Motor Learning Methods for Humanoid Control

Jun Morimoto
Dept . of Brain Robot Interface
ATR Computational Neuroscience Labs

Outline

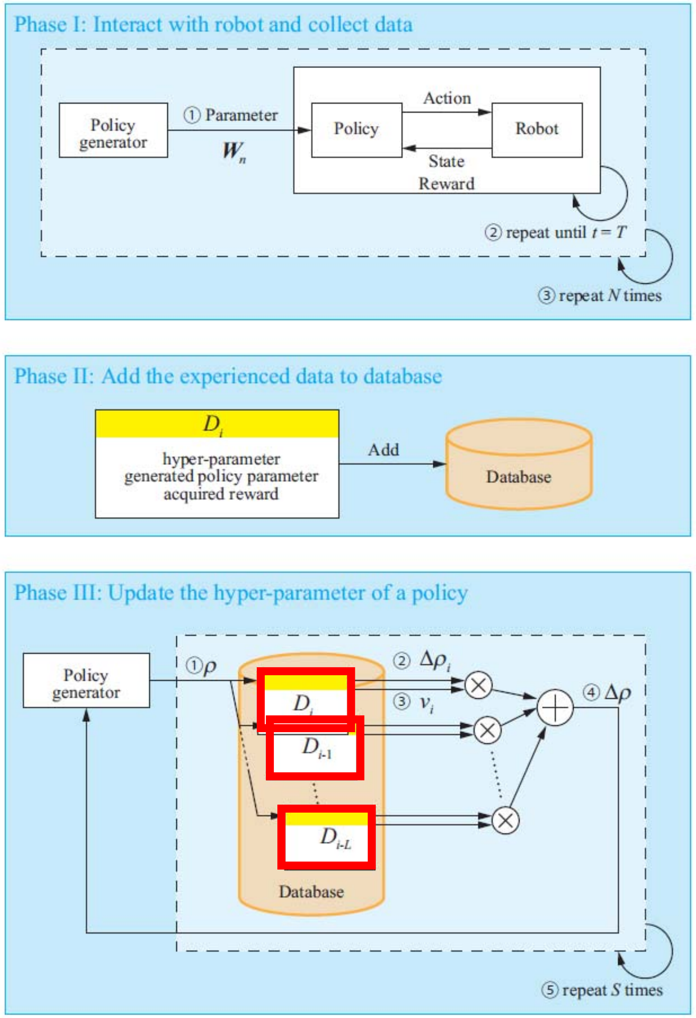
- Efficiently using collected data.
- Learning from demonstration.
- Hierarchical learning architecture.

Efficiently using collected data

Using Previous Experiences as Simulation Models

[Sugimoto, Sugiyama, Morimoto et al., IEEE Robot. Automat. Mag., 2016]

Dept. of Brain Robot Interface, ATR



Phase 1: Collect data from real environment using policy stochastically generated with policy generator (hyper-parameter).

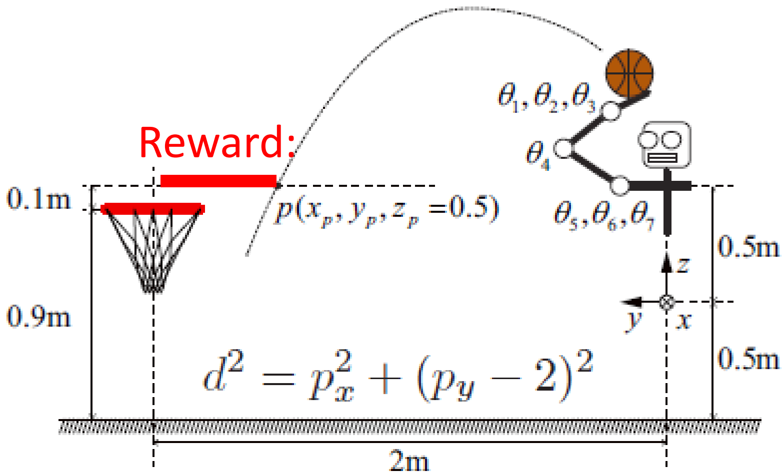
Phase 2: Add collected data to database.

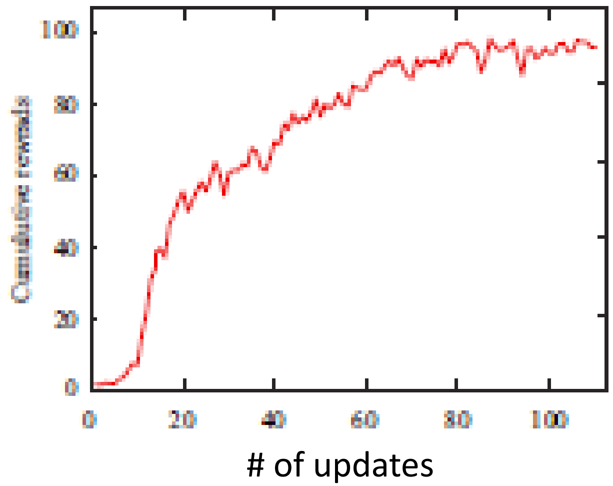
Phase 3: Update policy's hyper-parameter using stored data as simulation models.

Basketball shooting task

Basketball-shooting Task

Dept. of Brain Robot Interface, ATR





[Sugimoto, Sugiyama, Morimoto et al., IEEE Robot. Automat. Mag., 2016]

Policy model: $\underbrace{\mathbf{u}(t)}_{\text{Desired joint angle}} = \underbrace{\mathbf{w}}_{\text{Policy parameter}}{}^T \phi(t)$

Policy generator: $p(\mathbf{w} \mid \underbrace{\rho}_{\text{Hyper parameter}})$

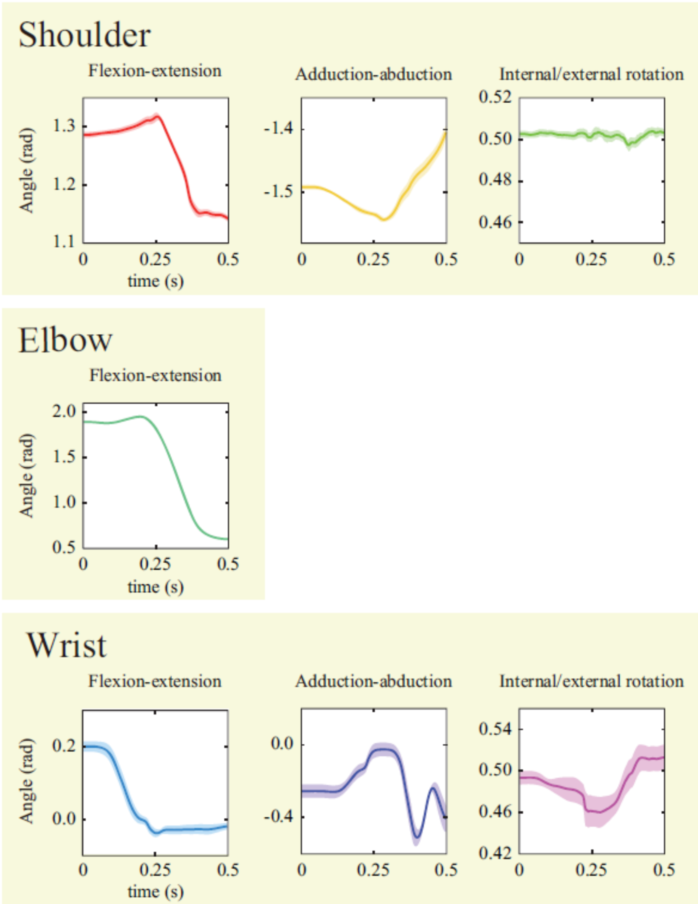
Gradient was estimated based on IW-PGPE. [T. Zhao, M. Sugiyama et al., Neural Computation, 2013]

Result

[Sugimoto, Sugiyama, Morimoto et al.,
IEEE Robot. Automat. Mag., 2016]

Dept. of Brain Robot Interface, ATR

Acquired Basket-ball Shooting Joint Trajectories

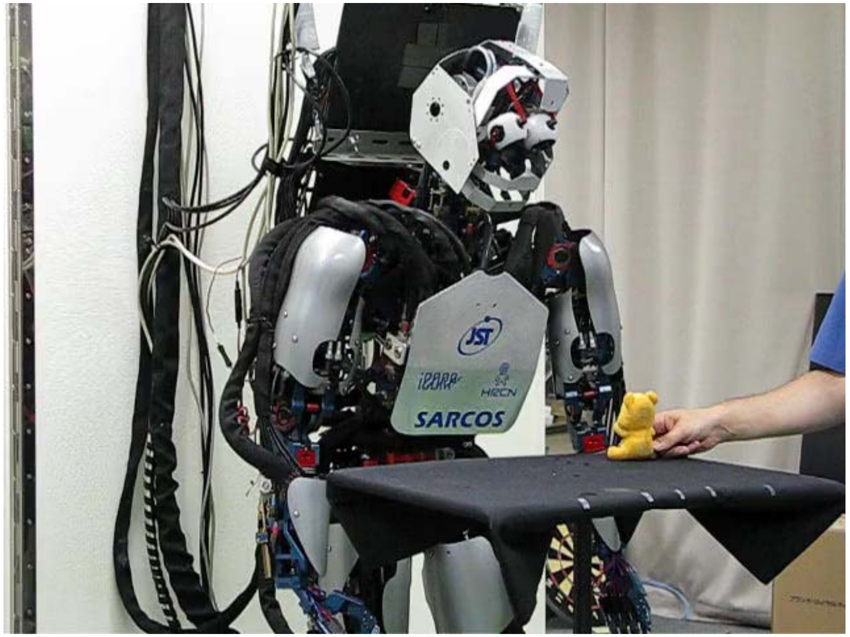


Learning from demonstration

Learning from Demonstration

Dept. of Brain Robot Interface, ATR

[Ude, Morimoto et al., IEEE Trans. on Robotics, 2010]



Learning from Demonstration

Dept. of Brain Robot Interface, ATR

Dynamic Movement Primitives

[Ijspeert et al., 2002]

[Ude, Morimoto et al., IEEE Trans. on Robotics, 2010]

Point attractor dynamics:

$$\tau \dot{z} = \alpha_z (\beta_z (g - y) - z) + \boxed{f(x)}$$

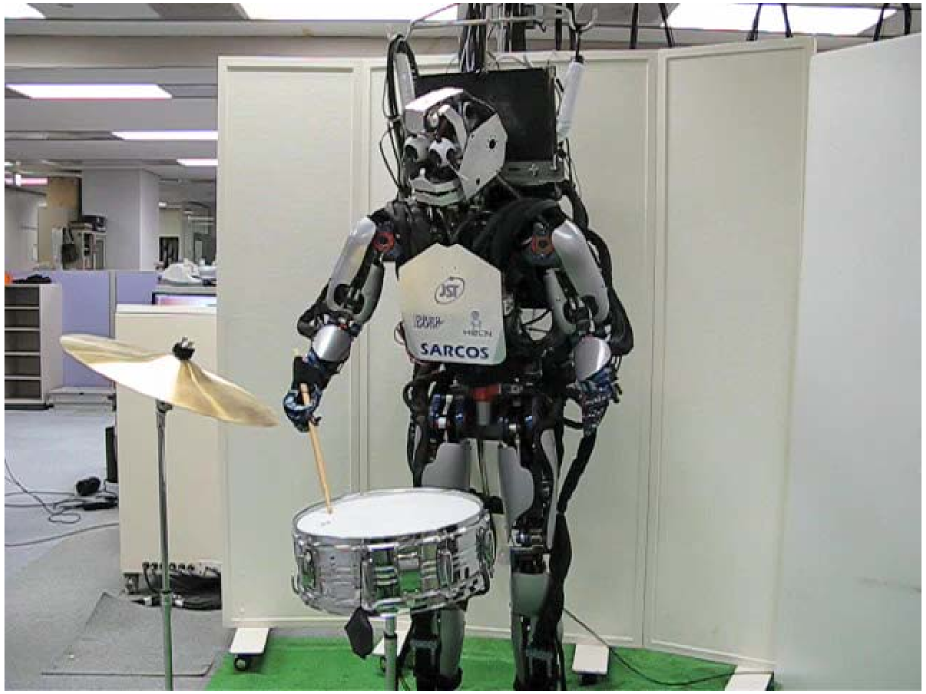
$$\tau \dot{y} = z$$

Phase-dependent modulation input:

$$\boxed{f(x)} = \frac{\sum_{i=1}^{N} w_i \Psi_i(x)}{\sum_{i=1}^{N} \Psi_i(x)} x, \ \Psi_i(x) = \exp\left(-h_i (x - c_i)^2\right)$$

Phase:

$$\tau \dot{x} = -\alpha_x x$$



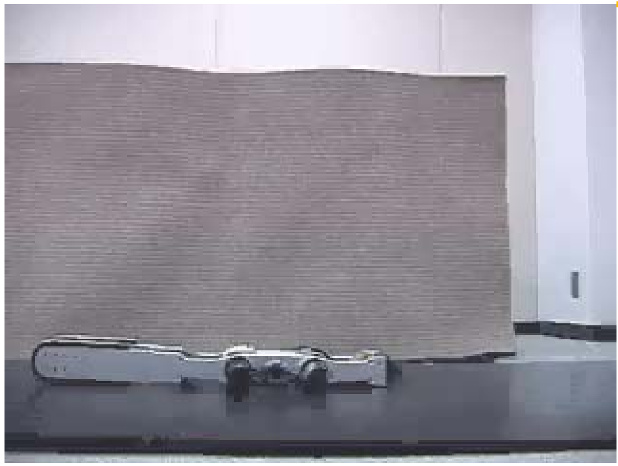
Hierarchical Reinforcement Learning: Application to Stand-up Movements

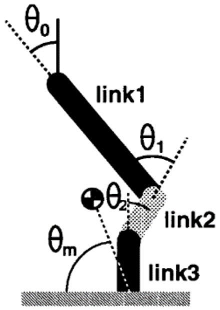
[Morimoto and Doya, 2001]

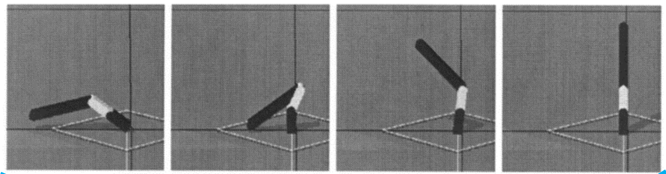
Dept. of Brain Robot Interface, ATR

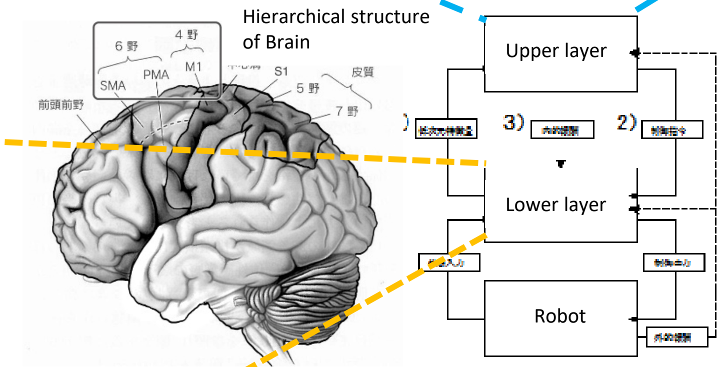
Learning trials

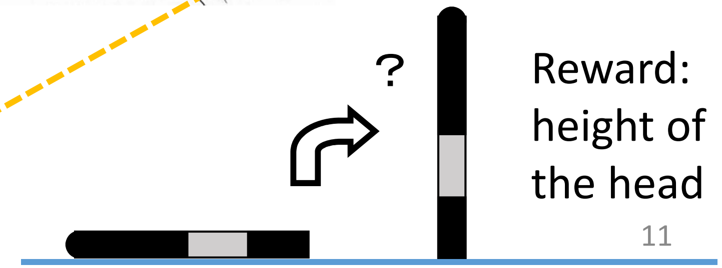
After learning trials











Reward: height of the head

Hierarchical learning architecture

Model Predictive Control (MPC)

Dept. of Brain Robot Interface, ATR

- Online derivation of the optimal control trajectory at each time step, using the first control output.

- Although each optimal control trajectory provides feedforward controller, MPC effectively works as *feedback* control policy due to the optimal control trajectory calculation at each time step.

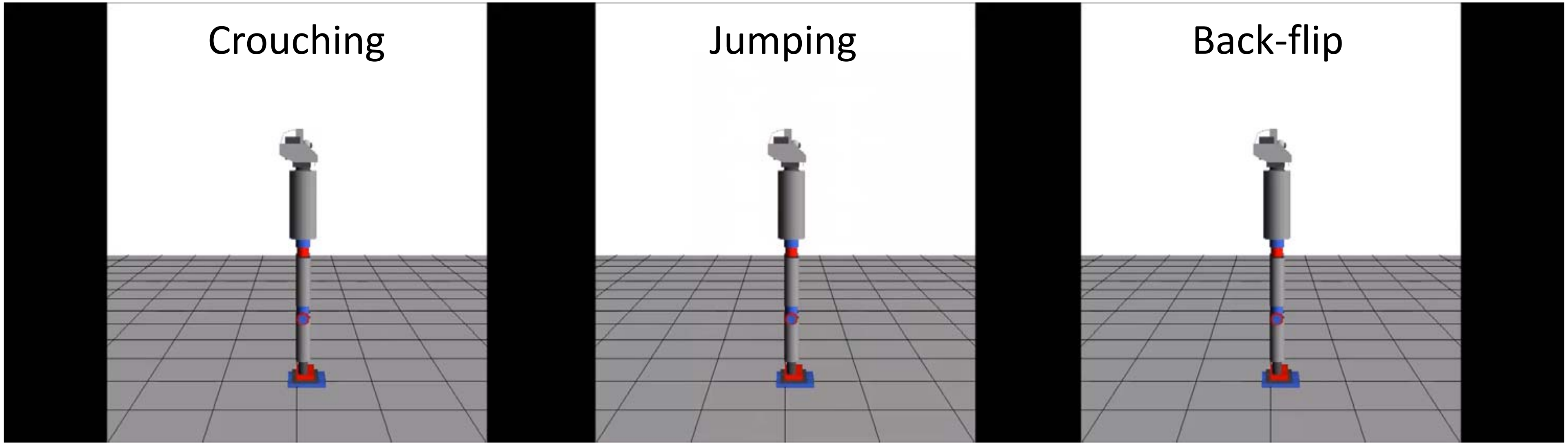
State
$\boldsymbol{x}_k$

$$\min_{\boxed{\boldsymbol{u}_k}, \cdots, \boldsymbol{u}_{N+k-1}} J_k = \sum_{i=k}^{N+k-1} \underbrace{r(\boldsymbol{x}_i, \boldsymbol{u}_i)}_{\text{Immediate cost}} + \underbrace{\Phi(\boldsymbol{x}_{N+k})}_{\text{Terminal cost}}$$

Optimized
control input
$\boxed{\boldsymbol{u}_k^*}$

MPC Can be Used to Generate Variety of Motions

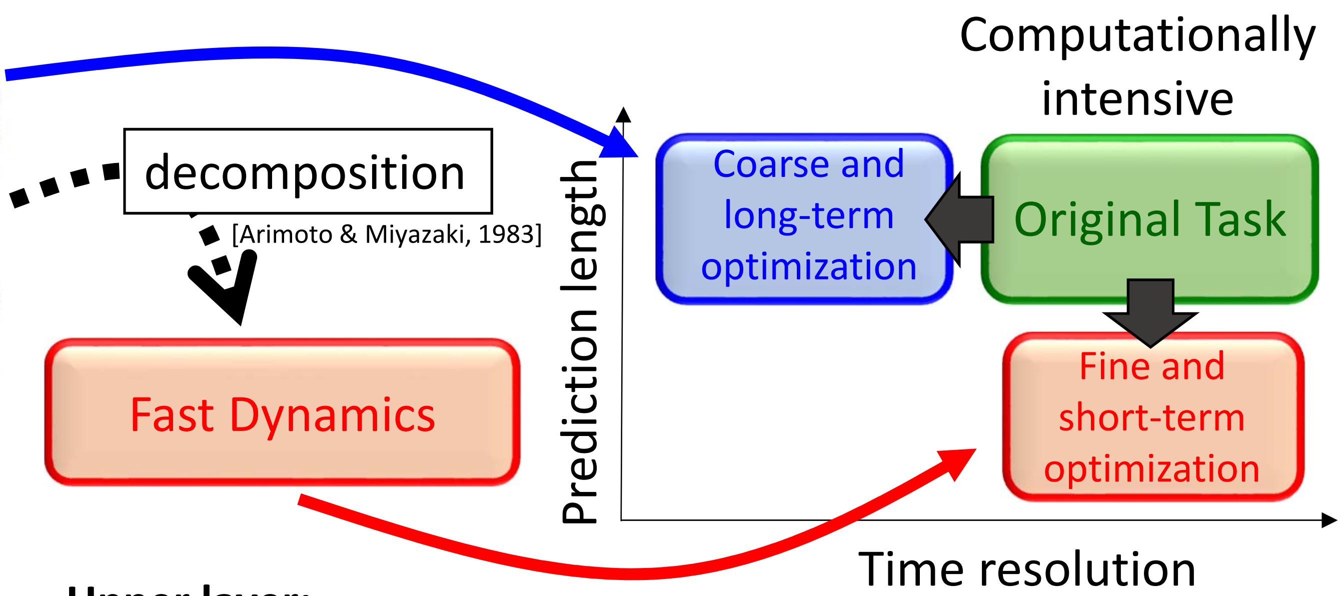
Dept. of Brain Robot Interface, ATR



- Optimal control trajectories derived at each time step: MPC is computationally intensive and not applicable to real-time/many-DoF robot control (e.g. humanoids) to generate fast and dynamic movements.

Hierarchical MPC Strategy with Singular Perturbed System

[Ishihara and Morimoto, Humanoids, 2015]



Upper layer:

- Longer time horizon with larger time step in MPC framework.
- Derive movement trajectory that involves longer-term effect but coarse planning in terms of control frequency.

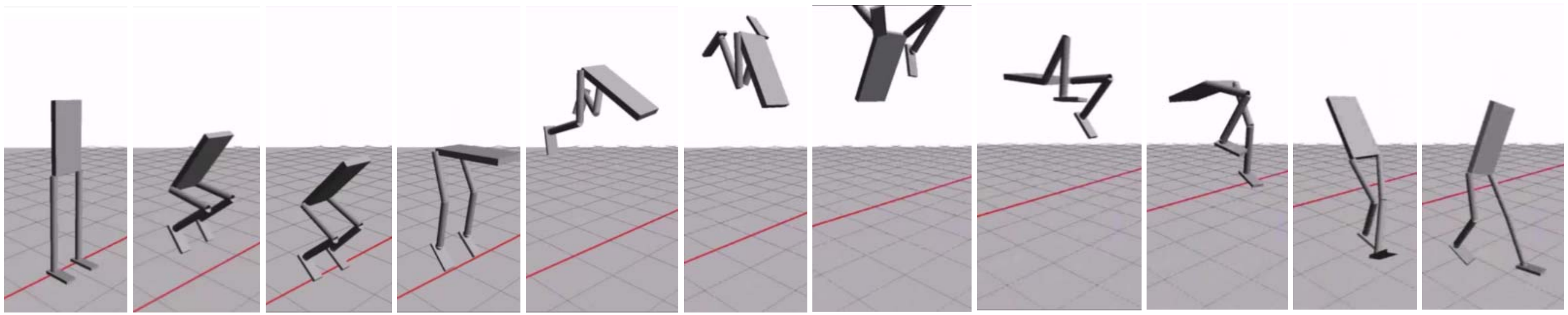
Lower layer:

- Shorter time horizon with smaller time step in MPC framework.
- Derive movement trajectory that involves shorter-term effect but fine planning in terms of control frequency.

Toward Agile Movement Like Human Experts

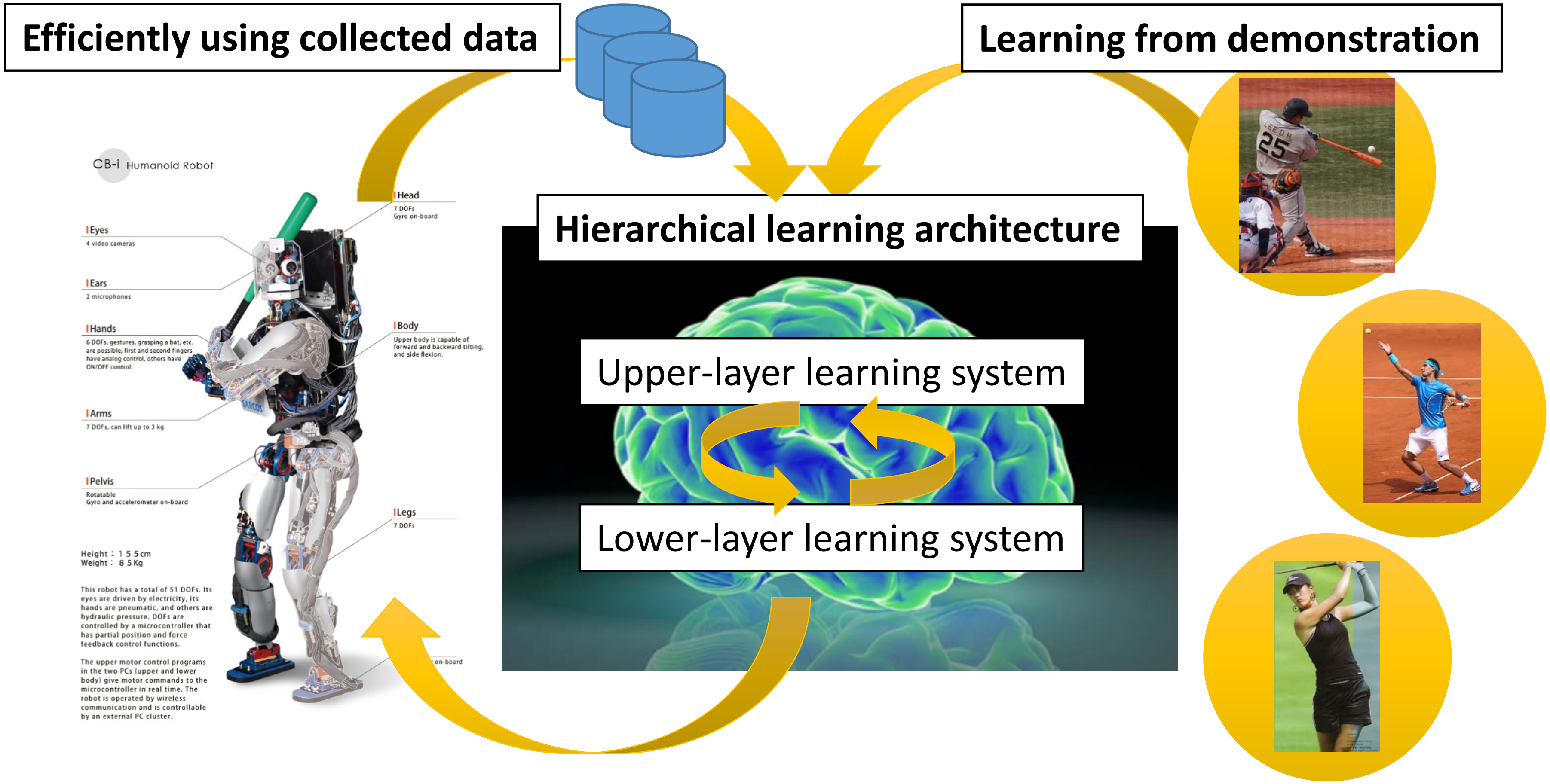
Dept. of Brain Robot Interface, ATR

Generated front flip movement by using the proposed method:



Discussion:
Stable Interactions between Different Learning Systems

Dept. of Brain Robot Interface, ATR



This study was supported by NEDO and JSPS KAKENHI.