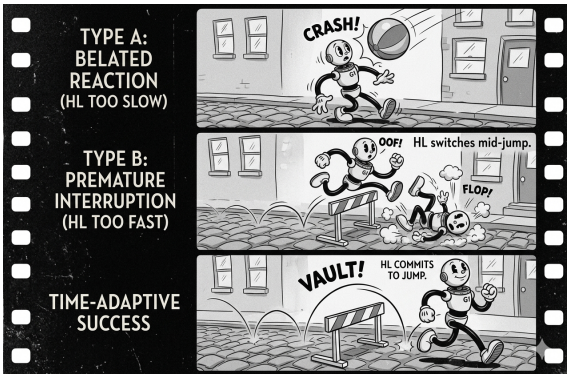
Commit or React: Time-Adaptive High-Level Control for Hierarchical Humanoid Motion Skills

Motivation:



In hierarchical humanoid control, a high-level (HL) policy selects from a library of pre-trained low-level (LL) motion skills—such as walking, jumping, or a tennis forehand—while each LL skill controls the robot at the simulation rate. This has been demonstrated for tennis [1], parkour [2], and martial arts [3] on the Unitree G1 [4].

A fundamental and largely overlooked problem is the **heterogeneity of skill execution durations**: a slip-recovery step lasts ∼150 ms; a tennis forehand takes ∼1.8 s; a jump arc lasts ∼800 ms. Existing HL controllers operate at a *fixed* decision frequency, creating two mutually incompatible failure modes:

- **Type A — Belated Reaction (HL too slow):** The decision interval exceeds the available reaction window. A perturbation cascades into a fall before a recovery skill fires; a ball arrives before the HL re-evaluates. The robot has the right skills—it simply cannot respond fast enough.
- **Type B — Premature Interruption (HL too fast):** The HL interrupts physically committed motions mid-execution. It switches to *walk* while the humanoid is airborne; a tennis swing is aborted at peak backswing. These failures look like policy quality problems but are structural timing problems.

We focuses on tasks that require both modes within the same episode. Sprinting toward a hurdle demands fast reactive initiation (∼200–300 ms window) followed by committed mid-air execution (switching mid-flight is catastrophic). *No single fixed HL frequency handles this correctly*—fast rates cause premature interruptions; slow rates miss reaction windows. This is the central problem this project addresses.

Goal:

We develop and validate a **time-adaptive high-level policy** for the Unitree G1. Rather than selecting only *which* skill to execute, the HL also learns *how long* to commit to it. Concretely, we aim to:

- Design and train an HL policy that outputs $(\text{skill}_{\text{id}}, \tau)$ tuples over a library of pre-trained G1 motion skills, where τ is the committed execution duration
- Demonstrate that this formulation resolves Type A, and B failure modes that are structurally unsolvable by any fixed-frequency baseline
- Validate the approach on a suite of athletic tasks in simulation, and deploy the time-adaptive controller on a physical Unitree G1

Approach:

Formulation. We formulate the HL policy as a Semi-Markov Decision Process (SMDP) [5]. At each decision point t_k, the policy outputs (z_k, τ_k), where $z_k \in \mathscr{Z}$ is a skill index and $\tau_k \in [\tau_{\min}, \tau_{\max}]$ is the committed duration. The LL skill π_{z_k} runs uninterrupted for τ_k; the HL is not queried until $t_{k+1} = t_k + \tau_k$. The SMDP return is:

$$R_k = \sum_{j=0}^{N_k-1} \gamma^j r_{t_k+j\Delta t}, \qquad y_k = R_k + \gamma^{N_k} V(s_{t_{k+1}})$$

where $N_k = \lfloor \tau_k / \Delta t \rfloor$. This adapts the TaCoS framework [6] to hierarchical control with a discrete pre-trained skill library. TARC [7] demonstrated physical transfer on quadrupeds and RC cars; this project extends it to humanoid skill hierarchies.

Duration parameterization. The HL actor outputs skill logits and a duration distribution $p(\tau_k \mid z_k, s_{t_k})$ as a squashed Gaussian over $[\tau_{\min}, \tau_{\max}]$, letting the policy learn that *jump* commits for $\sim$800 ms while *recovery* may terminate in $\sim$200 ms. Training uses PPO with SMDP-adjusted TD targets and separate entropy coefficients for each head.

HL–LL interface. Pre-trained LL policies are frozen and unmodified. A wrapper routes all simulation steps to π_{z_k} for duration τ_k; a short blending window (100–200 ms) interpolates joint configurations at skill boundaries.

Skill library from motion capture. LL skills are trained from retargeted motion capture using adversarial motion priors [8, 9]. The library draws from PHUMA [10] (92.7% sim-to-real transfer, native G1 support) and SMPLOlympics [11] (tennis, hurdling, parkour, martial arts), retargeted via GMR [12] where needed. Natural skill durations from MoCap clips initialize the HL duration bins.

General Details:

The student should bring along the following attributes:

1. Proficiency in Python and machine learning (PyTorch); experience with Isaac Gym / Isaac Lab is a plus.
2. Solid understanding of reinforcement learning, ideally including policy gradient methods (PPO) and reward design.
3. Familiarity with physics-based character animation or robot locomotion is strongly preferred.
4. Interest in hardware deployment; prior experience with real robots is a plus but not required.

Interested?

Reach out to Jin Cheng (jin.cheng@inf.ethz.ch) with your CV and transcripts.

References

[1] Zhikang Zhang, Jianyu Chen, et al. LATENT: Learning athletic humanoid tennis skills from imperfect human motion data. *arXiv preprint arXiv:2603.12686*, 2026. URL `https://arxiv.org/abs/2603.12686`.

[2] Ashish Kumar et al. Perceptive humanoid parkour: Chaining dynamic human skills via motion matching. *arXiv preprint arXiv:2602.15827*, 2026. URL `https://arxiv.org/abs/2602.15827`.

[3] TeleHuman Team. KungfuBot: Physics-based humanoid whole-body control for learning highly-dynamic skills. *arXiv preprint arXiv:2506.12851*, 2025. URL `https://arxiv.org/abs/2506.12851`.

[4] Unitree Robotics. Unitree G1 humanoid robot. `https://www.unitree.com/g1`. Accessed: 2026-04-01.

[5] Richard S. Sutton, Doina Precup, and Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1–2):181–211, 1999.

[6] Lenart Treven, Bhavya Sukhija, Yarden As, Florian Dörfler, and Andreas Krause. When to sense and control? a time-adaptive approach for continuous-time RL. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. URL `https://arxiv.org/abs/2406.01163`.

[7] Arnav Sukhija, Lenart Treven, Jin Cheng, Florian Dörfler, Stelian Coros, and Andreas Krause. TARC: Time-adaptive robotic control. *arXiv preprint arXiv:2510.23176*, 2025. URL https://arxiv.org/abs/2510.23176.

[8] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. AMP: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (SIGGRAPH)*, 40(4), 2021. URL https://arxiv.org/abs/2104.02180.

[9] Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. ASE: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Transactions on Graphics (SIGGRAPH)*, 41 (4), 2022. URL https://arxiv.org/abs/2205.01906.

[10] DAVIAN Robotics. PHUMA: Physically-grounded humanoid motion dataset. *arXiv preprint arXiv:2510.26236*, 2024. URL https://arxiv.org/abs/2510.26236.

[11] Zhengyi Luo, Jiashun Wang, Kangni Liu, Haotian Zhang, Chen Tessler, Zhiyuan Hu, Jingbo Wang, Ye Yuan, Jian Shi, Weinan Zhang, Kris Kitani, and Xue Bin Peng. SMPLOlympics: Sports environments for physically simulated humanoids. *arXiv preprint arXiv:2407.00187*, 2024. URL https://arxiv.org/abs/2407.00187.

[12] Yanjie Ze et al. GMR: General motion retargeting. 2026. URL https://github.com/YanjieZe/GMR.