|E_t - E_{t-1}|$ significantly contributed to the generation of continuous locomotion in the bipedal model. Before this term was included, no control solutions were found for walking and running, and the bipedal model exhibited acyclic, velocity-unstable, and fall-prone movements. Humans generate walking and running motions with small energy fluctuations by effectively using their body dynamics. During walking, humans behave like an inverted pendulum and exchange kinematic and potential energy to conserve mechanical energy. During running, humans store potential and kinematic energy in the spring elements of their bodies to reduce energy loss, resulting in less energy fluctuations. From Figure 4, it can be observed that the presented bipedal model employed the same strategy for the body dynamics observed in humans. An out-of-phase relationship was observed between kinetic and potential energy in the walking gait. In the running gait, during the initial single stance phases, elastic energy increased sharply, whereas both the kinematic and potential energies fell. Subsequently, the elastic energy gradually decreased with kinematic and potential energy increments. We consider that the first term in the reward function contributes to achieving walking and running in the bipedal model because this term helps the model mimic the energy variation in time, similar to how humans generate walking and running motions.

The constant parameters in the reward function, ω_E, ω_l, and C_3, need to be set carefully. When ω_E, which determines the penalty for the total energy variation, is set to a small value, the generated gaits appear awkward, as shown in the time evolution of energies for five thousand samples in Figure 7. Additionally, when ω_E is set to a large value from the beginning, no gait is generated, as shown in Figure 13F, because of the excessive penalty for the movement. We consider the approach of acquiring the basic gait pattern with a small penalty for energy in the initial phase, and acquiring the learning movements with a large penalty for energy as the training progresses, as effective in generating efficient gaits. When either ω_l or C_3 was set to a very small value, the agent did not tend to acquire a steady gait. However, when either ω_l or C_3 was set to a very large value, the trained controller exhibited a high kicking gait. f_{leg} ensures efficient learning but needs to be set carefully.

Figure 6 shows that although the generated walking was robust to the noise applied to the observed displacements of the thigh

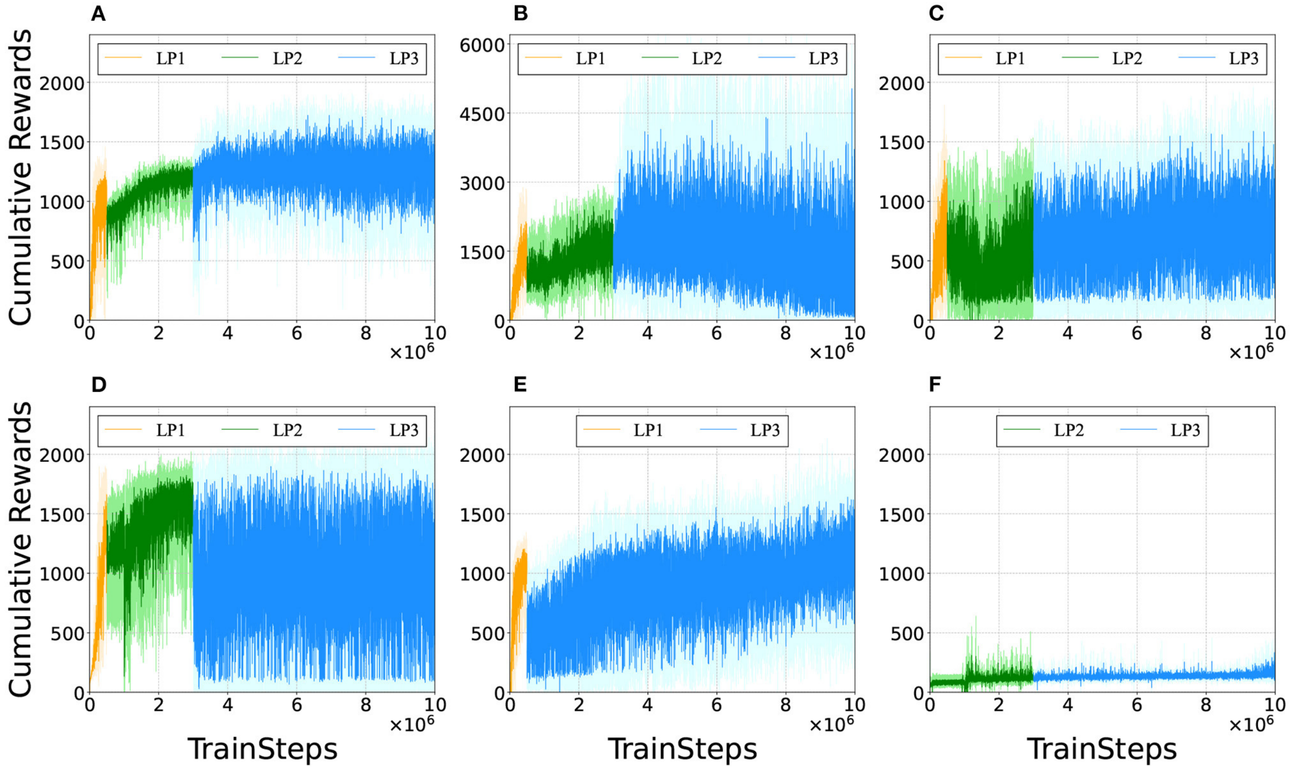


FIGURE 13
Comparison of learning curves. We evaluated the policies every 1,000 training time steps by testing the performance for one episode (the maximum length being 1,000 time steps). **(A)** Proposed method. **(B)** Excluding $1/(1+\omega_v)$ from Equation (2), i.e., the reward function was changed to $r(s_t, a_t) = -\omega_E|E_t - E_{t-1}| + \omega_v\dot{x} + f_{forward} + f_{alive} + f_{support}$. $1/(1+\omega_v)$ normalizes the value among the rewards that vary widely due to velocity term $\omega_v\dot{x}$. **(C)** The range of ω_v in LP1 and LP2 was set to [0,2.5]. The agent was trained to learn appropriate outputs for $\omega_v \in [0, 2.5]$ from the beginning. **(D)** The range of ω_v in LP1 and LP2 was changed to [1.5,2.5]. The agent was trained to learn appropriate outputs for larger ω_v first than smaller ω_v. **(E)** Without LP2. The reward function was set with LP1 parameters in the initial 500 thousand training time steps and with LP3 parameters in the rest of 9.5 million training time steps. **(F)** Without LP1. The reward function was set with LP2 parameters in the initial 3 million training time steps and with LP3 parameters in the rest of 7 million training time steps.

segments h, the generated running was extremely susceptible to the noise applied to h. In the IP model, which is a simple conceptual walking model, the stance leg is rigid and its length remains constant. However, in the SLIP model, which is a simple conceptual running model, the stance leg is represented as a spring and the leg length varies. Therefore, it can be inferred that the parameter h, which indicates the length of the leg of the biped model, is an important parameter in generating the running gait.

The gait speed obtained using the bipedal model was $0.51 \leq \dot{x} \leq 3.02\,\text{m/s}$. The proposed bipedal model has a wide range of speeds. These results indicate that by utilizing body dynamics effectively, the bipedal model can move over a wide range of speeds. It should be emphasized that the hip joints of the proposed bipedal model are completely passive. The actuation of the hip joint has a significant effect on locomotion velocity (Dzeladini et al., 2014; Bailey et al., 2017). For example, humans change the activity pattern of the muscles around the hip joint, e.g., gluteus maximus and rectus femoris, depending on their speed (Cappellini et al., 2006). To move faster, the proposed bipedal model leaned its body forward and vigorously moved its legs up and down to obtain a larger propulsion force in the stance phase.

Interestingly, the required time for the transition to the other limit cycles of the gait was different between the walk-to-run transition and run-to-walk transition: walk-to-run was faster, whereas run-to-walk required additional time steps. Moreover, Figure 10A shows that the kinetic energy increases with increase in ω_v in the walk-to-run transition, despite the time delay between the start of decrease in ω_v and decrease in the kinetic energy peak in the run-to-walk transition, as shown in Figure 10B. These observations can be attributed to hysteresis, i.e., the transition process dynamics depend on the previous gait pattern, which is attributed to the transition between different attractors (Diedrich and Warren, 1995). As shown in Figure 12, the differences in body posture owing to gait dynamics were confirmed. We assume that these differences in the attractor dynamics owing to the gait characteristics resulted in differences in the convergence process.

Comparing Figures 13A, B, it can be observed that reducing inter-reward variability improved the reliability of policy learning in multimodal locomotion. This is because the reward function without $1/(1+\omega_v)$ calculates widely different values, depending on the given ω_v as mentioned in designing the reward function. The agent was trained with a bias toward relatively high ω_v to obtain higher cumulative rewards. This fact suggests that the variation among rewards should be as small as possible during the policy-learning process to eliminate bias. The results also indicated that the ω_v range of the training progress affected the performance of the learned controller. It was also observed that the model failed to learn the walking gait when the value of ω_v was set to high in the initial learning stage (Figure 13D). Additionally, learning locomotion with a wide range of ω_v from the beginning makes training difficult (Figures 13C, E). These facts suggest that training to learn low-speed locomotion initially and then gradually transferring to high-speed locomotion is the key to effective multimodal locomotion learning.

In this study, the proposed method was validated exclusively through simulations. Therefore, building a hardware and verifying its feasibility in the real world would be of primary interest in future studies. Because the controller realized in this study is simple and is able to adapt to a new surface containing steps, we believe that the trained controller can be applied to suitable hardware. However, the gaps between the simulation environment and hardware need to be solved. A major gap is control latency. In this study, we did not consider the time delay between the sensors, controller, and actuators. In addition, as shown in Figure 6, some of the input parameters of the controller, namely, $\dot{\theta}$, φ, $\dot{\varphi}$, h, and θ were sensitive to noise in steady gait. Therefore, the hardware must be designed to accurately measure these parameters. Moreover, the bipedal model motion was achieved only in a forward straight line. Hence, the extension of the motion space into three dimensions and designing of a learning framework that can change the motion direction are other important issues that need to be resolved.

5. Conclusion

Gait generation in underactuated robots requires control solutions that can achieve stability with input from a limited number of active actuators. To reproduce multimodal locomotion, it is necessary to provide control solutions that generate motion patterns for drastically different modes in terms of dynamics, which is an extremely challenging optimization problem. Thus, multimodal locomotion using passive dynamics is an extremely challenging problem. Despite being limited to physical simulation, this study provided evidence that a bipedal model with completely passive hip joints was able to learn various motions, including walking, running, and gait transition, through DRL. Therefore, we believe that this study provides a framework that will enable walking and running with the efficient use of body morphology in bipedal robots.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

KK, DO, and MH conceived the study. SK developed a bipedal model and learning scheme. All the authors participated equally in the preparation of the manuscript.

Funding

This work was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research on Innovative Areas Hyper-Adaptability Project (JP20H05458 and JP22H04764) and Science of Soft Robots Project (JP21H00317, JP18H03167, JP19K22855, and JP20H04260).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbot.2022.1054239/full#supplementary-material

References

Alexander, R. M. (1996). Walking and running. *Math. Gazette* 80, 262–266. doi: 10.2307/3619558

Bailey, J. P., Mata, T., and Mercer, J. (2017). Is the relationship between stride length, frequency, and velocity influenced by running on a treadmill or overground? *Int. J. Exerc. Sci.* 10, 1067–1075.

Bhounsule, P. A., Cortell, J., and Ruina, A. (2012). "Design and control of ranger: an energy-efficient, dynamic walking robot," in *Proceedings of CLAWAR2012* (Maryland, MD), 441–448. doi: 10.1142/97898144159580057

Blickhan, R. (1989). The spring-mass model for running and hopping. *J. Biomech.* 22, 1217–1227. doi: 10.1016/0021-9290(89)90224-8

Brendan, T., Nicolas, H., and Akansel, C. (2020). Guided curriculum learning for walking over complex terrain. *arXiv:2010.03848 [cs.RO]*. doi: 10.48550/arXiv.2010.03848

Cappellini, G., Ivanenko, Y. P., Poppele, R. E., and Lacquaniti, F. (2006). Motor patterns in human walking and running. *J. Neurophysiol.* 95, 3426–3437. doi: 10.1152/jn.00081.2006

Castillo, G. A., Weng, B., Zhang, W., and Hereid, A. (2021). "Robust feedback motion policy design using reinforcement learning on a 3D digit bipedal robot," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Prague: IEEE), 5136–5143. doi: 10.1109/IROS51168.2021.9636467

Cavagna, G. (2006). The landing-take-off asymmetry in human running. *J. Exp. Biol.* 209, 4051–4060. doi: 10.1242/jeb.02344

Cavagna, G., Willems, P., and Heglund, N. (2000). The role of gravity in human walking: pendular energy exchange, external work and optimal speed. *J. Physiol.* 528(Pt 3), 657. doi: 10.1111/j.1469-7793.2000.00657.x

Cavagna, G. A., and Legramandi, M. A. (2020). The phase shift between potential and kinetic energy in human walking. *J. Exp. Biol.* 223, 232645. doi: 10.1242/jeb.232645

Cavagna, G. A., Thys, H., and Zamboni, A. (1976). The sources of external work in level walking and running. *J. Physiol.* 262, 639–657. doi: 10.1113/jphysiol.1976.sp011613

Collins, S. H., and Ruina, A. (2005). "A bipedal walking robot with efficient and human-like gait," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (Barcelona: IEEE), 1983–1988.

Dickinson, M. H., Farley, C. T., Full, R. J., Koehl, M., Kram, R., and Lehman, S. (2000). How animals move: an integrative view. *Science* 288, 100–106. doi: 10.1126/science.288.5463.100

Diedrich, F. J., and Warren Jr, W. H. (1995). Why change gaits? dynamics of the walk-run transition. *J. Exp. Psychol.* 21, 183. doi: 10.1037/0096-1523.21.1.183

Dzeladini, F., Van Den Kieboom, J., and Ijspeert, A. (2014). The contribution of a central pattern generator in a reflex-based neuromuscular model. *Front. Hum. Neurosci.* 8, 371. doi: 10.3389/fnhum.2014.00371

Farley, C. T., and Gonzalez, O. (1996). Leg stiffness and stride frequency in human running. *J. Biomech.* 29, 181–186. doi: 10.1016/0021-9290(95)00029-1

Fu, Z., Kumar, A., Malik, J., and Pathak, D. (2021). Minimizing energy consumption leads to the emergence of gaits in legged robots. *arXiv:2111.01674 [cs.RO*]. doi: 10.48550/arXiv.2111.01674

Haarnoja, T., Ha, S., Zhou, A., Tan, J., Tucker, G., and Levine, S. (2018a). Learning to walk via deep reinforcement learning. *arXiv preprint* arXiv:1812.11103. doi: 10.15607/RSS.2019.XV.011

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., et al. (2018b). Soft actor-critic algorithms and applications. *arXiv preprint* arXiv:1812.05905. doi: 10.48550/arXiv.1812.05905

Hansen, A. H., Childress, D. S., and Knox, E. H. (2004). Roll-over shapes of human locomotor systems: effects of walking speed. *Clin. Biomech.* 19, 407–414. doi: 10.1016/j.clinbiomech.2003.12.001

Hodgins, J. K. (1991). "Biped gait transitions," in *Proceedings of 1991 IEEE International Conference on Robotics and Automation*, Vol. 3 (Sacramento, CA: IEEE), 2092–2097. doi: 10.1109/ROBOT.1991.131936

Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, D., Tsounis, V., Koltun, V., et al. (2019). Learning agile and dynamic motor skills for legged robots. *Sci. Rob.* 4, eaau5872. doi: 10.1126/scirobotics.aau5872

Kang, G.-C., and Lee, Y. (2021). Finite state machine-based motion-free learning of biped walking. *IEEE Access* 9, 20662–20672. doi: 10.1109/ACCESS.2021.3055241

Kobayashi, T., Hasegawa, Y., Sekiyama, K., Aoyama, T., and Fukuda, T. (2016). "Unified bipedal gait for walking and running by dynamics-based virtual holonomic constraint in PDAC," in *2016 IEEE International Conference on Robotics and Automation (ICRA)* (Stockholm: IEEE), 1769–1775.

Kuo, A. D. (2007). The six determinants of gait and the inverted pendulum analogy: a dynamic walking perspective. *Hum. Mov. Sci.* 26, 617–656. doi: 10.1016/j.humov.2007.04.003

Kwon, O., and Park, J. H. (2003). "Gait transitions for walking and running of biped robots," in *2003 IEEE International Conference On Robotics and Automation (Cat. No. 03CH37422)*, *Vol. 1* (Taipei: IEEE), 1350–1355.

Li, Z., Cheng, X., Peng, X. B., Abbeel, P., Levine, S., Berseth, G., et al. (2021). "Reinforcement learning for robust parameterized locomotion control of bipedal robots," in *2021 IEEE International Conference on Robotics and Automation (ICRA)* (Xian: IEEE), 2811–2817. doi: 10.1109/ICRA48506.2021.9560769

McGeer, T. (1990). Passive dynamic walking. *Int. J. Robotics Res.* 9, 62–82. doi: 10.1177/027836499000900206

Minetti, A., and Alexander, R. M. (1997). A theory of metabolic costs for bipedal gaits. *J. Theor. Biol.* 186, 467–476. doi: 10.1006/jtbi.1997.0407

Nagasaka, K., Kuroki, Y., Suzuki, S., Itoh, Y., and Yamaguchi, J. (2004). "Integrated motion control for walking, jumping and running on a small bipedal entertainment robot," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004, Vol. 4* (New Orleans, LA: IEEE), 3189–3194.

Nikooyan, A. A., and Zadpoor, A. A. (2011). Mass-spring-damper modelling of the human body to study running and hopping-an overview. *Proc. Instit. Mech. Eng. H J. Eng. Med.* 225, 1121–1135. doi: 10.1177/0954411911424210

Okajima, S., Tournier, M., Alnajjar, F. S., Hayashibe, M., Hasegawa, Y., and Shimoda, S. (2018). Generation of human-like movement from symbolized information. *Front. Neurorobot.* 12, 43. doi: 10.3389/fnbot.2018.00043

Owaki, D., Koyama, M., Yamaguchi, S., Kubo, S., and Ishiguro, A. (2011). A 2-d passive-dynamic-running biped with elastic elements. *IEEE Trans. Rob.* 27, 156–162. doi: 10.1109/TRO.2010.2098610

Owaki, D., Osuka, K., and Ishiguro, A. (2008). "On the embodiment that enables passive dynamic bipedal running," in *2008 IEEE International Conference on Robotics and Automation* (Pasadena, CA: IEEE), 341–346.

Pfeifer, R., and Gómez, G. (2009). "Morphological computation-connecting brain, body, and environment," in *Creating Brain-Like Intelligence* (Berlin; Heidelberg: Springer Berlin Heidelberg), 66–83.

Pfeifer, R., and Scheier, C. (2001). *Understanding Intelligence*. MIT Press.

Ruina, A., Bertram, J. E., and Srinivasan, M. (2005). A collisional model of the energetic cost of support work qualitatively explains leg sequencing in walking and galloping, pseudo-elastic leg behavior in running and the walk-to-run transition. *J. Theor. Biol.* 237, 170–192. doi: 10.1016/j.jtbi.2005.04.004

Saputra, A. A., Ijspeert, A. J., and Kubota, N. (2020). "A neural primitive model with sensorimotor coordination for dynamic quadruped locomotion with malfunction compensation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Las Vegas, NV: IEEE), 3783–3788.

Shao, Y., Jin, Y., Liu, X., He, W., Wang, H., and Yang, W. (2021). Learning free gait transition for quadruped robots via phase-guided controller. *IEEE Rob. Autom. Lett.* 7, 1230–1237. doi: 10.1109/LRA.2021.3136645

Sharbafi, M. A., and Seyfarth, A. (2017). *Bioinspired Legged Locomotion*. Joe Hayton.

Siekmann, J., Godse, Y., Fern, A., and Hurst, J. (2021). "Sim-to-real learning of all common bipedal gaits via periodic reward composition," in *2021 IEEE International Conference on Robotics and Automation (ICRA)* (Xi'an: IEEE), 7309–7315.

Smit-Anseeuw, N., Gleason, R., Vasudevan, R., and Remy, C. D. (2017). The energetic benefit of robotic gait selection–a case study on the robot ramone. *IEEE Rob. Autom. Lett.* 2, 1124–1131. doi: 10.1109/LRA.2017.2661801

Sreenath, K., Park, H.-W., Poulakakis, I., and Grizzle, J. W. (2013). Embedding active force control within the compliant hybrid zero dynamics to achieve stable, fast running on mabel. *Int. J. Rob. Res.* 32, 324–345. doi: 10.1177/0278364912473344

Srinivasan, M., and Ruina, A. (2006). Computer optimization of a minimal biped model discovers walking and running. *Nature* 439, 72–75. doi: 10.1038/nature04113

Tajima, R., Honda, D., and Suga, K. (2009). "Fast running experiments involving a humanoid robot," in *2009 IEEE International Conference on Robotics and Automation* (Kobe: IEEE), 1571–1576.

Todorov, E., Erez, T., and Tassa, Y. (2012). "Mujoco: a physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (Vilamoura: IEEE), 5026–5033.

Xie, Z., Berseth, G., Clary, P., Hurst, J., and van de Panne, M. (2018). "Feedback control for cassie with deep reinforcement learning," *in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 1241–1246. doi: 10.1109/IROS.2018.8593722

Xie, Z., Ling, H. Y., Kim, N. H., and van de Panne, M. (2020). Allsteps: curriculum-driven learning of stepping stone skills. *Comput. Graph. Forum* 39, 213–224. doi: 10.1111/cgf.14115

Yu, W., Turk, G., and Liu, C. K. (2018). Learning symmetric and low-energy locomotion. *ACM Trans. Graph.* 37, 1–12. doi: 10.1145/3197517.3201397

Yue, Z., and Mester, J. (2002). A model analysis of internal loads, energetics, and effects of wobbling mass during the whole-body vibration. *J. Biomech.* 35, 639–647. doi: 10.1016/S0021-9290(01)00243-3