machines

MDPI

Article

Realization of a Human-like Gait for a Bipedal Robot Based on Gait Analysis †

Junsei Yamano [1,*], Masaki Kurokawa [1], Yuki Sakai [1] and Kenji Hashimoto [2,*]

[1] Mechanical Engineering Program, Graduate School of Science and Technology, Meiji University, Kawasaki 214-8571, Japan
[2] Graduate School of Information, Production and Systems, Waseda University, Kitakyushu 808-0135, Japan
* Correspondence: ce222089@meiji.ac.jp (J.Y.); kenji.hashimoto@waseda.jp (K.H.)
† This paper is an extended version of our paper published in 26th International Conference Series on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR 2023), Florianópolis, Brazil, 2–4 October 2023.

Abstract: There are many studies analyzing human motion. However, we do not yet fully understand the mechanisms of our own bodies. We believe that mimicking human motion and function using a robot will help us to deepen our understanding of humans. Therefore, we focus on the characteristics of the human gait, and the goal is to realize a human-like bipedal gait that lands on its heels and takes off from its toes. In this study, we focus on kinematic synergy (planar covariation) in the lower limbs as a characteristic gait seen in humans. Planar covariation is that elevation angles at the thigh, shank, and foot in the sagittal plane are plotted on one plane when the angular data are plotted on the three axes. We propose this feature as a reward for reinforcement learning. By introducing this reward, the bipedal robot achieved a human-like bipedal gait in which the robot lands on its heels and takes off from its toes. We also compared the learning results with those obtained when this feature was not used. The results suggest that planar covariation is one factor that characterizes a human-like gait.

Keywords: bipedal robot; deep reinforcement learning; gait analysis

check for updates

Citation: Yamano, J.; Kurokawa, M.; Sakai, Y.; Hashimoto, K. Realization of a Human-like Gait for a Bipedal Robot Based on Gait Analysis. *Machines* **2024**, *12*, 92. https://doi.org/10.3390/machines12020092

Academic Editors: Manuel F. Silva, Ebrahim Samer El Youssef, Mohammad Osman Tokhi and Leonardo Mejia Rincon

Received: 19 December 2023
Revised: 19 January 2024
Accepted: 23 January 2024
Published: 25 January 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

There are many studies that have analyzed the mechanisms and movements of the human body [1–4]. However, we do not yet fully understand the mechanisms and functions of our own bodies.

To date, robots have been used to analyze various angles to understand the mechanisms and movements of the human body. Kamikawa et al. developed a five-finger prosthetic hand with an underactuated mechanism optimized for human-like grasp force distribution that can grasp a variety of objects robustly [5]. This confirmed the effectiveness of human-like grasp force distribution in robust power grasping with the five-finger prosthetic hand. Hashimoto et al. developed a human-like foot mechanism in WABIAN-2R that mimics the elastic properties of the arch of the human's foot and changes in arch height during walking to clarify the function of the foot arch structure [6,7]. Several walking experiments were conducted using a gait stabilization control based on gait analysis to quantitatively clarify the function of the arch structure [8]. Cheng et al. developed the Computational Brain (CB) [9], which was created to explore the fundamental processing of the human brain while dealing with the real world. Ude et al. focused on visual attention models and eye movement, saccadic eye movements, and Vestibulo-ocular reflex (VOR), and implemented them in CBs, thereby achieving eye movement responses comparable to those of humans [10]. Kryczka et al. focused on the tendency of humans to naturally stabilize their head posture while walking and proposed a head stabilization controller based on an IMU installed in the head of KOBIAN [11,12]. Using this control, the vibration

of the robot's head was shown to be reduced during walking and when subjected to large-amplitude vibrations. To clarify the relationship between the behavior of torso motion during static and dynamic walking and the center of pressure (CoP) under each foot of the walker, Ferreira et al. used four force sensors under each foot to obtain the CoP. This confirmed the influence of the human torso angle during walking on the CoP position [13]. Based on the above previous studies, we believe that mimicking humans, not only in appearance, but also in movement and function, will help to deepen our understanding of humans.

In this study, we aim to understand the characteristics of the human gait by applying them to the bipedal robot and to realize a human-like bipedal gait in which the robot lands on its heels and takes off from its toes. We focus on kinematic synergy in the lower limb and aim to realize a human-like bipedal gait by reproducing kinematic synergy in simulation using deep reinforcement learning.

The article, first published in the proceedings of the 26th International Conference on Climbing and Walking Robots (CLAWAR 2023), is an extended description of the concept introduced in [14]. The paper is organized as follows: Section 2 describes the characteristics of human gait, the simulation environment, and the control method used in this study; Section 3 evaluates the gait locomotion and learning results; Section 4 summarizes the results and discusses future work.

2. Materials and Methods

2.1. Characteristics of the Human Gait

In this study, we focus on the kinematic synergy [15] (planar covariation), highlighted by Borghese et al., as a characteristic gait of humans. Planar covariation is the angular data being placed on a single plane when the elevation angles at the thigh, shank, and foot are plotted on three axes in the sagittal plane of a human body. This allows humans to lower and control the redundant degrees of freedom of the joints. This kinematic synergy is not dependent on the walking speed. And it is observed when walking on hard or soft surfaces, or when climbing stairs or hopping [16]. The contribution of the elevation angles to human walking on hard and soft surfaces was determined through principal component analysis and is shown in Table 1. The first and second principal components are the basis vectors of the approximate plane, and the third principal component is the normal vector of the approximate plane. It is known that the cumulative contribution ratio up to the second principal component is more than 99% in humans during walking and that the first and second principal components do not differ significantly between the left and right feet. It is also seen not only in humans but also in mammals, such as cats and Japanese macaques [17], so it is considered to be an essential characteristic of the gait.

Table 1. The contribution of the elevation angle when humans walk on hard and soft surfaces.

Surface	1st	2nd	3rd
Hard	77.4%	22.0%	0.6%
Soft	72.0%	27.4%	0.6%

Figure 1 shows the definition of elevation angles at the thigh, shank, and foot. As a concrete example, the elevation angles at the thigh, shank, and foot are plotted in 3D space when walking on hard and soft surfaces and are shown in Figure 2. As shown in Figure 2, there is no significant change in the three elevation angles during the heel strike and toe-off during the human gait. This indicates that the leg motion during walking is cyclically repeated.

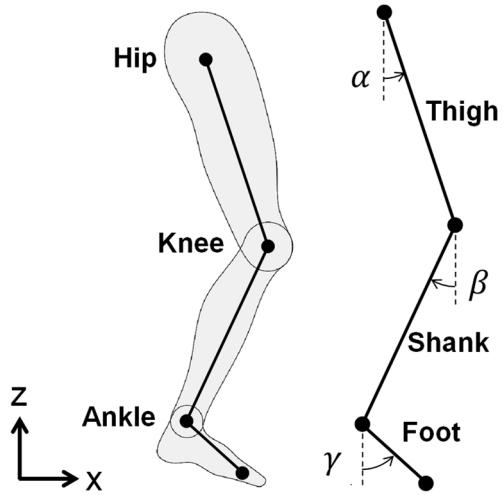
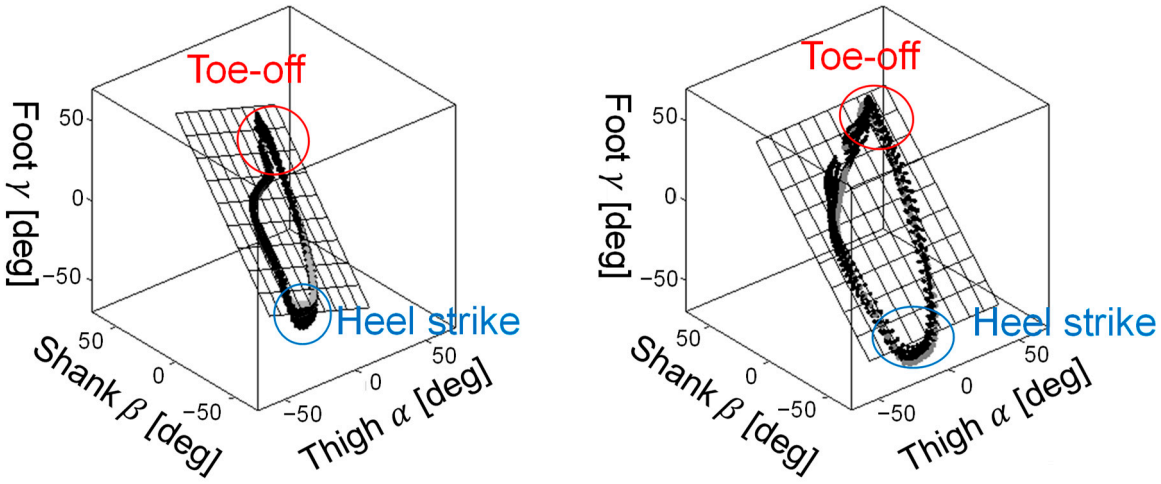


Figure 1. Definition of elevation angles at the thigh, shank, and foot.



(**a**) (**b**)

Figure 2. This figure plots the elevation angles at the thigh, shank, and foot during walking in 3-D space; (**a**) plotted data for walking on a hard surface; (**b**) plotted data for walking on a soft surface.

Although planar covariation has been used for robot control [18,19], these often use multiple controls in combination. This study aims to realize a human-like gait by setting a planar covariation as a reward function for reinforcement learning. The reward function is based on planar covariation and does not depend on other control elements. This method is expected to verify the hypothesis that planar covariation reflects kinematic synergies in a human-like gait.

2.2. Simulation Environment

In this study, we used PyBullet Gymperium (version: 0.1) [20], open-source software based on the physics simulator named PyBullet [21]. We used the environment Walker2DPyBulletEnv-v0, designed to make the bipedal robot walk. We use walker2d for the bipedal robot model. This robot model has three joints (hip joint, knee joint, and ankle joint) for each of the two legs. The total of six joints in both legs is controlled by applying torque. Table 2 shows the range of motion of each joint, which is based on that of a human [22].

Table 2. The joint's range of motion.

Joint Name	Angle [deg]
Hip Joint	−20~135
Knee Joint	−140~0
Ankle Joint	−45~25

2.3. Control Method

2.3.1. Deep Reinforcement Learning

The main walking control methods for the bipedal robots are model-based control and learning control. In this study, we use learning control, which has potential for future development and has been actively studied [23–26].

Humans learn to walk as infants by adjusting their movement methods through trial and error according to their abilities and the characteristics of the terrain [27]. This learning process is considered similar to the process of reinforcement learning. In reinforcement learning, the actions that maximize future value are learned through trial and error. The neural network used in deep learning is said to mimic the mechanism of human neurons [28]. For these reasons, among learning controls, we believe that deep reinforcement learning can be employed to incorporate human gait characteristics as rewards for acquiring a human-like bipedal gait in various environments.

We first explain reinforcement learning. In reinforcement learning, the agent executes actions according to the policy, and then, the environment returns rewards to the agent as feedback based on the state of the position, velocity, posture, joint angles, etc., and the agent's choice of action. The agent improves the policy so that the cumulative reward received in the sequence of actions becomes larger. The above is repeated to find the optimal solution.

Reinforcement learning has the problem that it can only handle discrete values of the action space. Therefore, we combine the high-function approximation ability of neural networks with the action selection of reinforcement learning. As a result, the continuous action space can be learned with continuous values without artificially discretizing it, and a system that can respond more flexibly to changes in the state can be realized.

In this study, we use the policy gradient method, in which the policy is a function expressed by a certain parameter, and the policy can be learned directly by learning the parameter. Moreover, using the REINFORCE algorithm and Gaussian policy, the policy gradient methods can be used in a continuous action space [29].

2.3.2. Policy Gradient Methods

The update equation of the policy parameter θ can be expressed as follows:

$$\theta_{t+1} = \theta_t + \alpha \nabla_\theta J(\theta_t) \tag{1}$$

where α is the learning rate factor, and ∇_θ denotes the partial differential vector with respect to θ, where θ is a multidimensional vector.

In the policy gradient method, we consider the problem of maximizing the objective function, the expected return $J(\theta)$. We define the objective function $J(\theta)$ as the value function $v_\pi(s_0)$ calculated under the policy $\pi(a|s_0, \theta)$ at the time of the start of learning, since this is the expected return at state s_0.

$$J(\theta) = v_\pi(s_0) \equiv \mathbb{E}_\pi[G_t|S_t = s_0] \tag{2}$$

where G_t is the cumulative discounted reward, and S_t is the state variable at time t. The gradient $\nabla_\theta J(\theta)$ with respect to the policy parameters of the objective function is expressed

as follows for any policy $\pi(a|s_0, \theta)$, differentiable with respect to the parameter θ, and for the objective function $J(\theta)$, defined by Equation (2) [30].

$$\nabla_\theta J(\theta) = \mathbb{E}_\pi[\nabla_\theta \log \pi(a|s, \theta) q_\pi(s, a)] \tag{3}$$

where a is the action and $q_\pi(s, a)$ is the action value function. There are two problems to calculate Equation (3): "It is difficult to calculate the expected value" and "Estimation of the action value function is necessary".

The solution to the first problem is to use Monte Carlo approximation, which is a method for approximating the expected value. This allows the expected value to be calculated even when the probability distribution is complex. Based on the probabilistic policy π, the action is executed for T steps, and the gradient is approximated from the obtained state, action, and reward observations.

$$\nabla_\theta J(\theta) \approx \frac{1}{T} \sum_{t=0}^{T-1} (\nabla_\theta \log \pi(A_t|S_t, \theta)) q_\pi(S_t, A_t) \tag{4}$$

The solution to the second problem is to use the REINFORCE algorithm, which does not require direct estimation of the action value function but approximates it with the actual cumulative discounted reward or reward obtained. The action value function $q_\pi(s, a)$ in Equation (4) is approximated based on the cumulative discounted reward G_t.

$$G_t = \sum_{k=1}^{T-t} \gamma^{k-1} R_{t+k} \tag{5}$$

$$\nabla_\theta J(\theta) \approx \frac{1}{T} \sum_{t=0}^{T-1} (\nabla_\theta \log \pi_\theta(A_t|S_t)) G_t \tag{6}$$

where γ is the discount rate factor, and R_t is the reward variable at time t. The REINFORCE algorithm reduces the variance of the action value function $q_\pi(s, a)$ by introducing a baseline, which is a function $b(s)$ that provides a reference for the action value function $q_\pi(s, a)$. This makes learning easier to converge. The equation with the baseline is shown below.

$$\nabla_\theta J(\theta) = \mathbb{E}_\pi[(\nabla_\theta \log \pi(a|s, \theta))(q_\pi(s, a) - b(s))] \tag{7}$$

The value function $v_\pi(s, a)$ is a weighted average function of the action value function $q_\pi(s, a)$ with the policy probability $\pi(a|s)$, by definition. Therefore, the value function $v_\pi(s, a)$ is used as the baseline in this case. We define the advantage function, which represents the quantity of the action value measured with respect to its mean value, as follows.

$$A^\pi(s, a) = q_\pi(s, a) - v_\pi(s) \tag{8}$$

The expected value can be expressed by substituting Equation (8) into Equation (7) and computing the Monte Carlo approximation as follows.

$$\begin{aligned} \nabla_\theta J(\theta) &= \mathbb{E}_\pi[\nabla_\theta \log \pi(a|s, \theta) A^\pi(s, a)] \\ &\approx \tfrac{1}{T} \sum_{t=0}^{T-1} \nabla_\theta \log \pi(A_t|S_t, \theta) A^\pi(S_t, A_t) \end{aligned} \tag{9}$$

where A_t is the action variable at time t. The action value function $q_\pi(s, a)$ included in the advantage function is approximated based on the cumulative discounted reward G_t using the REINFORCE algorithm. Then, the update formula for θ is calculated as follows.

$$\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta) \tag{10}$$

$$\nabla_\theta J(\theta) = \nabla_\theta \log \pi(A_t|S_t, \theta) A^\pi(S_t, A_t) \tag{11}$$

$$A^{\pi}(s,a) = G_t - v_{\pi}(s) \tag{12}$$

2.3.3. Neural Network

In this study, we use two types of neural networks.

The first is a neural network used for the Gaussian policy. To clarify the objective function, the derivative operator is taken from Equation (11) and multiplied by -1 to obtain Equation (13) as the loss function. RMSprop is used as the optimization algorithm, and the parameters are updated to minimize this loss function. The input, hidden, and output layers are listed in Table 3.

$$-\log \pi(A_t|S_t,\theta)A^{\pi}(S_t,A_t) \tag{13}$$

Table 3. Neural network for the Gaussian Policy.

Layer	Detail	Dimension
Input	Body position	3
	Body velocity	3
	Body posture	2
	Joint angle	6
	Joint angular velocity	6
	Foot ground state	2
Hidden	Layer 1	220
	Layer 2	114
	Layer 3	60
Output	Mean	6
	Variance	6

The second is a neural network for estimating the state value function. This neural network is designed to output the cumulative discounted rewards, using Equation (15) as the loss function. Adam is used as the optimization algorithm, and the parameters are updated to minimize this loss function. The input, hidden, and output layers are listed in Table 4.

$$\sum_{t=0}^{T-1}(V - G_t)^2 \tag{14}$$

Table 4. Neural network for estimating the state value function.

Layer	Detail	Dimension
Input	Body position	3
	Body velocity	3
	Body posture	2
	Joint angle	6
	Joint angular velocity	6
	Foot ground state	2
Hidden	Layer 1	220
	Layer 2	33
	Layer 3	5
Output	State value function	1

A schematic diagram of the learning algorithm used in this study is shown in Figure 3. We define one step as the next action to be taken in a certain state according to Gaussian policy and execute that action to obtain the next state and reward. One episode is defined as repeating the process until the termination conditions are met: the bipedal robot falls over, or the maximum number of steps is reached. When the termination condition is met, the neural network of the state value function is updated based on the history of states,

actions, and rewards, and the cumulative discounted reward is calculated. The advantage function is then updated, and the neural network of the Gaussian policy is updated. In this study, this learning process is repeated for 500,000 episodes.

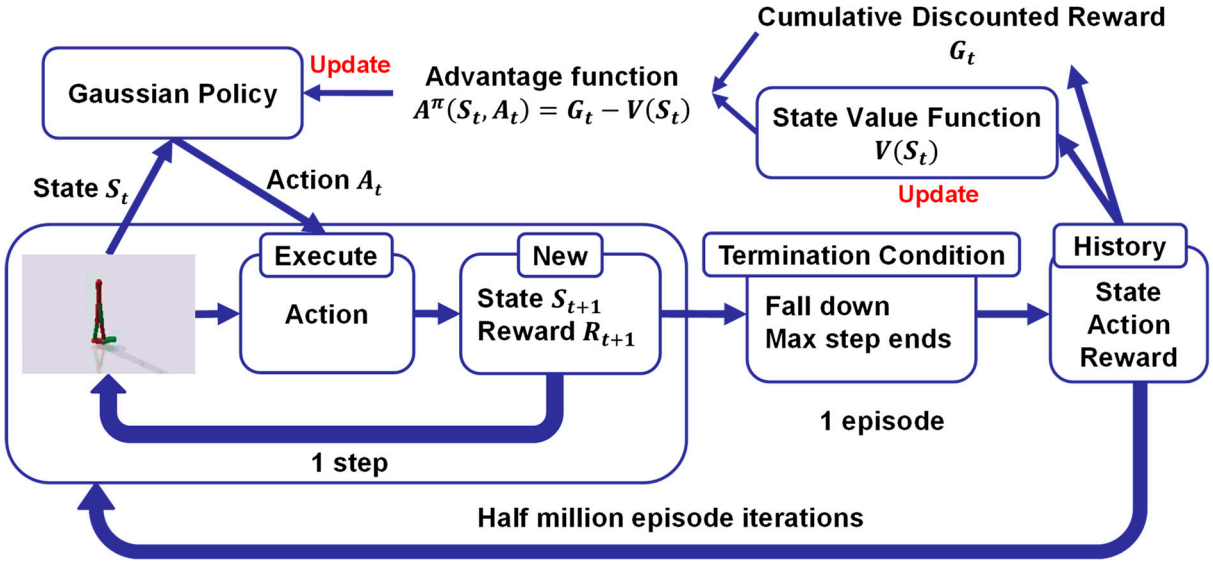


Figure 3. A schematic diagram of the learning algorithm.

2.3.4. Probabilistic Policy with the Gaussian Model

The Gaussian model is a typical example of a probabilistic policy for control in continuous space. This model samples a K-dimensional action vector a from a K-dimensional normal distribution with a mean $\mu(s)$ and covariance matrix $\sum(s)$ as parameters in state s. The equation is expressed as follows.

$$\pi(a|s,\theta) \propto \frac{1}{\sqrt{\det \sum(s)}} \exp\left(-\frac{1}{2}(a-\mu(s))^{\top} \sum(s)^{-1}(a-\mu(s))\right) \tag{15}$$

If the covariance matrix $\sum(s)$ is not diagonal, the different components of the action vector will interact. This complicates the implementation of the policy function in a neural network. Therefore, we assume that the components of the action vectors are independent and that the covariance matrix consists only of diagonal components. Let the k-th diagonal component of the covariance matrix be $\sigma_k^2(s)$, then the K-dimensional normal distribution can be decomposed into an independent 1-dimensional normal distribution formula as follows.

$$\begin{aligned} \pi(a|s,\theta) &= \prod_{k=1}^{K} \pi(a_k|s,\theta) \\ &\propto \prod_{k=1}^{K} \frac{1}{\sqrt{\sigma_k^2(s)}} \exp\left(-\frac{(a_k-\mu_k(s))^2}{2\sigma_k^2(s)}\right) \end{aligned} \tag{16}$$

where $\mu_k(s)$ and $\sigma_k^2(s)$ are functions with state s as input, obtained based on the neural network introduced in Section 2.3.3.

Using the Gaussian policy, $\log \pi(A_t|S_t,\theta)$ in Equation (13) can be calculated as follows.

$$\begin{aligned} \log \pi(A_t|S_t,\theta) &= \log \prod_{k=1}^{K} \pi(A_{tk}|S_{tk},\theta) \\ &= \sum_{k=1}^{K} \log \pi(A_{tk}|S_{tk},\theta) \\ &= \frac{1}{2} \sum_{k=1}^{K} \left(-\log \sigma_k^2(S_t) - \frac{(A_{tk}-\mu_k(S_t))^2}{2\sigma_k^2(S_t)}\right) \end{aligned} \tag{17}$$

2.3.5. Rewards

This study used the following four rewards for reinforcement learning.

- r_{alive}

This is the reward that keeps the upper body at a certain height and reduces the rotation of the body in the direction of the pitch. The equation is expressed as follows.

$$r_{alive} = \begin{cases} 1, \left(z_{body} > 0.8[m] \cap \left| \theta_{body} \right| < 1.0[rad] \right) \\ -1, (else) \end{cases} \tag{18}$$

where z_{body} is the height of the center of the body and θ_{body} is the angle of the center of the body along the pitch axis.

- $r_{progress}$

This is the reward for the distance advanced by the center of the body. First, we find the distance p_t from the position at time t to the target position.

$$p_t = -\sqrt{(x_d - x_t)^2 + (y_d - y_t)^2} \tag{19}$$

where x_d and y_d are the target positions in the x and y directions, respectively. The advanced distance d is expected based on the following equation.

$$d = p_t - p_{t-1} \tag{20}$$

The reward $r_{progress}$ is expressed by the following equation.

$$r_{progress} = k_{progress} \times d \tag{21}$$

where $k_{progress} = 121.2$ is the weight coefficient.

- r_{joint}

This is the reward for ensuring that each joint does not exceed its range of motion. The reward r_{joint} is expected based on the following equation.

$$r_{joint} = -k_{joint} \times n_{joint} \tag{22}$$

where $k_{joint} = 0.1$ is the weight coefficient, n_{joint} is the number of joints beyond the range of motion.

- r_{planar}

This is the reward for planar covariation, which is a characteristic of the human gait. Since the kinematic synergy is such that the elevation angles at the thigh, shank, and foot are placed on a single plane, it is sufficient to ensure that the three calculated elevation angles are placed on a single plane. The following is a detailed description of the application method and the calculation process.

First, we explain how the angles are read and when the rewards are executed. The angle of each joint is read each time an action is executed according to the policy. When the heel of the bipedal robot lands on the ground, the reward r_{planar} is executed and the angle data are reset.

Next, we explain how we take a plane from the angle readings. In this study, the plane is found by using the least-squares plane. The least-squares plane is a plane that is the least-squares distance from all the points in a 3-dimensional point cloud. The least-squares plane can be expressed as follows.

$$z = Ax + By + C \tag{23}$$

where A, B, and C are coefficients. We can obtain the unknowns A, B, and C using the lower–upper (LU) decomposition method. A least-squares plane is obtained by substituting the calculated coefficients into Equation (23). In this study, x is the elevation angle at the thigh, y is the elevation angle at the shank, and z is the elevation angle at the foot, so for simplicity, α is used for x, β for y, and γ for z. The plane $\gamma\prime$ is obtained by substituting each variable into Equation (23).

$$\gamma\prime = A\alpha + B\beta + C \tag{24}$$

The average of the squares ϵ of the difference between the plane and the elevation angle at the foot should be zero for the angle data to be placed on a single plane. Therefore, we train so that ϵ is close to zero. ϵ and the reward r_{planar} are expressed by the following equations, respectively.

$$\epsilon = \frac{\sum_{i=i}^{n} \left(\gamma_i' - \gamma_i\right)^2}{n_{data}} \tag{25}$$

$$r_{planar} = \begin{cases} k_{planar} \times \frac{0.5}{\epsilon}, (\epsilon \geq 1) \\ k_{planar} \times (1 - 0.5\epsilon), (\epsilon < 1) \end{cases} \tag{26}$$

where n_{data} is the number of the angle data, γ_i is the elevation angle at the i-th foot, and $k_{planar} = 3.0$ is the weight coefficient. Since planar covariation is valid for each of the left and right legs, the rewards in Equation (26) are also applied to each of the left and right legs.

3. Results

3.1. Walking Motion

3.1.1. Proposed Method

Figure 4 shows the walking motion of one step when the proposed method described in Section 2.3.5 is executed, walking in the order from (1) to (5). Figure 4 shows that the bipedal robot landed on its heels and took off its toes during walking. The elevation angles at the thigh, shank, and foot are plotted in 3-D space, and the views from the side and from above are shown in Figure 5. The blue points in each graph represent the timing of the heel strike, the red points represent the timing of the toe-off, and the yellow line in Figure 5a represents the least-squares plane. The contribution of the elevation angle is shown in Table 5.

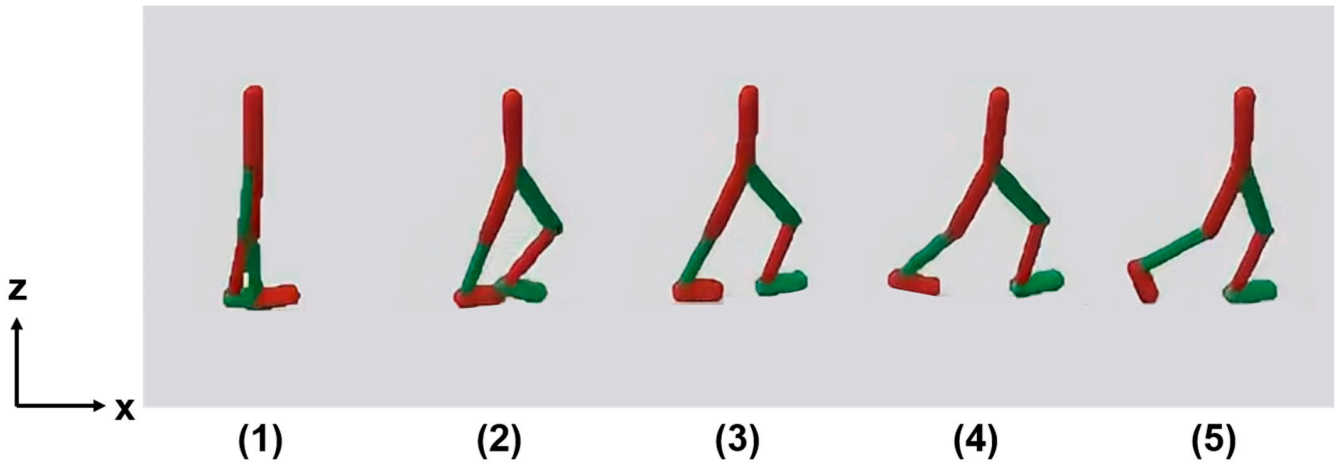


Figure 4. Walking motion with the proposed method. The red foot is the left side.

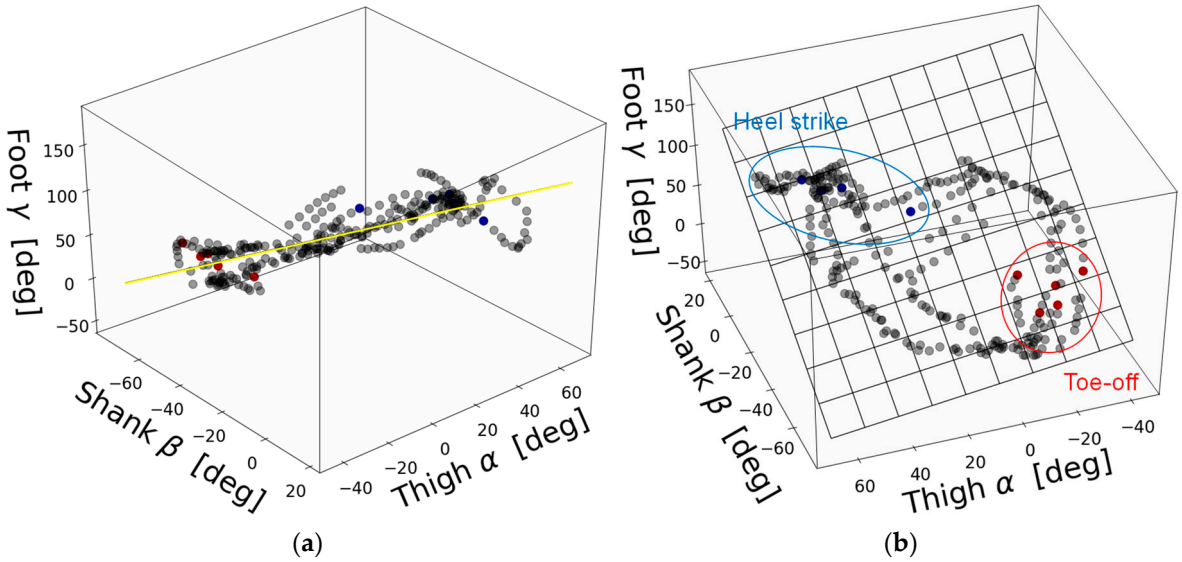


Figure 5. The elevation angles at the thigh, shank, and foot plotted in 3-D space using the proposed method; (**a**) viewed from the side with respect to the least-squares plane of the point cloud; (**b**) viewed from above with respect to the least-squares plane of the point cloud.

Table 5. The contribution of the elevation angle using the proposed method.

Leg	1st	2nd	3rd
Left	70.0%	27.1%	2.9%
Right	72.8%	24.0%	3.2%

3.1.2. Comparative Method

We use the rewards of Walker2DPyBulletEnv-v0 for a comparison with our proposed method. The rewards are the same for the body posture, the distance advanced by the center of the body, and the range of motion of the joints. Instead of a reward for the planar covariation, we use a deduction reward that prevents excessive torque and excessive power consumption. The following equations are used for each of these rewards.

$$r_{torque} = -k_{torque} \times \frac{\sum_{i=1}^{6} (a_{ti})^2}{6} \tag{27}$$

$$r_{electricity} = -k_{electricity} \times \frac{\sum_{i=1}^{6} |a_{ti} \times \omega_{ti}|}{6} \tag{28}$$

where a_{ti} and ω_{ti} are the action and angular velocity of each joint at a certain time t, and $k_{torque} = 0.1$ and $k_{electricity} = 2.0$ are the weight coefficient.

The walking motion of one step when the comparison method was executed is shown in Figure 6, with the steps taken in the order (1) through (5). As shown in Figure 6, the bipedal robot walked with its entire foot landing parallel to the road surface. Figure 7 shows the elevation angles at the thigh, shank, and foot plotted in 3-D space, viewed from the side and from above with respect to the least-squares plane of the point cloud. The blue points in each graph represent the timing of the heel strike, the red points represent the timing of the toe-off, and the yellow line in Figure 7a represents the least-squares plane. The contribution of the elevation angle is shown in Table 6.

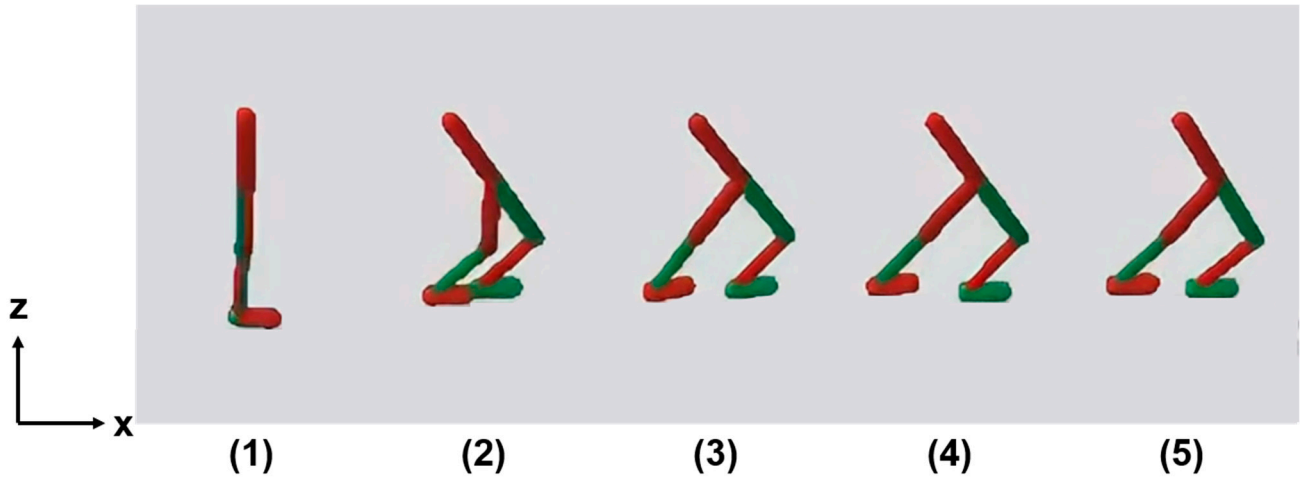


Figure 6. Walking motion with the comparative method. The red foot is the left side.

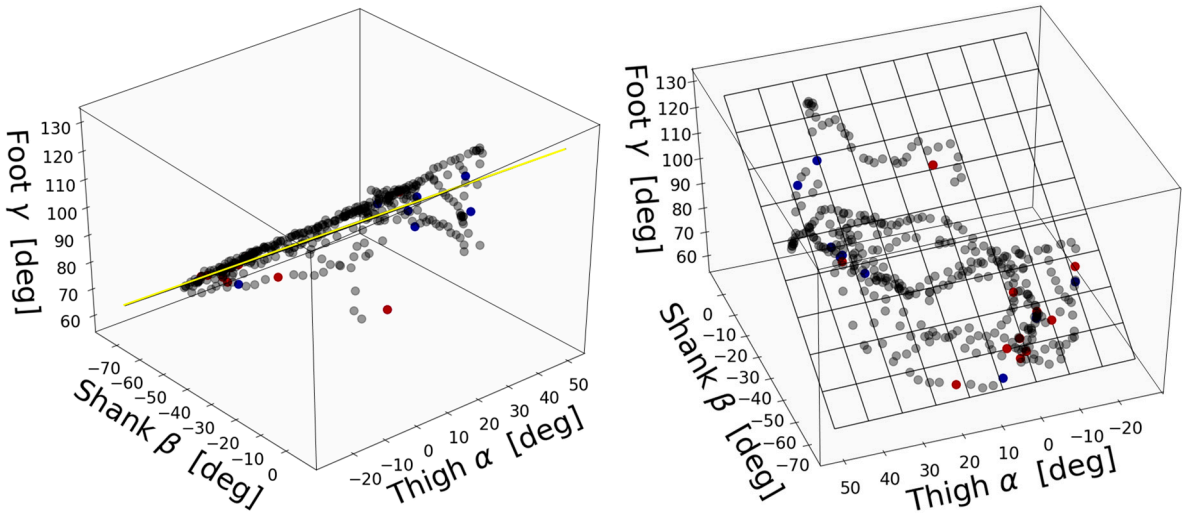


Figure 7. The elevation angles at the thigh, shank, and foot plotted in 3-D space using the comparative method; (**a**) viewed from the side with respect to the least-squares plane of the point cloud; (**b**) viewed from above with respect to the least-squares plane of the point cloud.

Table 6. The contribution of the elevation angle using the comparative method.

Leg	1st	2nd	3rd
Left	64.9%	33.3%	1.8%
Right	78.8%	17.8%	3.4%

3.2. Learning Curve

Figures 8 and 9 show the learning curves of the number of steps, cumulative reward, and loss function of the neural network representing the Gaussian policy in one episode using the proposed method and the comparison method, respectively. From these figures, it is clear that more rewards are acquired, and the loss function decreases as the learning progresses. However, the loss function in Figure 8 has not converged compared to Figure 9, suggesting that the learning is insufficient.

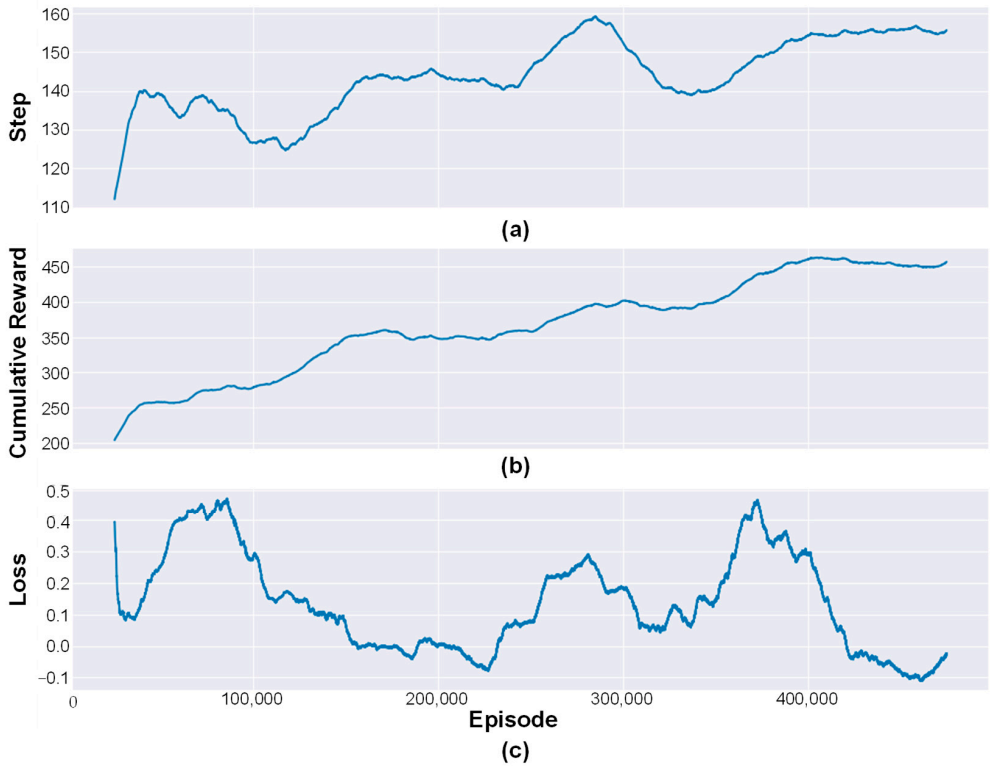


Figure 8. Learning curve for the proposed method. (**a**) number of steps per episode; (**b**) the cumulative reward obtained in one episode; (**c**) loss function of the neural network representing Gaussian policy.

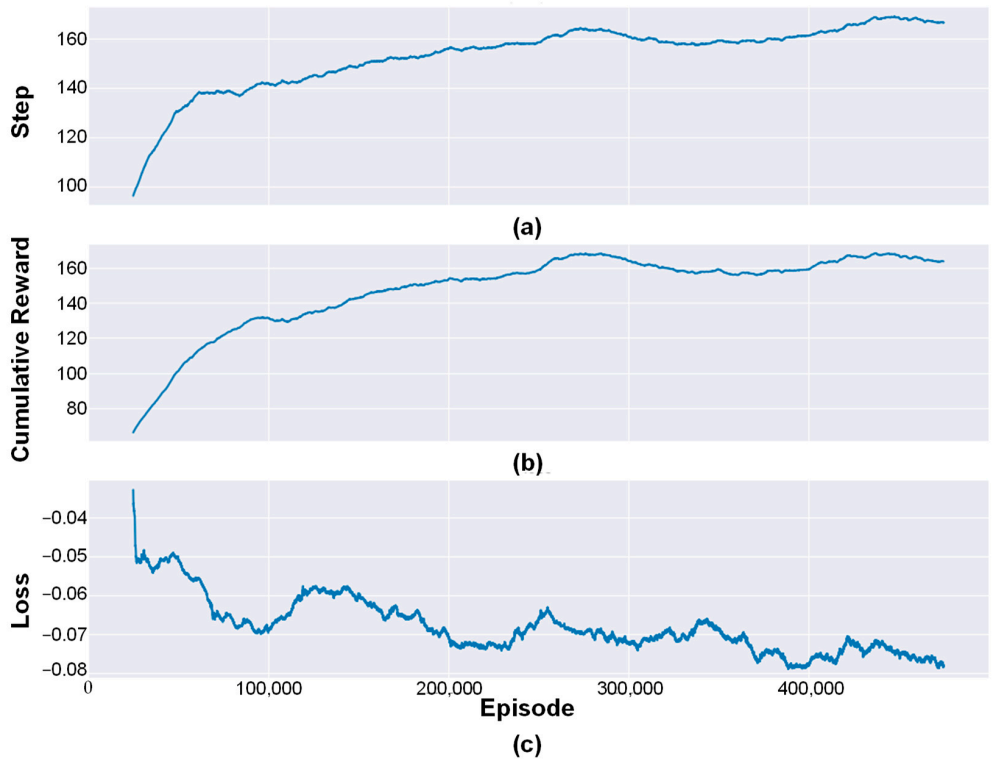


Figure 9. Learning curve for the comparative method. (**a**) number of steps per episode; (**b**) the cumulative reward obtained in one episode; (**c**) loss function of the neural network representing Gaussian policy.

4. Discussion and Future Work

This paper describes the gait control of a bipedal robot that aims to realize a human-like bipedal gait, in which the robot lands on its heels and takes off from its toes, using deep reinforcement learning in the simulation. We achieved a human-like gait by incorporating planar covariation as a part of the reinforcement learning reward.

The cumulative contribution of the proposed method and the comparison method to the second principal component is 97% and 97~98%, respectively. However, the comparison method did not produce a human-like gait, suggesting that planar covariation has other effects besides the fact that the elevation angles at the thigh, shank, and foot are placed on a single plane.

Comparing Figures 5 and 7, it is clear that the proposed method does not significantly change the three elevation angles during heel strike and toe-off, as is the case with humans. On the other hand, in the comparative method, the red and blue points are unevenly distributed, suggesting that different movements are generated each time.

It is also known that human gait consists of continuous periodic and symmetrical movements produced by a precise series of coordinated movements, alternating between one leg and the other [31]. In the human and the proposed method, the contribution ratio of each principal component does not differ significantly for the left and right legs. On the other hand, the contribution rates of the first and second principal components differ by nearly 15% for the left and right legs in the comparison method. This suggests that the comparison method cannot consider the symmetry of the movement of the left and right legs.

These results indicate that planar covariation has an effect of symmetry of the left and right legs, in which the three elevation angles do not change significantly during heel strike and toe-off. In the future, we will consider not only the lower limbs but also the upper limbs, head, and other human characteristics. We would like to explore coordinated movements of the whole body to characterize a human-like gait.

Author Contributions: Conceptualization, J.Y., M.K., Y.S. and K.H.; methodology, J.Y., M.K., Y.S. and K.H.; software, J.Y., M.K. and Y.S.; validation, J.Y.; formal analysis, J.Y.; investigation, J.Y.; data curation, J.Y.; writing—original draft preparation, J.Y., M.K. and Y.S.; writing—review and editing, J.Y. and K.H.; visualization, J.Y.; supervision, K.H.; project administration, K.H.; funding acquisition, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI Grant Numbers JP20H04267 and JP21H05055. This work was also supported by a Waseda University Grant for Special Research Projects (Project number: 2023C-179).

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: This study was conducted with the support of the Information, Production and Systems Research Center, Waseda University; Future Robotics Organization, Waseda University; and as a part of the humanoid project at the Humanoid Robotics Institute, Waseda University.

Conflicts of Interest: M.K. and Y.S. graduated from Meiji University in March 2023, and all authors declare no conflict of interest.

References

1. Tomomitsu, M.S.V.; Alonso, A.C.; Morimoto, E.; Bobbio, T.G.; Greve, J.M.D. Static and dynamic postural control in low-vision and normal-vision adults. *Clinics* **2013**, *68*, 517–521. [CrossRef] [PubMed]
2. Pozzo, T.; Berthoz, A.; Lefort, L. Head stabilization during various locomotor tasks in humans. *Exp. Brain Res.* **1990**, *82*, 97–106. [CrossRef]
3. Venkadesan, M.; Yawar, A.; Eng, C.M.; Dias, M.A.; Singh, D.K.; Tommasini, S.M.; Haims, A.H.; Bandi, M.M.; Mandre, S. Stiffness of the human foot and evolution of the transverse arch. *Nature* **2020**, *579*, 97–100. [CrossRef] [PubMed]
4. Bohm, S.; Mersmann, F.; Santuz, A.; Schroll, A.; Arampatzis, A. Muscle-specific economy of force generation and efficiency of work production during human running. *eLife* **2021**, *10*, e67182. [CrossRef] [PubMed]

5. Kamikawa, Y.; Maeno, T. Underactuated Five-Finger Prosthetic Hand Inspired by Grasping Force Distribution of Humans. In Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Nice, France, 22–26 September 2008; pp. 717–722.
6. Ogura, Y.; Aikawa, H.; Shimomura, K.; Morishima, A.; Lim, H.O.; Takanishi, A. Development of a New Humanoid Robot WABIAN-2. In Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA), Orlando, FL, USA, 15–19 May 2006; pp. 76–81.
7. Hashimoto, K.; Takezaki, Y.; Hattori, K.; Kondo, H.; Takashima, T.; Lim, H.O.; Takanishi, A. A Study of Function of Foot's Medial Longitudinal Arch Using Biped Humanoid Robot. In Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Taipei, Taiwan, 18–22 October 2010; pp. 2206–2211.
8. Hashimoto, K.; Takezaki, Y.; Motohashi, H.; Otani, T.; Kishi, T.; Lim, H.O.; Takanishi, A. Biped Walking Stabilization Based on Gait Analysis. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, MN, USA, 14–18 May 2012; pp. 154–159.
9. Cheng, G.; Hyon, S.H.; Morimoto, J.; Ude, A.; Colvin, G.; Scroggin, W.; Jacobsen, S.C. CB: A Humanoid Research Platform for Exploring NeuroScience. In Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots, Genova, Italy, 4–6 December 2006; pp. 182–187.
10. Ude, A.; Wyar, V.; Lin, L.H.; Cheng, G. Distributed Visual Attention on a Humanoid Robot. In Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan, 5 December 2005; pp. 381–386.
11. Endo, N.; Takanishi, A. Development of whole-body emotional expression humanoid robot for ADL-assistive RT services. *J. Robot. Mechatron.* **2011**, *23*, 969–977. [CrossRef]
12. Kryczka, P.; Falotico, E.; Hashimoto, K.; Lim, H.O.; Takanishi, A.; Laschi, C.; Dario, P.; Berthoz, A. A Robotic Implementation of a Bio-Inspired Head Motion Stabilization Model on a Humanoid Platform. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 2076–2081.
13. Ferreira, J.P.; Crisóstomo, M.M.; Coimbra, A.P. Human Gait Acquisition and Characterization. *IEEE Trans. Instrum. Meas.* **2009**, *58*, 2979–2988. [CrossRef]
14. Yamano, J.; Kurokawa, M.; Sakai, Y.; Hashimoto, K. Walking Motion Generation of Bipedal Robot Based on Planar Covariation Using Deep Reinforcement Learning. In Proceedings of the 26th International Conference on Climbing and Walking Robots (CLAWAR 2023), Florianópolis, Brazil, 2–4 October 2023.
15. Borghese, N.A.; Bianchi, L.; Lacquaniti, F. Kinematic determinants of human locomotion. *J. Physiol.* **1996**, *494*, 863–879. [CrossRef] [PubMed]
16. Ivanenko, Y.P.; d'Avella, A.; Poppele, R.E.; Lacquaniti, F. On the Origin of Planar Covariation of Elevation Angles During Human Locomotion. *J. Neurophysiol.* **2008**, *99*, 1890–1898. [CrossRef] [PubMed]
17. Ogihara, N.; Kikuchi, T.; Ishiguro, Y.; Makishima, H.; Nakatsukasa, M. Planar covariation of limb elevation angles during bipedal walking in the Japanese macaque. *J. R. Soc. Interface* **2012**, *9*, 2181–2190. [CrossRef] [PubMed]
18. Ha, S.S.; Yu, J.H.; Han, Y.J.; Hahn, H.S. Natural Gait Generation of Biped Robot based on Analysis of Human's Gait. In Proceedings of the 2008 International Conference on Smart Manufacturing Application, Goyangi, Republic of Korea, 9–11 April 2008; pp. 30–34.
19. Ghiasi, A.R.; Alizadeh, G.; Mirzaei, M. Simultaneous design of optimal gait pattern and controller for a bipedal robot. *Multibody Syst. Dyn.* **2010**, *23*, 410–429. [CrossRef]
20. GitHub Benelot/Pybullet-Gym. Available online: https://github.com/benelot/pybullet-gym (accessed on 8 December 2023).
21. PyBllet. Available online: https://pybullet.org/wordpress/ (accessed on 8 December 2023).
22. Kapandji, A.I.; Owerko, C. *The Physiology of the Joints: 2 The Lower Limb*, 7th ed.; Handspring Publishing Limited: London, UK, 2019.
23. Wang, S.; Chaovalitwongse, W.; Babuska, R. Machine learning algorithms in bipedal robot control. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2012**, *42*, 728–743. [CrossRef]
24. Xie, Z.; Berseth, G.; Clary, P.; Hurst, J.; van de Panne, M. Feedback Control For Cassie With Deep Reinforcement Learning. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1241–1246.
25. Tsounis, V.; Alge, M.; Lee, J.; Farshidian, F.; Hutter, M. DeepGait: Planning and Control of Quadrupedal Gaits Using Deep Reinforcement Learning. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3699–3706. [CrossRef]
26. Wang, Z.; Wei, W.; Xie, A.; Zhang, Y.; Wu, J.; Zhu, Q. Hybrid Bipedal Locomotion Based on Reinforcement Learning and Heuristics. *Micromachines* **2022**, *13*, 1688. [CrossRef]
27. Adolph, K.E.; Bertenthal, B.I.; Boker, S.M.; Goldfield, E.C.; Gibson, E.J. Learning in the Development of Infant Locomotion. *Monogr. Soc. Res. Child Dev.* **1997**, *62*, 1–140. [CrossRef]
28. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **2018**, *4*, e00938. [CrossRef] [PubMed]
29. Sutton, R.S.; Barto, A.G. *Reinforcement Learning*, 2nd ed.; The MIT Press: Cambridge, MA, USA, 2018.

30. Sutton, R.S.; McAllester, S.; Singh, S.; Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In Proceedings of the 12th International Conference on Neural Information Processing Systems, Denver, CO, USA, 29 November–4 December 1999.
31. Castermans, T.; Duvinage, M.; Cheron, G.; Dutoit, T. Towards Effective Non-Invasive Brain-Computer Interfaces Dedicated to Gait Rehabilitation Systems. *Brain Sci.* **2014**, *4*, 1–48. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.