Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Speed and Morphology Conditioned Fourier Gait Synthesis for Bipedal Locomotion Learning

Yusuf Baran Ateş

baran.ates@bilkent.edu.tr

Bilkent University

Ömer Morgül

Bilkent University

Research Article

Keywords: Frequency domain motion priors, Harmonic analysis, Quasi-periodic signals, Gait synthesis, Imitation learning, Bipedal locomotion

Posted Date: September 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10638551/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Speed and Morphology Conditioned Fourier Gait Synthesis for Bipedal Locomotion Learning

Yusuf Baran Ateş[1*] and Ömer Morgül[1]

[1]Department of Electrical and Electronics Engineering, Bilkent University, Ankara, 06800, Türkiye.

*Corresponding author(s). E-mail(s): baran.ates@bilkent.edu.tr;
Contributing authors: morgul@ee.bilkent.edu.tr;

Abstract

Human gait is a quasi-periodic multi-channel signal whose cycle duration changes with walking speed, which makes it hard to feed to a model with a fixed-width input or output. A fixed-duration segment of joint angle data is sampled from a motion capture corpus and its discrete Fourier transform is taken, so a segment of any cycle duration becomes a fixed-dimensional coefficient vector and the reconstruction is periodic by construction rather than learned. A small network conditioned on walking speed and the two leg lengths is trained to predict these coefficients and inverse discrete Fourier transform is used to reconstruct the reference signal for joint angles. Candidate training objectives are formulated and compared by a single validation metric computed in joint space so that losses defined in different domains stay comparable. On held-out subjects, our proposed network trained on frequency domain performs better than spline interpolation, recorded-cycle playback, a coupled oscillator, and a residual variant on pointwise error, dynamic time warping distance, and harmonic-magnitude fidelity. The network is then used as the reference signal generator for a torque-controlled **7**-DoF planar biped trained with Proximal Policy Optimization (PPO) using imitation learning, and the resulting policies are tested off-distribution and compared against baseline methods.

Keywords: Frequency domain motion priors, Harmonic analysis, Quasi-periodic signals, Gait synthesis, Imitation learning, Bipedal locomotion

1 Introduction

Human walking produces a multi-channel signal that is periodic with respect to the gait phase (the progress through a single step cycle), but not periodic with respect to time.

While each joint traces a repeating motion pattern over every stride, the time it takes to complete a stride isn't fixed—it shrinks from about three seconds at a slow pace to under one second at a fast pace, depending on leg length and walking speed. A recorded corpus is therefore a set of segments of unequal length encoding the same cyclic structure at different time scales, and a model with fixed-width input or output must resolve that mismatch first. The usual fixes — zero-padding, cutting every cycle down to the shortest one, or resampling in time — either add spectral content that is not in the signal or throw away the link between speed and duration that makes the corpus informative.

The representation of gait signal can be formulated in different ways. Playback of recorded cycles

is exact at the speeds recorded and undefined between them, and spline interpolation over a speed-indexed library smooths the gaps but stays anchored to the same discrete samples. Coupled-oscillator models are periodic and compact, but their waveform is fixed by the oscillator equations rather than fitted to measurement, so they result in unnatural gait patterns. Recurrent models emit cycles of any length, but periodicity is then something the network must hold, and small state errors accumulate into phase drift and eventually result in a non-periodic signal.

Resampling each cycle onto a fixed grid in gait phase before transforming removes this mismatch rather than absorbing it. Every cycle then maps to a coefficient vector of the same length, whatever its duration. The inverse transform of any such vector is periodic by construction, so no extrapolation can produce a drifting output.

This paper makes four contributions. We define the phase-normalized spectral representation. We train a compact encoder that maps walking speed and limb lengths to the harmonic coefficients, and pick among candidate training objectives with a single validation metric that stays comparable across them. We compare the representation against four alternatives at the same code size on held-out subjects. Finally we use it as the reference signal for a torque-controlled walking policy, and measure how far that policy carries outside the conditions it was trained on.

2 Related Work

Treating a gait waveform as a truncated Fourier series is long-established in biomechanics. Ren et al. [1] build a predictive model of a complete gait cycle on a harmonic parameterization of the joint kinematics, and Chau [2] surveys harmonic descriptions of gait waveforms alongside the alternatives. The appeal is always the same: a walking waveform is dominated by its first few stride harmonics, so a handful of coefficients captures most of its energy.

Functional principal component analysis reaches the same compaction from data, estimating an orthogonal basis from the waveform ensemble [3] rather than fixing a sinusoidal one; Deluzio and Astephen [4] separate clinical populations on knee gait waveforms with a few such components. That basis is mean-square optimal for the training ensemble but is estimated rather than being analytic, and its components carry no explicit periodicity.

Another approach is to use Central Pattern Generators (CPG) to produce rhythm from coupled nonlinear oscillators without a reference waveform. Taga et al. [5] obtain stable planar walking from neural oscillators coupled to a musculo-skeletal model, and robotic formulations built on Matsuoka's model [6] span a low-dimensional, hand-tuned manifold of waveforms. There are studies that use this approach to provide reference signals to tune bipedal walking controllers [7, 8].

Using a reference signal to tune bipedal walking controllers is a common approach [9, 10]. DeepMimic [11] rewards joint-space tracking of a single recorded clip and introduces reference-state initialization and early termination; the same recipe transfers to legged hardware when the demonstrations are retargeted animal motions [12]. Adversarial motion priors [13, 14] replace explicit clip tracking with a discriminator score over a dataset and are competitive with hand-designed reward functions. Closest to the present setting, Su and Huang [15] produce multi-speed bipedal walking by imitating human motion capture, and surveys of learned bipedal control place such methods alongside purely task-driven ones [16].

The present study differs from these works in that it defines a compact frequency-domain network that produces gait patterns for different walking speeds and morphologies, and demonstrates its effectiveness as a reference generator in a torque-controlled bipedal walking task.

3 Motion-capture Corpus and Data Preparation

The source signals are the multimodal gait recordings of Schreiber and Moissenet [17] : 1143 trials from 50 participants (24 female, 26 male; age 37.0 $\pm$ 13.6 yr, height 1.74 $\pm$ 0.09 m, mass 71.0 $\pm$ 12.3 kg), captured optically at 100 Hz across five walking-speed regimes. Notably, the dataset preserves the individual body dimensions of each participant, including leg lengths, unlike other motion capture datasets that retarget every recording onto a canonical skeleton. Keeping the

original limb lengths is what allows the network to be conditioned on morphology as well as speed.

Figure 1(a) summarizes the preparation. Inverse kinematics gives six sagittal joint angles, each low-pass filtered at 5 Hz — a cutoff in the range conventionally used for lower-limb kinematics [18] — and decimated to 10 Hz, and foot-strike events segment the recording into cycles labelled with mean forward velocity v. Each cycle is then resampled to $N = 32$ points in gait phase rather than in time, so its transform bins are stride harmonics rather than frequencies in hertz and bin k picks out the same feature of the gait at every speed. A real 32-point transform leaves 17 unique bins, giving 34 numbers per joint and $D = 204$ over the six joints. The cycle duration is stored beside the code, read off the segmentation and never estimated by a model.

Two consequences motivate the construction. The code length does not depend on how long the cycle lasted, so a fixed-width regressor can be fitted across the corpus without padding or truncation; and the inverse transform of any coefficient vector repeats exactly over one cycle, so the synthesized reference is periodic by construction rather than by training. The cutoff frequency and sample count are coupled (32 samples per cycle at a 5 Hz cutoff yields a maximum unaliased cycle duration of 3.2 s), which accommodates the corpus's maximum cycle length of 3.01 s for the slowest walking speed.

4 Conditioned Spectral Encoder and Reconstruction

4.1 Mapping, architecture, and objectives

The encoder is the regression

$$f_\theta : \; x = [\, v,\; \ell_r,\; \ell_l \,] \longmapsto \hat{C} \in \mathbb{R}^{204}, \qquad (1)$$

with the forward speed normalized by 2.4 m/s and ℓ_r, ℓ_l the right and left leg lengths. Both sides are fixed-size, a direct consequence of the frequency domain representation. The trunk is three fully connected layers of 512 units with LeakyReLU(0.1) activations (Fig. 1(b)) read by a linear head; a second head serving a four-joint variant outside the scope of this paper shares the trunk and does not affect the results here. The model has fewer than 0.6 M parameters and evaluates in well under a millisecond on CPU, so it can be queried inside a control loop.

For the training of the network, three candidate objectives are formulated,

$$\mathcal{L}_{\text{freq}} = \text{MSE}\big(\hat{C},\, C\big), \qquad (2)$$

$$\mathcal{L}_{\text{time}} = \text{MSE}\big(\hat{q},\, q\big), \qquad (3)$$

$$\mathcal{L}_{\text{comb}} = \alpha\, \mathcal{L}_{\text{freq}} + (1 - \alpha)\, \mathcal{L}_{\text{time}}, \qquad (4)$$

Here $\hat{C}$ and C are the predicted and recorded harmonic coefficients — the frequency-domain quantities — while $\hat{q}$ and q are the corresponding joint-angle waveforms over one cycle, obtained from those coefficients by inverse transform, $\alpha \in (0, 1)$ is the mixing weight, and MSE stands for Mean Squared Error.

Because the three losses are measured in different units, their validation values cannot be compared, and a model scored by its own training loss is scored on the thing it was fitted to. Model selection therefore uses one external criterion,

$$\mathcal{V}_{\text{fair}} = \text{MSE}\big(\hat{q}_{\text{val}},\, q_{\text{val}}\big) \quad [\text{rad}^2], \qquad (5)$$

where q_{val} is the recorded joint-angle waveform on the validation split and $\hat{q}_{\text{val}}$ the inverse transform of the coefficients predicted for those same cycles. It is computed in joint space whatever domain the configuration was trained in, and drives early stopping, hyperparameter selection, and the ranking across configurations alike.

4.2 Training protocol

Two architectural variants — direct prediction, and a residual variant predicting a correction to a cubic-spline prior — are crossed with the three objectives, and each of the six configurations is searched over hidden width $\{256, 512\}$, learning rate $\{3 \times 10^{-4}, 10^{-3}\}$, and batch size $\{32, 64\}$, for 48 runs. All use Adam with weight decay 10^{-5}, unit-norm gradient clipping, a ReduceLROnPlateau schedule stepped on $\mathcal{V}_{\text{fair}}$, and early stopping with patience 250 up to 4,000 epochs under 10-fold cross-validation. The selected configuration is direct prediction with $\mathcal{L}_{\text{freq}}$: three layers of 512 units, 3×10^{-4}, batch 64 (Table 1).

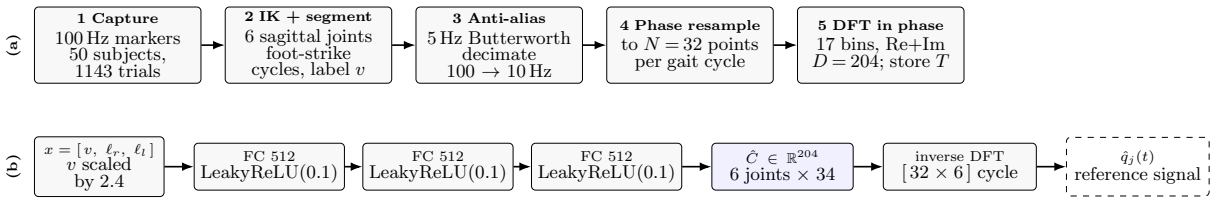


Fig. 1 (a) The encoding pipeline. Stage 4 resamples each cycle to $N = 32$ points in gait phase; stage 5 takes the DFT of the phase-resampled cycle, so its bins are stride harmonics and not frequencies in Hz. (b) The conditioned encoder: speed and the two leg lengths in, harmonic code out, inverse transformed into the reference waveform the controller consumes.

Table 1 Representation constants, encoder architecture, and training configuration.

Representation		*Encoder*		*Training*	
Phase samples N	32 per gait cycle	Input	$[\,v,\ \ell_r,\ \ell_l\,]$, v scaled by 2.4	Objective	$\mathcal{L}_{\text{freq}}$ on coefficients
Unique bins	17, Re+Im	Hidden layers	3 fully connected, 512 units	Optimizer	Adam, $\eta = 3 \times 10^{-4}$
Code dimension D	204 (6 joints × 34)	Activation	LeakyReLU(0.1)	Weight decay / clip	10^{-5} / $\|\nabla\|_2 \le 1$
Low-pass / decimation	5 Hz Butterworth, 100→10 Hz	Output head	$\hat{C} \in \mathbb{R}^{204}$, linear	Batch size	64
Longest admissible cycle	3.2 s	Parameters	< 0.6 M	LR schedule	Plateau on $\mathcal{V}_{\text{fair}}$
Cycle duration T	from segmentation, not predicted	Grid search	width, lr, batch (48 runs)	Epochs	$\le$ 4,000, patience 250

4.3 Reconstruction results

Every result here is computed on held-out *subjects*, not held-out cycles: with held-out cycles, recorded-cycle playback would be retrieving neighbours of its own training data and would score artificially well. Test cycles are grouped into seven speed bins spanning 0.35–2.2 m/s, and aggregates are sample-weighted means over bins because the bin populations are uneven. Three different test metrics are used. First, per-joint mean squared error in rad^2 for pointwise fidelity. Second, dynamic time warping distance [19] for shape fidelity after optimal monotone realignment, which separates a genuine shape error from a small phase offset. Third, spectral fidelity S, the mean harmonic-magnitude error over the first three stride harmonics, which measures the error in the same domain as the representation.

Figure 2 shows example outputs on four held-out subjects, one per speed setting. The four cycles differ widely in duration yet are encoded by vectors of the same length, which is the practical effect of the phase normalization.

Table 2 and Fig. 3 compare the learned harmonic code against alternative representations at a matched budget of $D = 204$ on the same subject split. Recorded-cycle playback retrieves the nearest-speed held-out-subject cycle; the cubic spline interpolates a speed-bucketed library; the coupled oscillator is a Matsuoka-style model [5, 6]; and the residual variant regresses a correction to the spline prior.

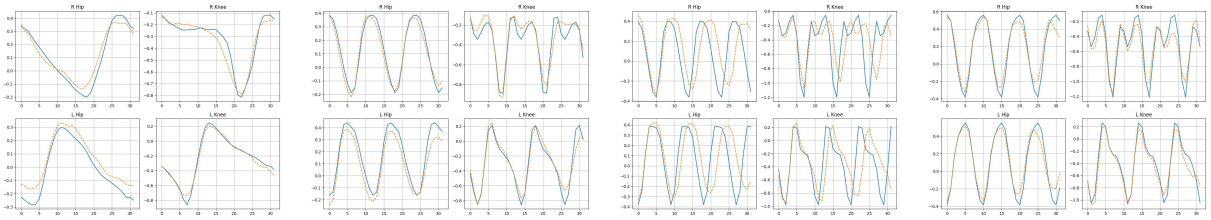
Table 2 Aggregate accuracy across representation bases at matched code budget $D = 204$, sample-weighted over held-out subjects; lower is better throughout. Conditioning on $[\,v, \ell_r, \ell_l\,]$ is applied wherever the basis admits it.

Representation	MSE [rad^2]	DTW	S
Phase-norm. spectral (ours)	**0.0185**	**0.129**	**0.378**
Residual spectral MLP	0.0212	0.198	0.553
Cubic spline	0.0284	0.232	0.592
Recorded-cycle playback	0.0311	0.253	0.620
Coupled oscillator	0.0728	0.637	2.270

Three readings follow. First, the learned harmonic code is the most accurate representation on all three metrics, with a mean MSE about 40% below both the spline and playback, and the lead widens at the ends of the speed range, where a sample-anchored representation has least to interpolate between.

Second, what separates the learned code from the coupled oscillator is the *fitted spectrum*, not periodicity. The oscillator is exactly periodic by construction and still scores worst on every metric. Due to its inability to produce natural gait patterns, it is the only method whose error rises monotonically with speed.

Third, the result survives a change of architecture: the residual variant also outperforms the other methods, proving that the learned spectral prior is effective in improving the reconstruction accuracy.



(a) 0.31 m/s (b) 0.91 m/s (c) 1.39 m/s (d) 1.93 m/s

Fig. 2 Example network outputs on four held-out subjects; solid: decoded from the predicted code, dashed: recorded cycle, for right and left hip and knee. The horizontal axis is the phase-sample index, identical in all four panels whatever the cycle duration.

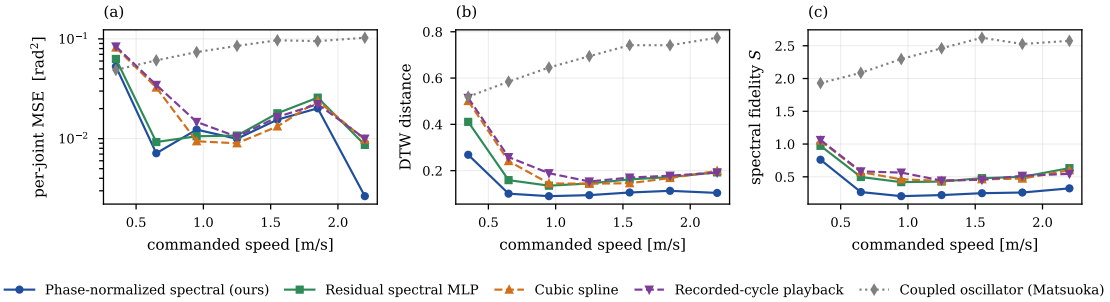


Fig. 3 Reconstruction error against commanded speed on held-out subjects: (a) per-joint MSE, log scale; (b) DTW distance; (c) spectral fidelity S.

5 Reinforcement Learning Control on a Planar Biped

The gait generator network is tested inside a feedback loop with a controller to test its ability to be used as a reference to achieve human-like walking. The controller loop can be seen in figure 4.

5.1 Simulation, observation, and control scheme

The plant is a 7-DoF planar biped: a torso–hip joint plus bilateral hip, knee, and ankle. Link masses, lengths, and inertias are those of the SIMBICON model [20], with head, arms, and trunk lumped into a single upper-body segment. All experiments run in PyBullet, with the physics integrated at 1 kHz and the policy queried at 100 Hz. The torque commands pass through a first-order low-pass filter before reaching the actuators, which bridges the two clock domains and suppresses the chatter a raw torque command at 100 Hz would otherwise produce (Fig. 4).

The observation is a compact history-plus-preview vector, chosen empirically for training stability and following the practice of Peng et al. [12]. It combines the immediate past (q_{t-1}, a_{t-1}), the present (q_t, $\dot{q}_t$, v_{com}), and the future, in the form of the decoded reference at the current step and at two horizons ahead. The result is 55-dimensional, given in table 3; references are 6-D while the measured vectors are 7-D, the difference being the torso joint, which has no reference. At the 100 Hz policy rate, previewing at t+1 and t+10 looks 10 ms and 100 ms ahead: the near-term preview informs the immediate torque, while the longer-horizon preview signals the upcoming part of the cycle so the policy can prepare for it instead of merely reacting. The action space is a vector of desired torques for the left and right hip, knee, and ankle joints and the torso joint, totaling 7 dimensions.

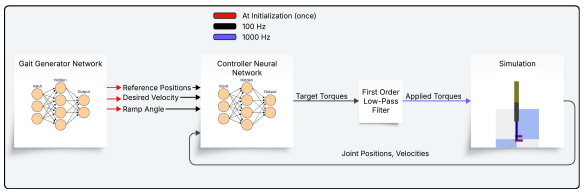


Fig. 4 The control loop. The encoder is queried once per episode with the commanded speed and the robot's leg lengths; its code is decoded at 100 Hz into the reference preview the policy consumes.

Table 3 Components of the 55-D observation s_t.

Component	Dim.	Description
v_{ds}	1	commanded forward speed
r_{angle}	1	ramp pitch angle (rad)
v_{com}	1	measured CoM forward speed
q_{t-1}, q_t	7+7	past and current joint angles
$\hat{q}_t, \hat{q}_{t+1}, \hat{q}_{t+10}$	6+6+6	current and preview reference
a_{t-1}	7	previous torques
$\dot{q}_t$, $\dot{\hat{q}}_t$	7+6	actual and reference velocities

5.2 Reward design

The reward blends imitation of the decoded reference with task and stability terms,

$$r(t) = \omega_{\text{im}}(t)\, r_{\text{im}}(t) + \omega_{\text{gait}}(t)\, r_{\text{gait}}(t). \quad (6)$$

The first term makes the policy follow the trajectories the encoder supplies and is responsible for the human-likeness of the gait; the second rewards forward progress, contact consistency, and energy economy, and is responsible for the robustness. The weights $\omega_{\text{im}}(t)$ and $\omega_{\text{gait}}(t)$ set the relative importance of the two terms. In different training methods, the weights are not fixed as detailed in equation 12. From here on the time index t is dropped from r_{im}, r_{gait}, and their constituent terms for brevity; every quantity below is still evaluated per step. In the imitation term hips and knees are matched strongly and ankles softly, in both position and velocity:

$$r_{\text{im}} = \sum_{j\in\{h,k,a\}} w_j\, e^{-5\|\Delta q_j\|} + \tfrac{1}{5}\, w_j\, e^{-0.2\|\Delta \dot{q}_j\|}, \quad (7)$$

where Δq_j and $\Delta \dot{q}_j$ stack the left and right position and velocity errors of joint pair $j \in \{\text{hip}, \text{knee}, \text{ankle}\}$ and $w_h = w_k = 0.75$, $w_a = 0.25$. The asymmetry is deliberate: the ankle is the joint through which the robot answers local slope, surface noise, and impact, so constraining it tightly degrades ground interaction on ramps and rough terrain, whereas hip and knee trajectories govern centre-of-mass progression and stance height and tolerate being held harder. The task term collects four contributions,

$$r_{\text{gait}} = r_{\text{alive}} + r_{\text{contact}} + r_{\text{speed}} - r_{\text{torque}}. \quad (8)$$

With centre-of-mass height z_{com} and torso angle q_{torso}, the posture term rewards a safe band and terminates the episode outside it,

$$r_{\text{alive}} = \begin{cases} +0.5, & z_{com} \in [0.75,\, 1.25], \\ -100, & |q_{\text{torso}}| > 0.9 \text{ or } z_{com} > 1.25 \\ & \text{or } z_{com} < 0.75, \end{cases} \quad (9)$$

the last case also raising the termination flag, so a fall costs the remainder of the episode as well as the penalty. Speed tracking uses a finite-difference forward velocity $\dot{x}_t = (x_t - x_{t-\Delta})/\Delta t$ with $\Delta t = 10$ ms and an exponential score $r_{\text{speed}} = 0.6\, e^{-3|\dot{x}_t - \dot{x}^\star|}$, and energy use is penalized by $r_{\text{torque}} = \frac{1}{1000}\sum_{j=1}^{7} |a_j|$.

Each foot carries four contact points. The gait phase ϕ is inferred from the vertical ground reaction forces F_L, F_R rather than from a clock,

$$\phi = \begin{cases} \text{double support}, & F_R \geq 10 \text{ N},\ F_L \geq 10 \text{ N}, \\ \text{right swing}, & F_R < 10 \text{ N},\ F_L \geq 10 \text{ N}, \\ \text{left swing}, & F_R \geq 10 \text{ N},\ F_L < 10 \text{ N}, \\ \text{flight}, & F_R < 10 \text{ N},\ F_L < 10 \text{ N}, \end{cases} \quad (10)$$

and the contact reward is keyed to it. Writing c_L, c_R for the active contact counts, x_L, x_R, z_L, z_R for foot positions, $z_{\text{plane}}(x)$ for the terrain height, σ for the logistic function, and taking 0.15 m as the desired swing clearance,

$$r_{\text{contact}} = \begin{cases} \sigma\big(2[(c_L + c_R) - 4]\big), & \text{double support}, \\ \tfrac{1}{2}\sigma\big(-20|z_R - z_{\text{plane}}(x_R) - 0.15|\big) + \tfrac{1}{2}\sigma\big(2(c_L - 2)\big), & \text{right swing}, \\ \tfrac{1}{2}\sigma\big(-20|z_L - z_{\text{plane}}(x_L) - 0.15|\big) + \tfrac{1}{2}\sigma\big(2(c_R - 2)\big), & \text{left swing}, \\ 0, & \text{flight}. \end{cases} \quad (11)$$

Double support rewards loading all eight contact points; each single-support phase rewards swing-foot clearance and stance-foot loading equally. Flight earns nothing, since a walking gait should not have one.

5.3 Training methodology

The policy is optimized with PPO [21] for 15 M environment steps under the hyperparameters of Table 4. The training domain randomizes ramp inclination over $[-5°, +5°]$ together with flat ground, and commanded speed over $[0, 2.0]$ m/s. Steeper slopes and all rough terrain are held out for evaluation, so every terrain result reported below is off-distribution.

Table 4 PPO hyperparameters. Both schedules decay linearly over training.

Parameter	Value
Learning rate	$3 \times 10^{-4} \rightarrow 1 \times 10^{-4}$
Entropy coefficient	$1 \times 10^{-3} \rightarrow 1 \times 10^{-4}$
Minibatch size	256
Rollout horizon	8192
PPO epochs per update	5
Clip range ε	0.15
Target KL	0.2
Total timesteps	15 M

Three policy architectures are trained under identical settings. Configuration 1 is two feedforward layers of 256 units; Configuration 2 prepends an LSTM layer of 64 units to a single feedforward layer of 256; Configuration 3 prepends the same LSTM to two feedforward layers of 256.

Three training techniques are used on top of PPO. *Reference-state initialization* [11] sets the robot at the start of each episode to a randomly chosen phase of the reference motion rather than always to the beginning of the cycle; without it the policy must learn the cycle sequentially and cannot make progress on the later part of the stride until it has mastered the earlier part. *Early termination* ends the episode as soon as the posture band of (9) is violated, which stops the policy accumulating reward while falling. *Imitation decay* relaxes the imitation weight over training,

$$\omega_{\text{im}}(t) = 1 - \alpha \frac{t}{T_{\text{tot}}}, \quad \omega_{\text{gait}}(t) = 1 + \alpha \frac{t}{T_{\text{tot}}}, \tag{12}$$

with $\alpha \in [0, 1)$, t the current training step, and T_{tot} the total number of training steps. The schedule keeps the human-like structure early and shifts priority to task performance later, which matters here because the commanded speed can exceed the 2.4 m/s ceiling of the corpus, where the representation is extrapolating. Without decay both weights stay fixed at 1.

5.4 Evaluation and results

Four scenarios probe different failure modes. *Speed matching* sweeps the commanded speed from 0 to 2.2 m/s on flat ground against the realized speed. The *ramp demonstration* sweeps ramp angle from $-15°$ to $+15°$ at those same speeds, three times the training range. The *noisy-plane demonstration* uses two stochastic terrains generated with the method of Hamzaçebi et al. [22]: the noise level ω scales the height perturbation from 1 to 20 cm, and the resolution $\gamma \in \{0.25, 0.5, 1.0, 2.0\}$ m sets how rapidly the surface varies. Finally, *velocity tracking* varies the commanded speed online within a single rollout.

Ablations toggle reference-state initialization and imitation decay. For comparison, two additional baselines are trained: a Covariance Matrix Adaptation estimation (CMA-ES) tuned SIMBICON controller [23], and a PPO baseline that keeps the reference states but drops r_{im}. Eight configurations are trained in total.

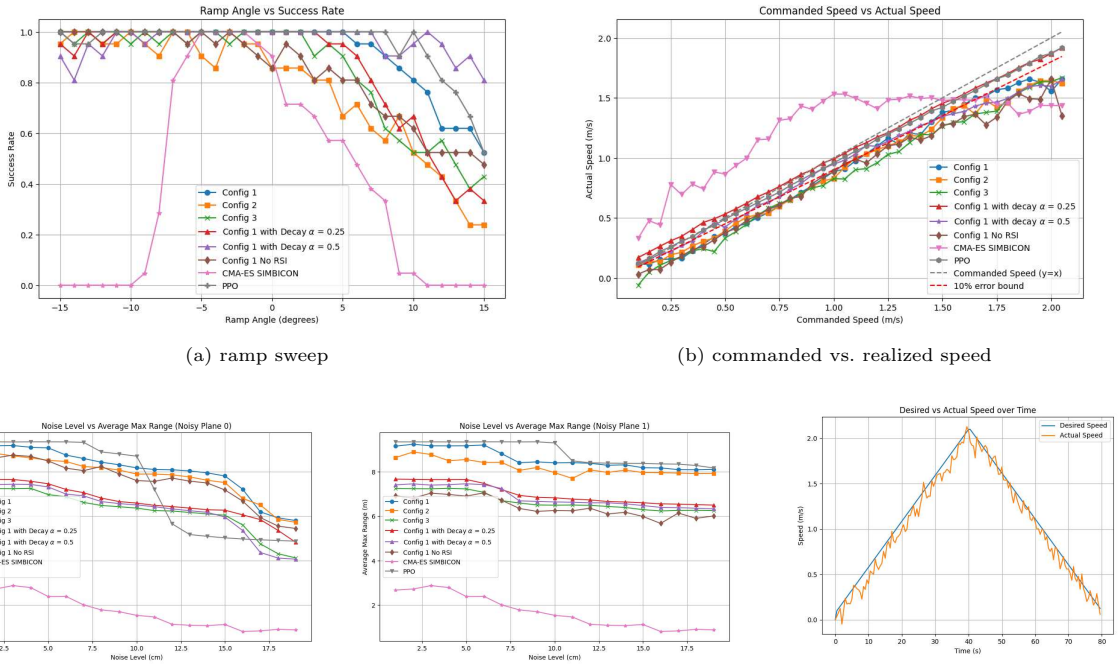
Imitation-guided configurations outperform both the PPO baseline and the CMA-ES SIMBICON controller across the aggregate metrics of Table 5. Imitation decay trades accuracy against adaptability: $\alpha = 0.25$ gives the lowest velocity-tracking MSE while $\alpha = 0.5$ gives the highest ramp success, and the fixed-weight Configuration 1 gives the best rough-terrain behaviour. Disabling reference-state initialization degrades every column.

On flat ground all PPO-based controllers track the commanded velocity inside a $\pm 10\%$ band up to roughly 1.3–1.5 m/s, with fixed-weight Configuration 1 holding that error to about 1.8 m/s, while the recurrent policies and the no-RSI ablation saturate early and CMA-ES SIMBICON does not reproduce the trend at all. Ramps are the sharpest test, because the reference is synthesized for level walking at every angle: success stays close to unity across the whole decline range, and on inclines the $\alpha = 0.5$ variant retains roughly 80% success at $+15°$, three times the training range, while CMA-ES SIMBICON fails consistently beyond $+5°$. On rough ground, unseen in training, range separates the configurations — Configuration 1 traverses farthest and degrades most gracefully, the recurrent policies stay competitive to about 8 cm of perturbation, and SIMBICON collapses at 2–3 m. Configuration 1 also follows an online speed command with a lag of only 0.1–0.2 m/s.

The recurrent policies perform worse across every test: 15 points lower ramp success and over a third less noisy-terrain success than the feedforward baseline. This is expected. A recurrent state exists to carry information from the past, most importantly gait phase; here, phase is already supplied explicitly by the preview terms of Table 3.

Table 5 Closed-loop performance across all eight configurations, averaged over commanded speeds, ramp angles, and noise amplitudes. NP0 and NP1 are independent noisy-terrain realizations; range is mean maximum traversal distance in metres. Best per column in bold.

Configuration	Vel. MSE ↓	Succ. (Rot.) ↑	Succ. (NP0) ↑	Succ. (NP1) ↑	Range (NP0) ↑	Range (NP1) ↑
Config. 1 (feedforward)	0.0309	0.923	**0.654**	**0.826**	**8.057**	8.586
Config. 2 (LSTM + 1 FF)	0.0372	0.776	0.416	0.501	7.779	8.204
Config. 3 (LSTM + 2 FF)	0.0521	0.843	0.455	0.562	6.226	6.665
Config. 1, decay $\alpha = 0.25$	**0.0036**	0.853	0.321	0.584	6.644	7.002
Config. 1, decay $\alpha = 0.5$	0.0290	**0.962**	0.396	0.598	6.297	6.847
Config. 1, no RSI	0.0604	0.839	0.346	0.531	7.599	6.409
CMA-ES SIMBICON	0.1548	0.401	0.163	0.163	1.693	1.693
PPO baseline (no r_{im})	0.0041	0.951	0.285	0.680	7.349	**8.869**



(a) ramp sweep

(b) commanded vs. realized speed

(c) traversal range, NP0

(d) traversal range, NP1

(e) online speed tracking

Fig. 5 Closed-loop behaviour. (a) Success rate against ramp angle, against a ±5° training range. (b) Commanded against realized speed with the ±10% band. (c), (d) Mean maximum traversal range against terrain noise amplitude on Noisy Planes 0 and 1. (e) Online speed variation within one episode.

Recurrence therefore adds optimization difficulty without adding new information.

Robustness scores do not show whether the motion still looks human. The joint trajectories in Fig. 6 prove that our controller stays on the motion-capture mean across every joint, while the PPO baseline, which keeps the reference states but drops the imitation reward, and the no-reference variant both diverge. Their gaits lose the smooth stance–swing transitions and the phase consistency of human motion, and SIMBICON oscillates at roughly twice the reference stride frequency. Remove either ingredient and the gait breaks, so both the imitation reward and the reference-state conditioning are load-bearing.

6 Discussion and Limitations

The two halves of the framework carry different limitations.

Everything is sagittal. Six joint channels and a 7-DoF planar plant exclude the lateral balance problem that dominates control on a real biped.

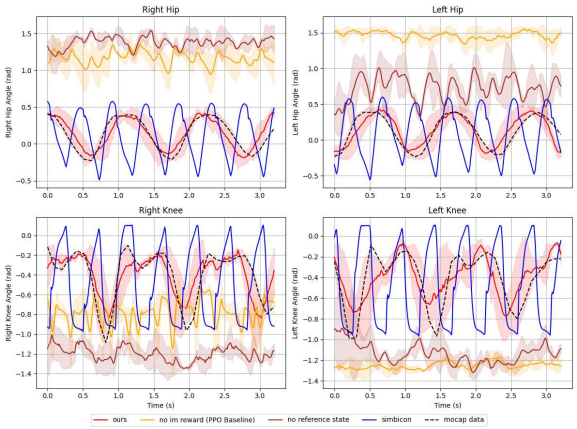


Fig. 6 Joint trajectories at 1.2 m/s for the imitation-guided controller, its ablations, SIMBICON, and motion capture. Curves are means over three seeds, shaded bands ±SD; the motion-capture curve is the mean of five trials at 1.195–1.205 m/s.

Speeds past 2.4 m/s and unusual limb geometries are extrapolation, a bound imposed by the regressor rather than the transform.

Periodicity is an assumption, not an approximation. Starts, stops, turns, and stumbles fall outside the domain rather than merely being harder, and gaits slower than 3.2 s would need a lower cutoff or more phase samples.

The reference is terrain-blind. Synthesized for level walking at every slope, it leaves terrain adaptation entirely to the policy — which is what makes the ramp and rough-ground results informative, but leaves a terrain-conditioned reference untested.

Validation is in simulation, and says nothing about actuator bandwidth, latency, or sensor noise on hardware.

7 Conclusion

This paper developed a two-stage framework for bipedal locomotion: a frequency-domain gait generator conditioned on walking speed and limb geometry, and a torque-controlled PPO policy that tracks its output under a blended imitation and task reward.

Treating a gait cycle as a spectrum rather than a sequence of poses yields a reference that is periodic by construction and fixed in size regardless of cycle duration. On held-out subjects it reconstructs human walking more faithfully than interpolation, playback, and oscillator-based alternatives at a matched budget. Placed inside the control loop of a planar biped, the same network that generates a compact frequency-domain representation drives policies to robust, human-like walking: they hold speed across the commanded range, outperform model-based and reward-only baselines, and stay upright on slopes and rough ground well outside the training distribution.

The contribution of this study is a compact frequency-domain neural network that generates gait patterns for a range of walking speeds and morphologies, together with its use as a reference signal generator in an imitation learning framework for a torque-controlled bipedal walking task, whose effectiveness is demonstrated across a variety of scenarios. Further studies will explore this framework on more complex humanoid models and in real-world hardware experiments.

Declarations

Funding. Not applicable. The authors did not receive support from any organization for the submitted work.

References

[1] Ren, L., Jones, R.K., Howard, D.: Predictive modelling of human walking over a complete gait cycle. Journal of Biomechanics **40**(7), 1567–1574 (2007)

[2] Chau, T.: A review of analytical techniques for gait data. part 2: neural network and wavelet methods. Gait & Posture **13**(2), 102–120 (2001)

[3] Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer, New York (2005)

[4] Deluzio, K.J., Astephen, J.L.: Biomechanical features of gait waveform data associated with knee osteoarthritis: An application of principal component analysis. Gait & Posture **25**(1), 86–93 (2007)

[5] Taga, G., Yamaguchi, Y., Shimizu, H.: Self-organized control of bipedal locomotion by

neural oscillators in unpredictable environment. Biological Cybernetics **65**(3), 147–159 (1991)

[6] Matsuoka, K.: Sustained oscillations generated by mutually inhibiting neurons with adaptation. Biological Cybernetics **52**(6), 367–376 (1985)

[7] Çatalbaş, B., Morgül, Ö.: Two-legged robot motion control with recurrent neural networks. Journal of Intelligent & Robotic Systems **104**(4), 59 (2022)

[8] Çatalbaş, B., Çatalbaş, B., Morgül, Ö.: System identification and model reference adaptive control of bipedal locomotion with neural networks. Robotics and Autonomous Systems **198**, 105331 (2026)

[9] Xie, Z., Berseth, G., Clary, P., Hurst, J., Panne, M.: Feedback control for cassie with deep reinforcement learning. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1241–1246 (2018). IEEE

[10] Itahashi, N., Itoh, H., Fukumoto, H., Wakuya, H.: Reinforcement learning of bipedal walking using a simple reference motion. Applied Sciences **14**(5), 1803 (2024)

[11] Peng, X.B., Abbeel, P., Levine, S., Panne, M.: DeepMimic: Example-guided deep reinforcement learning of physics-based character skills. ACM Transactions on Graphics **37**(4), 1–14 (2018)

[12] Peng, X.B., Coumans, E., Zhang, T., Lee, T.-W., Tan, J., Levine, S.: Learning agile robotic locomotion skills by imitating animals. arXiv preprint arXiv:2004.00784 (2020)

[13] Peng, X.B., Ma, Z., Abbeel, P., Levine, S., Kanazawa, A.: AMP: Adversarial motion priors for stylized physics-based character control. ACM Transactions on Graphics **40**(4), 1–20 (2021)

[14] Escontrela, A., Peng, X.B., Yu, W., Zhang, T., Iscen, A., Goldberg, K., Abbeel, P.: Adversarial motion priors make good substitutes for complex reward functions. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 25–32 (2022). IEEE

[15] Su, M., Huang, Y.: Multi-speed walking gait generation for bipedal robots based on reinforcement learning and human motion imitation. In: 2023 42nd Chinese Control Conference (CCC), pp. 4815–4821 (2023). IEEE

[16] Li, Z., Peng, X.B., Abbeel, P., Levine, S., Berseth, G., Sreenath, K.: Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control. The International Journal of Robotics Research **44**(5), 840–888 (2025)

[17] Schreiber, C., Moissenet, F.: A multimodal dataset of human gait at different walking speeds established on injury-free adult participants. Scientific Data **6**(1), 111 (2019)

[18] Yu, B., Gabriel, D., Noble, L., An, K.-N.: Estimate of the optimum cutoff frequency for the Butterworth low-pass digital filter. Journal of Applied Biomechanics **15**(3), 318–329 (1999)

[19] Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing **26**(1), 43–49 (1978)

[20] Yin, K., Loken, K., Panne, M.: SIMBICON: Simple biped locomotion control. ACM Transactions on Graphics **26**(3), 105 (2007)

[21] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)

[22] Hamzaçebi, H., Uyanik, I., Morgül, Ö.: On the analysis and control of a bipedal legged locomotion model via partial feedback linearization. Bioinspiration & Biomimetics **19**(5), 056004 (2024)

[23] Wang, J.M., Fleet, D.J., Hertzmann, A.:

Optimizing walking controllers. In: ACM SIGGRAPH Asia 2009 Papers, pp. 1–8 (2009)