

APRENDIZAJE POR REFUERZO INSPIRADO BIOLÓGICAMENTE PARA LOCOMOCIÓN: MPO + CPG

ADAN YUSSEFF DOMÍNGUEZ RUIZ, ÓSCAR LOYOLA
(DIRECCIÓN)

Introducción

La locomoción bípeda es un reto en robótica y biomecánica debido a la complejidad de mantener el equilibrio y estabilidad con alta complejidad dinámica y grados de libertad redundantes [1]. Métodos convencionales requieren cálculos complejos y constantes del centro de masa (CoM)[2], no son fácilmente adaptables a diferentes entornos y producen movimientos poco similares a la naturaleza[3]. Los Generadores Centrales de Patrones (CPGs) son modelos biológicamente inspirados que generan patrones rítmicos sin retroalimentación sensorial continua, permitiendo movimientos cílicos naturales y eficientes, como los observados en seres vivos[4], [5]. Han sido efectivos en robótica para controlar la locomoción y generar patrones en dispositivos protésicos[6]. Por otro lado, el Aprendizaje por Refuerzo (RL) optimiza políticas mediante la interacción con el entorno[7]. Integrar RL con CPGs mejora la suavidad, estabilidad y eficiencia del movimiento, proporcionando un enfoque robusto y adaptable para el control de locomoción bipeda[8].

Objetivos

- Desarrollar un marco de trabajo de RL+CPG para controlar la locomoción cíclica en diferentes entornos.
- Evaluar el movimiento cíclico generado y compararlo con otros métodos de RL, incluyendo SAC y MPO, en términos de velocidad, suavidad del movimiento, consumo energético y sincronización entre extremidades.
- Analizar la correlación cruzada entre las extremidades, midiendo la estabilidad y simetría del patrón de marcha generado por cada algoritmo.

Metodología

El experimento se llevó a cabo en el entorno simulado Walker-2d de Gymnasium, en el que se evalúa el desempeño de un robot bípedo en la ejecución de tareas de locomoción sobre un plano horizontal.

Variables involucradas:

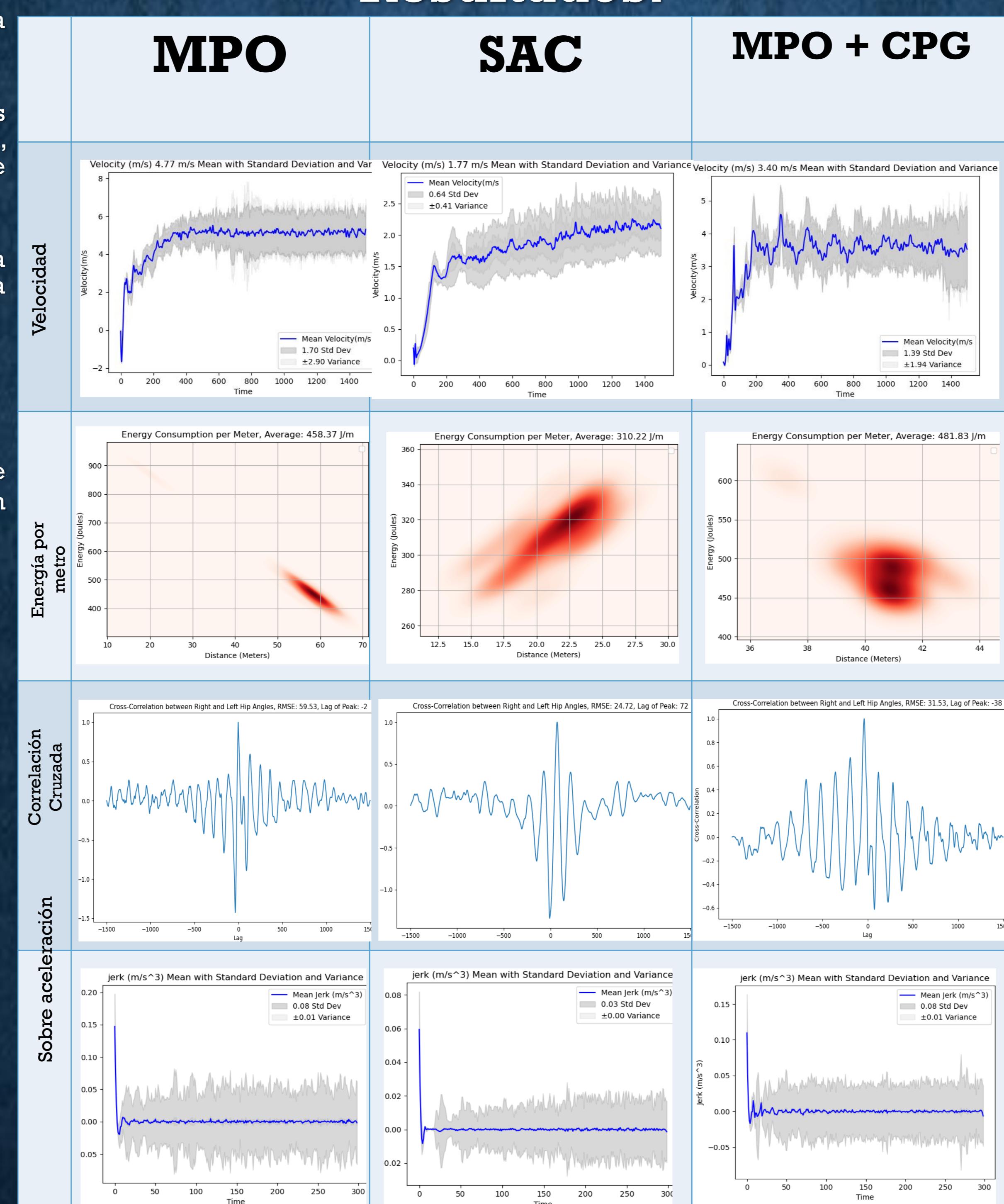
- Velocidad Media (m/s)
- Energía Consumida (J/m)
- Correlación Cruzada
- Jerk (Suavidad del Movimiento)
- Distancia Máxima Recorrida (m) s.

Metodología

Procedimiento:

- El agente se desplazó en línea recta hacia la derecha, y la recompensa fue otorgada únicamente cuando se lograba movimiento hacia adelante.
- Se realizaron 40 ejecuciones del experimento para cada enfoque (SAC, MPO y MPO + CPG), capturando solo los episodios donde el agente no caía.
- Los datos fueron recopilados en cada paso, incluyendo la velocidad en el eje X, el torque de cada articulación, la energía consumida y las aceleraciones.

Resultados:



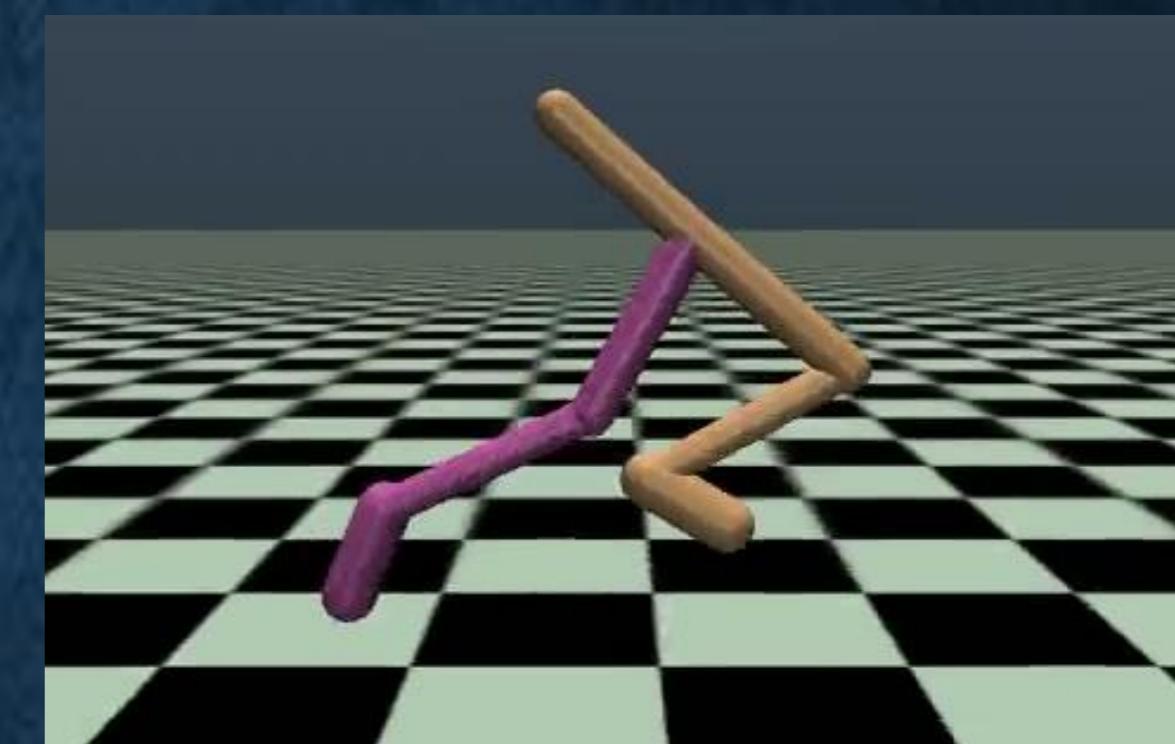
Trabajo a Futuro

Se propone validar los resultados obtenidos en simulaciones mediante pruebas experimentales en robots físicos, lo que permitirá comprobar la robustez y aplicabilidad del modelo en entornos reales. Otro aspecto a explorar es la inclusión de mecanismos de retroalimentación sensorial, que ayuden a adaptar dinámicamente el control del sistema a entornos reales, creando un entorno simulación/realidad para gemelos digitales.

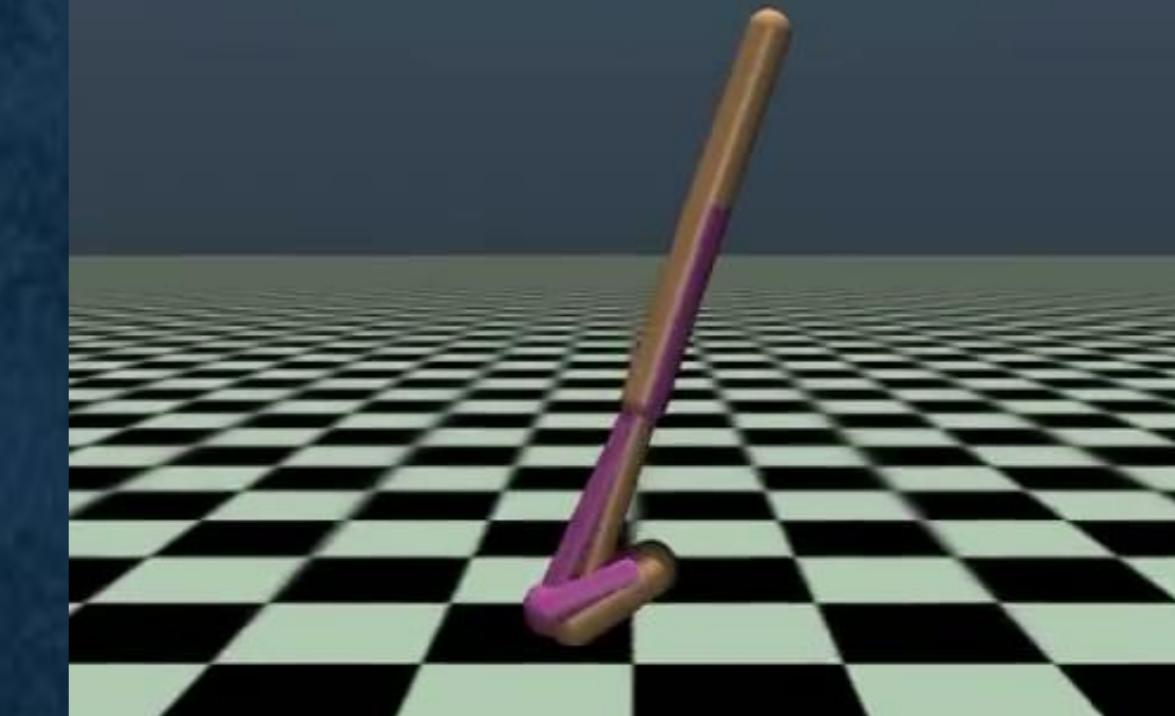
Finalmente, se buscará integrar modelos musculoesqueléticos más detallados y aumentar la complejidad de las tareas de locomoción, con la finalidad de avanzar hacia el desarrollo de prótesis inteligentes que no solo repliquen el movimiento, sino que también se adapten activamente a las necesidades del usuario.

Discusiones y Conclusiones

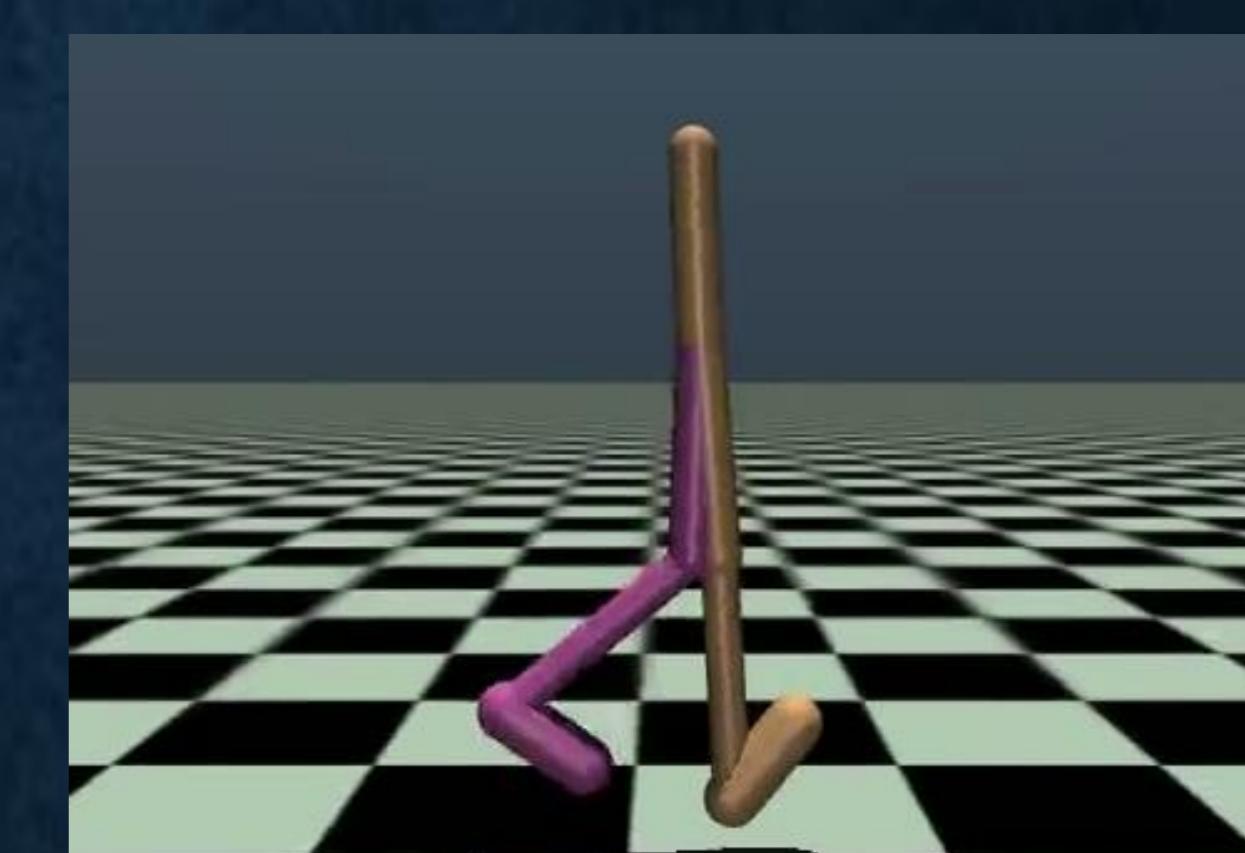
El algoritmo SAC mostró un patrón de movimiento inconsistente, caracterizado por saltos de ambas piernas simultáneamente, lo que resultó en una alta variabilidad en la distancia recorrida. Aunque alcanzó una velocidad promedio aceptable de 1.77 m/s, la eficiencia energética fue relativamente baja (310.22 J/m) y la suavidad del movimiento fue la más baja entre los enfoques, con un jerk de 0.03.



El enfoque MPO logró una velocidad promedio significativamente mayor (4.77 m/s), pero a costa de un mayor consumo energético (458.37 J/m). A pesar de la velocidad, el movimiento presentó picos pronunciados y fluctuaciones en la estabilidad, como lo demuestra un alto valor de jerk (0.08). Este enfoque favoreció la carrera sobre el caminado, lo cual aumentó la distancia recorrida, pero disminuyó la consistencia del movimiento.



El enfoque MPO + CPG mostró una clara ventaja en cuanto a la suavidad del movimiento. Aunque la velocidad fue intermedia (3.40 m/s), el patrón de marcha fue más consistente y energéticamente eficiente (481.83 J/m), con una menor variabilidad entre episodios. La correlación cruzada indicó un mejor sincronismo entre las piernas en comparación con los otros métodos, con un valor de LAG de -38, lo que demuestra un patrón cíclico más estable.



La combinación de MPO con CPG mostró una mejora significativa en la consistencia y estabilidad del movimiento, lo que indica que la integración de modelos biológicamente inspirados en controladores de RL puede optimizar la locomoción bípeda en entornos simulados. A pesar de que el enfoque MPO + CPG no alcanzó la mayor velocidad, su capacidad para generar movimientos más naturales y eficientes desde el punto de vista energético sugiere que es una estrategia prometedora para la robótica asistiva y el control de prótesis. La mejora en la suavidad del movimiento, medida mediante la reducción del jerk y la correlación cruzada, también valida la importancia de los CPGs en la generación de patrones rítmicos. Estos resultados ofrecen una base sólida para futuras investigaciones que integren modelos biológicos en algoritmos de control más complejos para dispositivos protésicos y exoesqueletos.

Referencias

- J. Reher and A. D. Ames, "Dynamic Walking: Toward Agile and Efficient Bipedal Robots," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. Volume 4, 2021, pp. 535–572, May 2021, doi: 10.1146/annurev-control-071020-045021.
[2] C. Kaymak, A. Ucar, and C. Guzelis, "Development of a New Robust Stable Walking Algorithm for a Humanoid Robot Using Deep Reinforcement Learning with Multi-Sensor Data Fusion," *Electronics*, vol. 12, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/electronics12030568.
[3] N. Heess et al., "Emergence of Locomotion Behaviours in Rich Environments," Jul. 10, 2017, arXiv: arXiv:1707.02286. Accessed: Sep. 06, 2024. [Online]. Available: http://arxiv.org/abs/1707.02286
[4] K. Matsuoka, "Mechanisms of frequency and pattern control in the neural rhythm generators," *Biol. Cybernetics*, vol. 56, no. 5–6, pp. 345–353, Jul. 1987, doi: 10.1007/BF00319514.
[5] A. J. Ijspeert, "Central pattern generators for locomotion control in animals and robots: A review," *Neural Networks*, vol. 21, no. 4, pp. 642–653, May 2008, doi: 10.1016/j.neunet.2008.03.014.
[6] Y. Wang, X. Xue, and B. Chen, "Matsuoka's CPG With Desired Rhythmic Signals for Adaptive Walking of Humanoid Robots," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 613–626, Feb. 2020, doi: 10.1109/TCYB.2018.2870145.
[7] L. De Vree and R. Carloni, "Deep Reinforcement Learning for Physics-Based Musculoskeletal Simulations of Healthy Subjects and Transformed Prostheses' Users During Normal Walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 607–618, 2021, doi: 10.1109/TNSRE.2021.3063015.
[8] G. Li, A. Ijspeert, and M. Hayashibe, "AI-CPG: Adaptive Imitated Central Pattern Generators for Bipedal Locomotion Learned Through Reinforced Reflex Neural Networks," *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–8, Jun. 2024, doi: 10.1109/LRA.2024.3388842.
[9] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," Aug. 08, 2018, arXiv: arXiv:1801.01290, doi: 10.48550/arXiv.1801.01290.
[10] A. Abdolmaleki, J. T. Springenberg, Y. Tassa, R. Munos, N. Heess, and M. Riedmiller, "Maximum a Posteriori Policy Optimisation," Jun. 14, 2018, arXiv: arXiv:1806.06920. Accessed: Aug. 30, 2023. [Online]. Available: http://arxiv.org/abs/1806.06920