University of Tsukuba
筑波大学

Toward Coaching a Robot by Utilizing Bioelectrical Signals

Kenji Suzuki
University of Tsukuba

Imitation and Coaching in Humanoid Robots Workshop
December 1, 2008

IEEE/RAS Humanoids 2008
Daejeon, Korea

Outline
Coaching a robot by utilizing bioelectrical signals

- An approach to coach a robot
- Toward coaching by utilizing bioelectrical signals
- A Case-study: A small humanoid experiment
  - Balancing and walking experiment
- A model of learning from success and failure
  - Formulation
  - Simulation experiments
- Conclusion

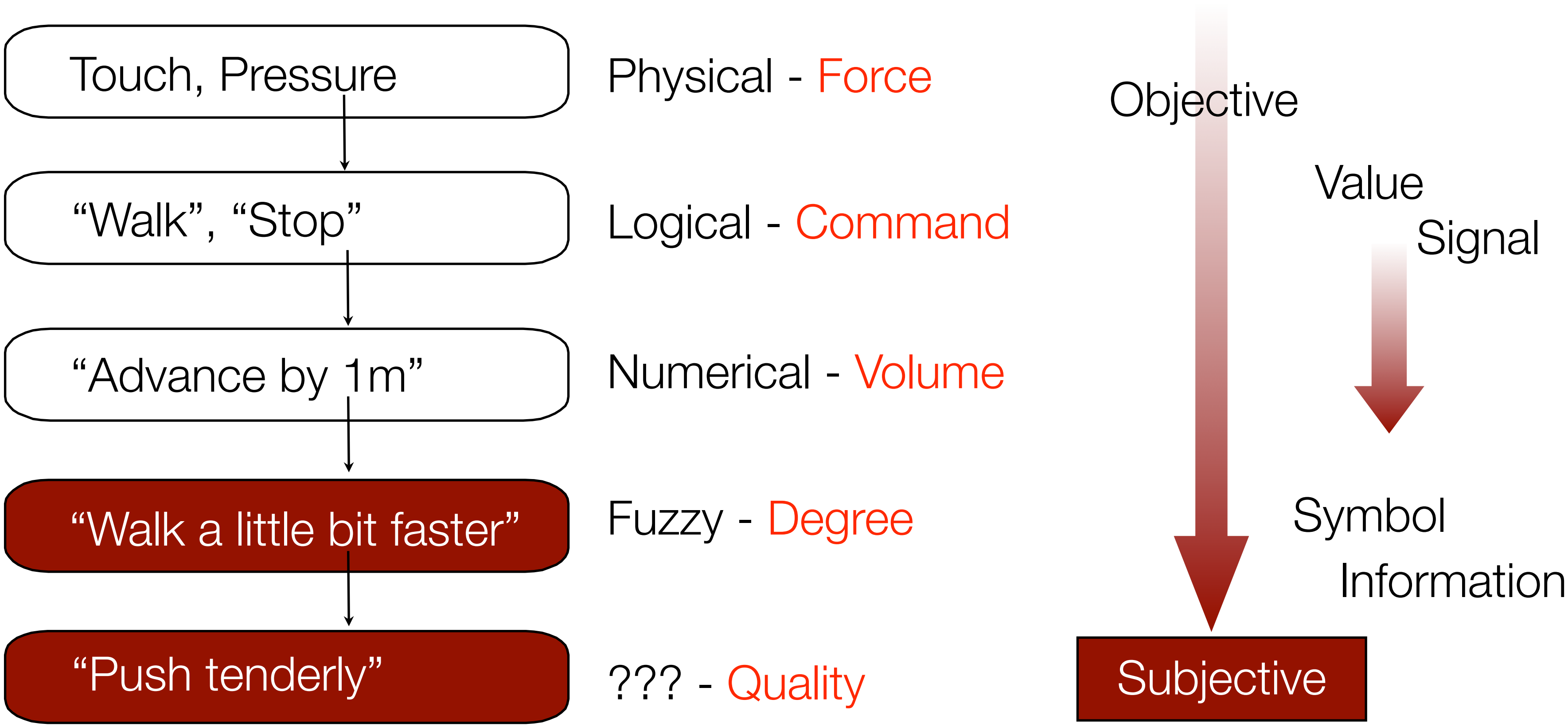
Coaching a robot

- Provide an **intuitive and interactive ways** to program robot behaviors, especially for non-experts
  - Coaching as a human-robot interaction
  - Guidance/instruction as minimum as possible
- Representation for **faster learning**
  - Decrease the learning iterations to accomplish the desired control task
- Utilizing human knowledge and subjective evaluation

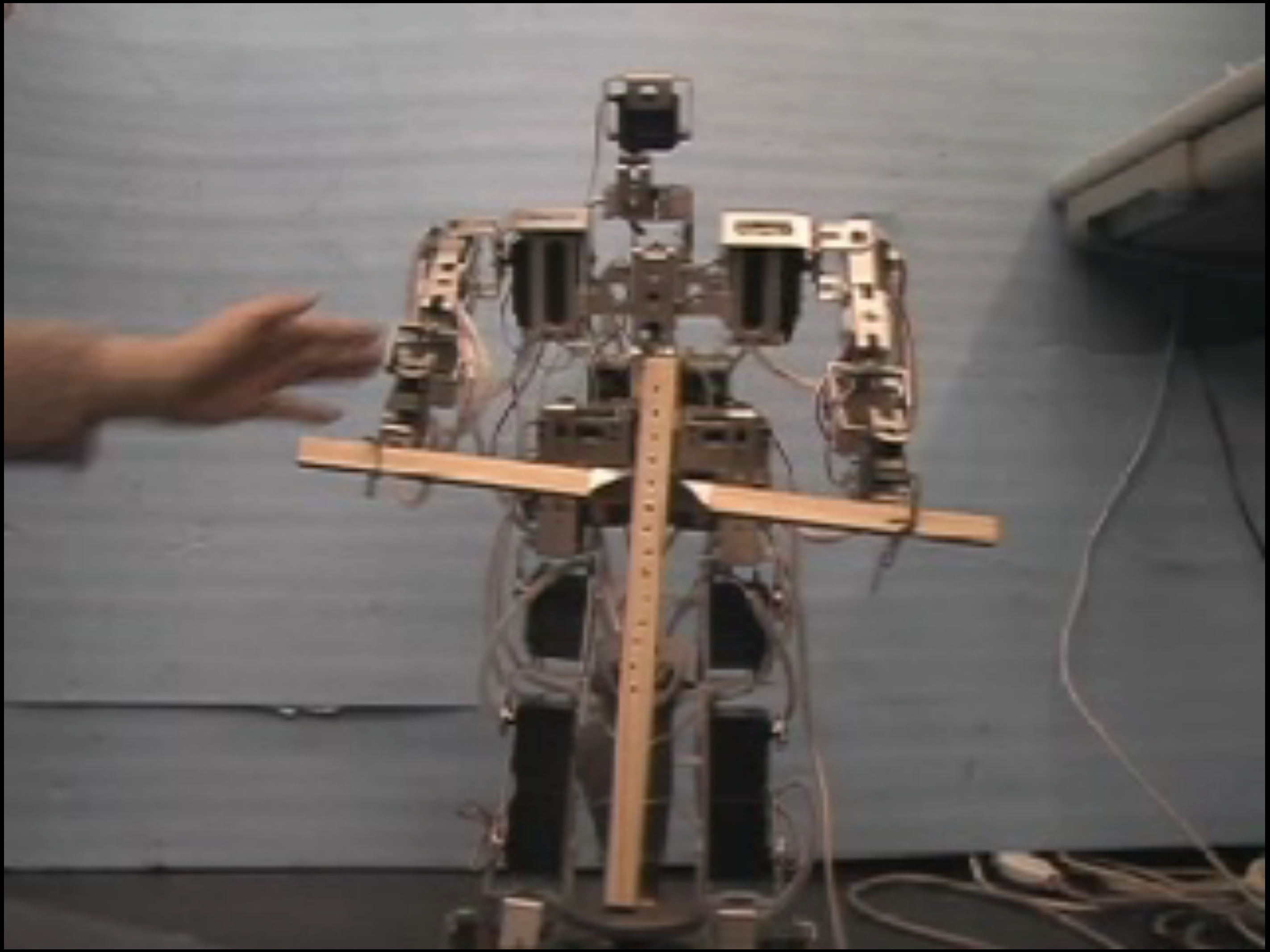
Related Works

- Direct teaching / Practical applications
- **Interactive Evolutionary Computation**: Fusion of the Capabilities of EC Optimization and Human Evaluation, [Takagi, Proc. of IEEE, 89, 9, 2001]
- **Subjective-Evaluation Oriented Teaching Scheme** [Humanoids2003]
- **Coaching:** An Approach to Efficiently and Intuitively Create Humanoid Robot Behaviors [Riley, M., Humanoids2006]

Commands to the machine/robot



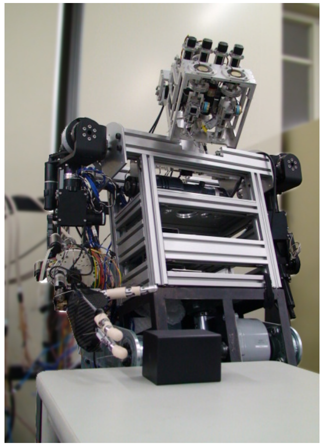
Limited few dimensions
"Good" or "Poor"

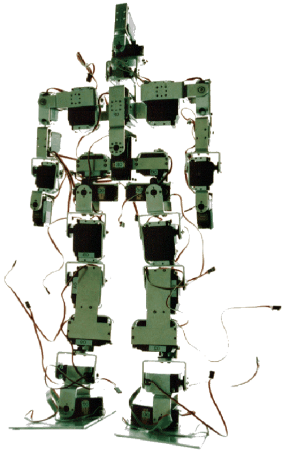


A small humanoid to keep balancing

Potential Applications

- **Control tasks for humanoid robot**
  - High DoF and redundant manipulator
  - Human-like and daily-life tasks
  - Task with no explicit evaluation function

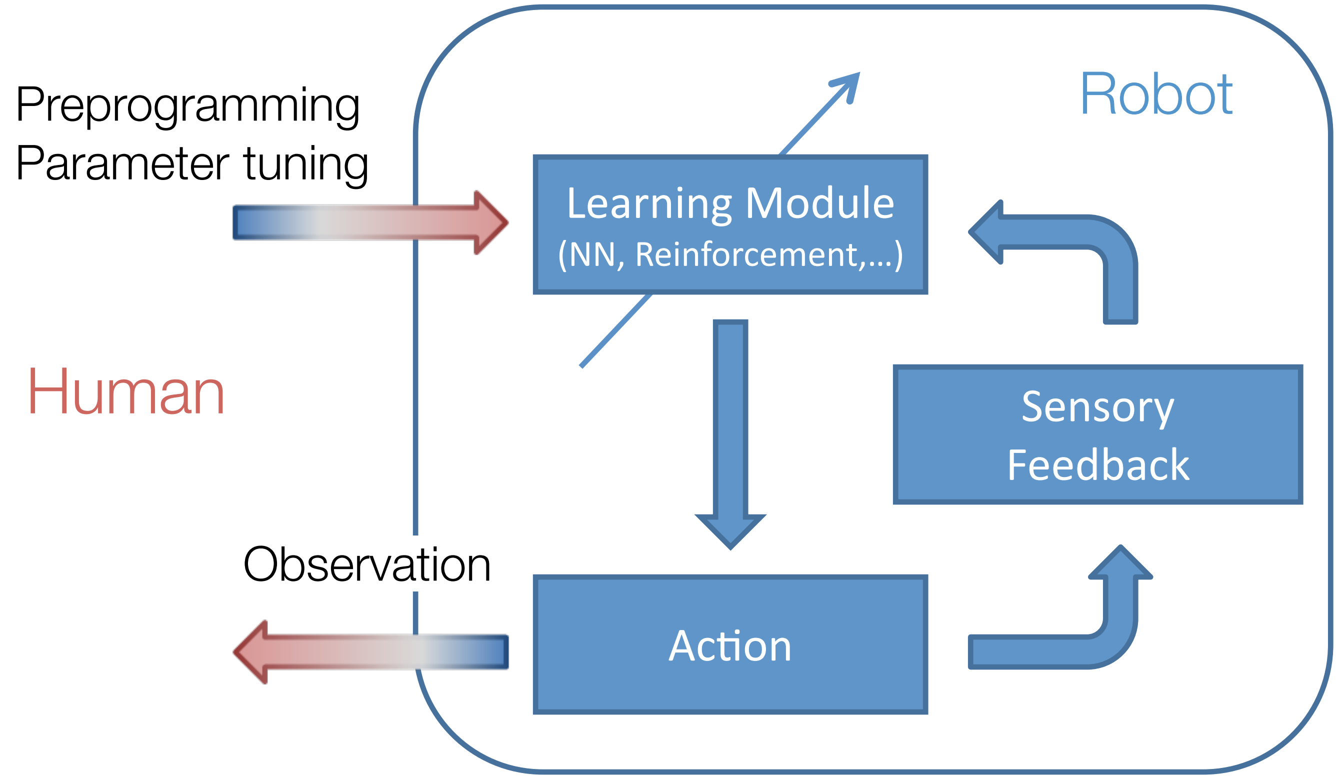




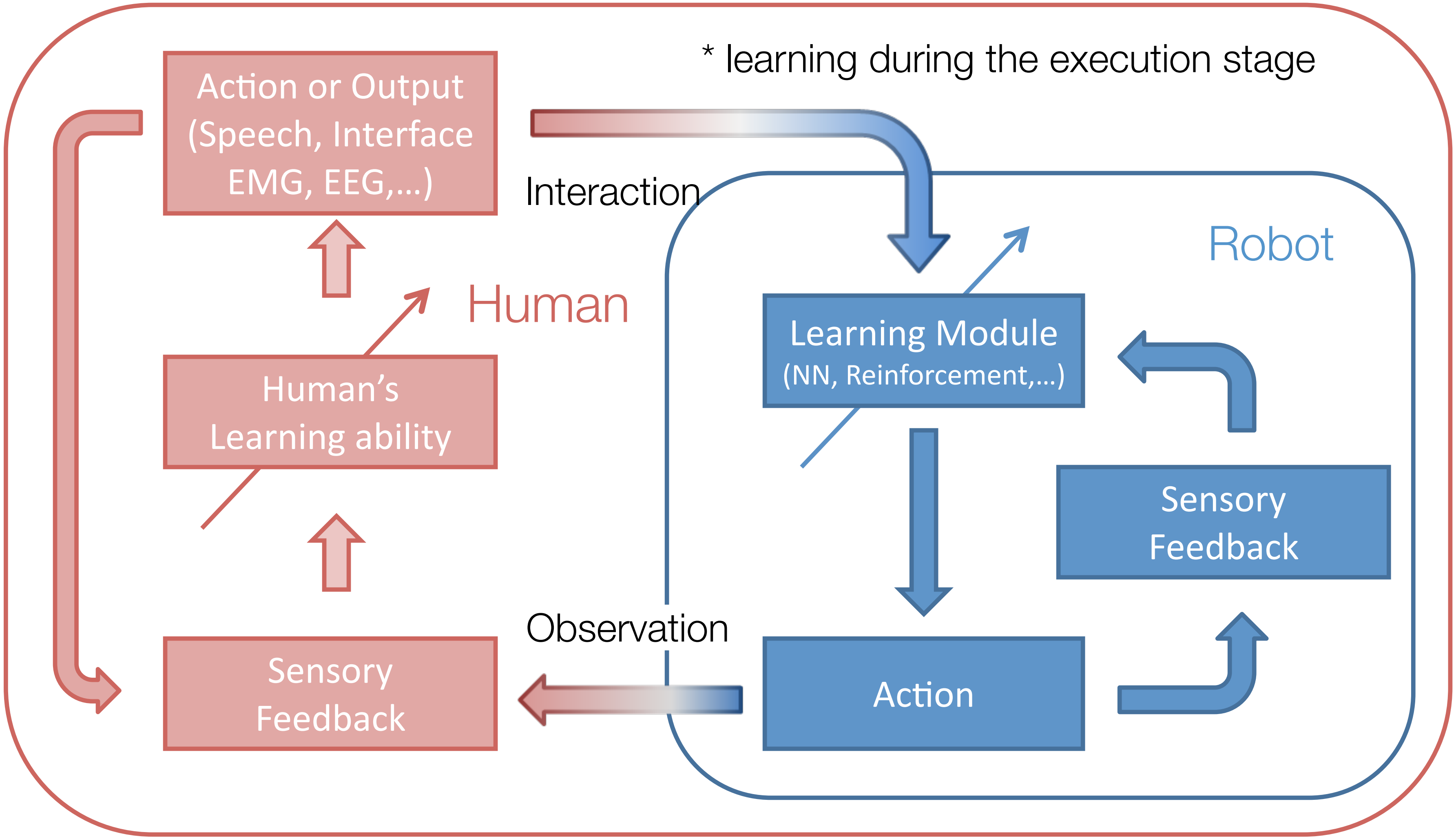
- **Wearable / rehabilitation robot**
  - Individual difference
  - Based on subjective evaluation
  - Adaptation to the interpersonal timing of control

Coaching as a human-robot interaction

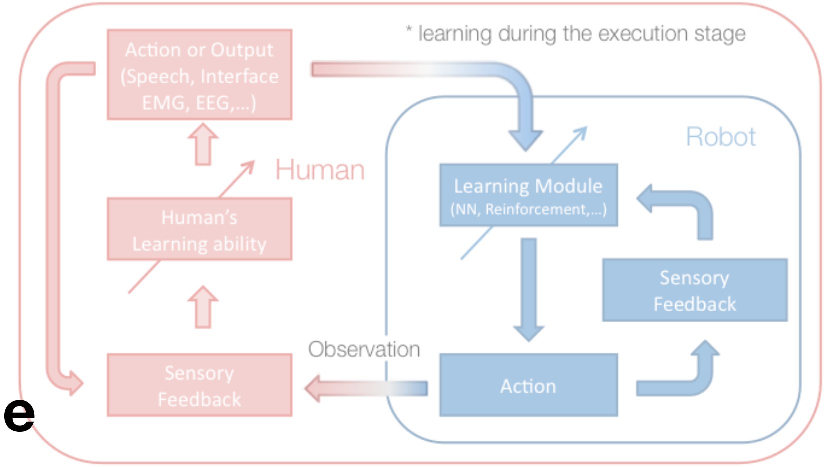
Traditional control/learning of the control target



Coaching as a human-robot interaction

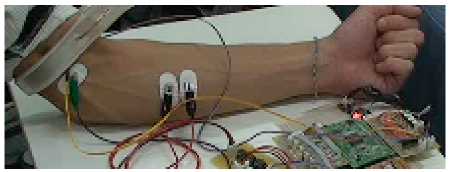


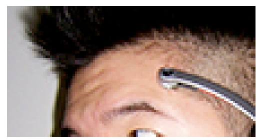
Important Factors



- **Good and/or Poor Learning**
  - Instruction **as simple as possible**
  - Humans are likely to give his/her **subjective evaluation**
  - Give a **quick and intuitive** evaluation to the control system
  - **Robot trainer - Everyone**
- **Feedback during the action execution**
  - Understanding human intentions that are often ambiguous and difficult to be quantified
  - Temporal continuity and contingency
  - Coaching based on **bioelectrical signals**

Why bioelectrical signals?





1ch EEG

EMG

EEG+fNIRS

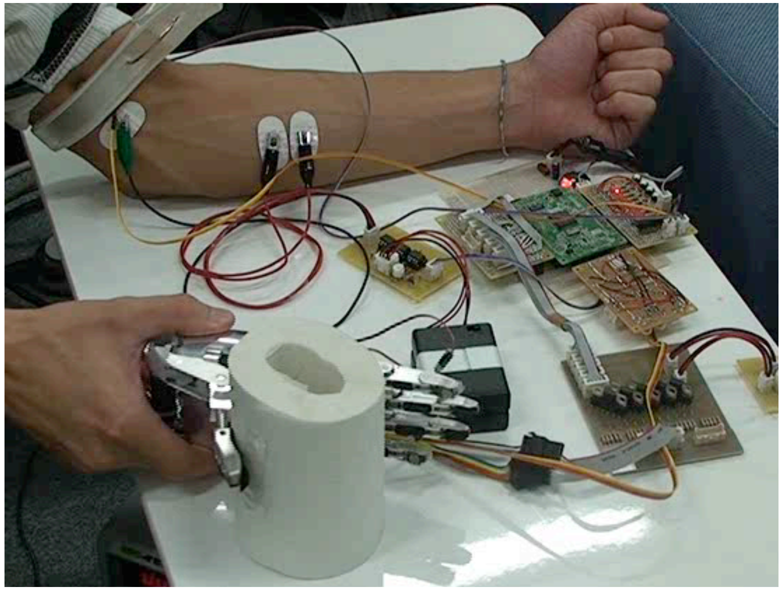
fNIRS

Fast response ⟷ Low response

- **Quantitative measurement** of the activation
- Robust recognition: Independent of location, environment, noise/lighting conditions, compared to traditional interfaces
- **Obtrusive** and **individual calibration**

Bioelectrical signal processing

- **Response time**
  - EMG or EEG related signals
  - **Quite limited** channels
  - Enough quality compared to vision or speech commands
- **Considerations**
  - Electrode locations
  - Individual differences
  - Simplified wearable device
    - EMG or EEG



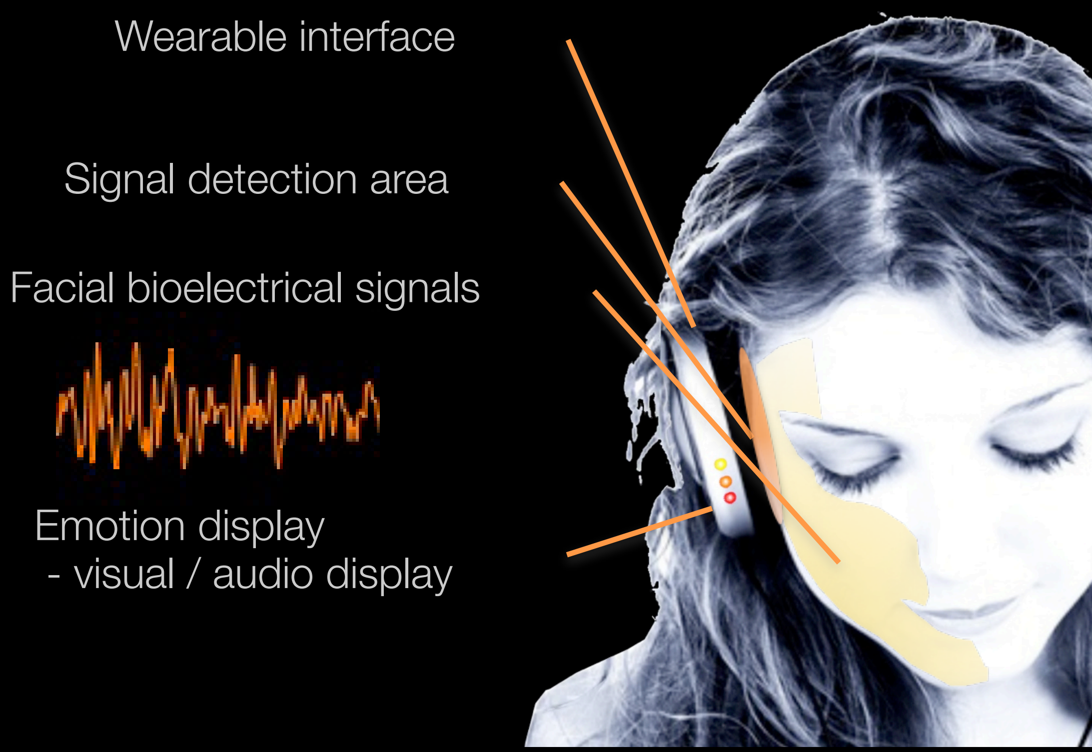
Blind control of robot hand

(No visual but tactile feedbacks are provided)

Not trying to understand natural intention but utilizing it

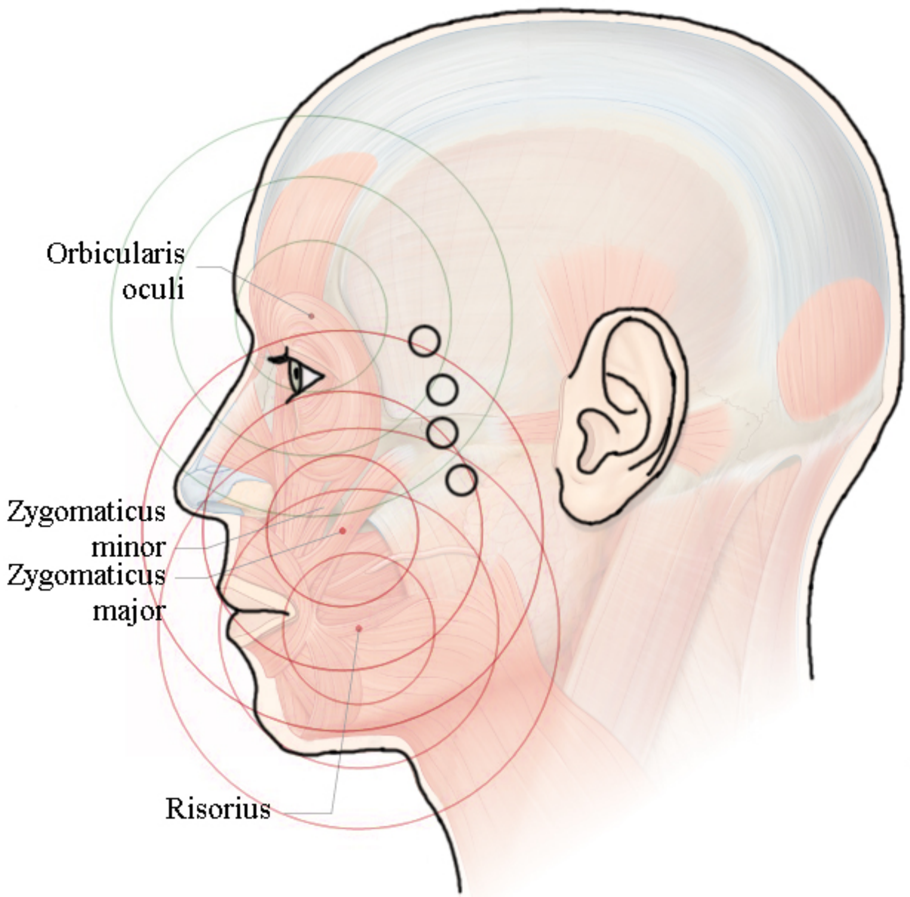
Emotion Reading

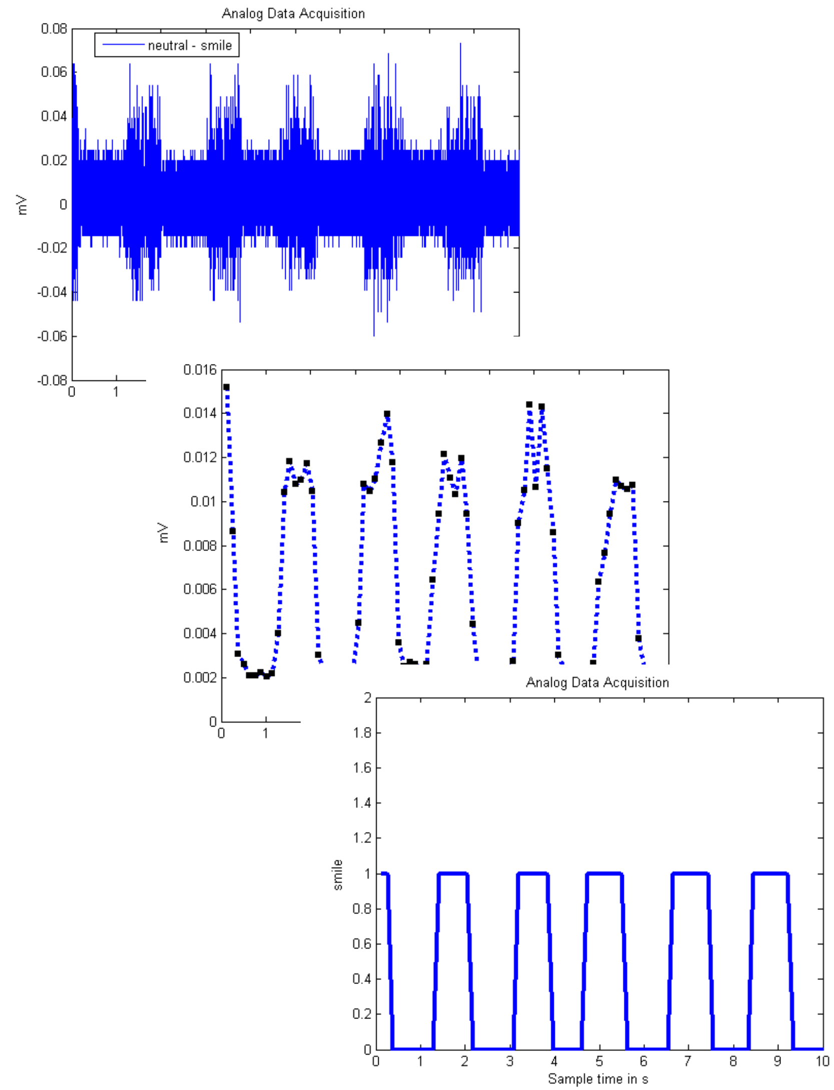
Members: Anna Gruebler, Kenji Suzuki



- Therapeutic Application
- Communication aid
- Facial expression training
- Games or entertainment
- Rehabilitation
- Quality of life

Cybernics
fusion of human, machine and information systems



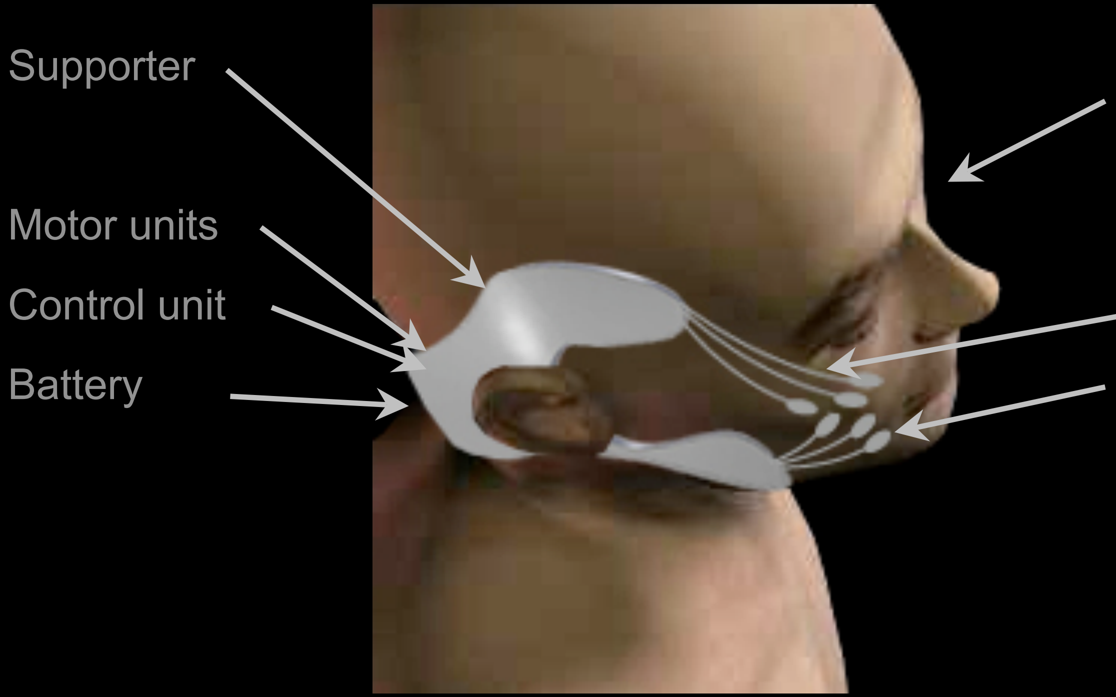


Robot Mask

Members: Dushyantha Jayatilake, Kenji Suzuki

Cybernics
fusion of human, machine and information systems

SHISEIDO



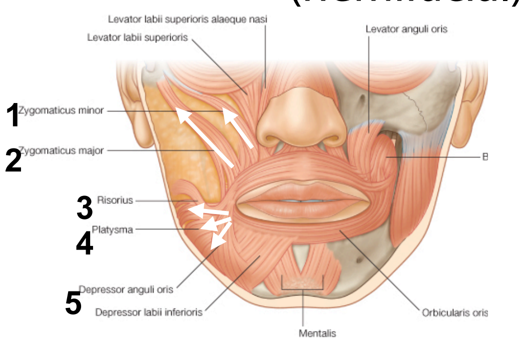
Robot Mask
To support expressive functions of the face

"A Soft Actuator Based Expressive Mask for Facial Paralyzed Patients," **IEEE IROS 2008**, Nice, France, 2008.

~~asymmetric~~

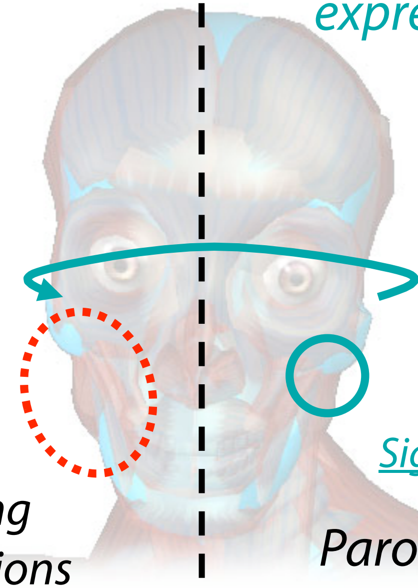
Symmetrical expression

one-side facial paralysis (hemifacial)



© Elsevier. Drake et al: Gray's Anatomy for Students - www.studentconsult.com
Oral group of facial muscles.

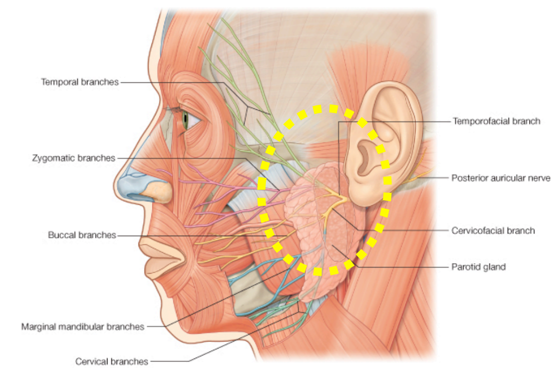
(1) Zygomaticus minor
(2) Zygomaticus major
(3) Risorius
(4) Platysma
(5) Depressor anguli oris



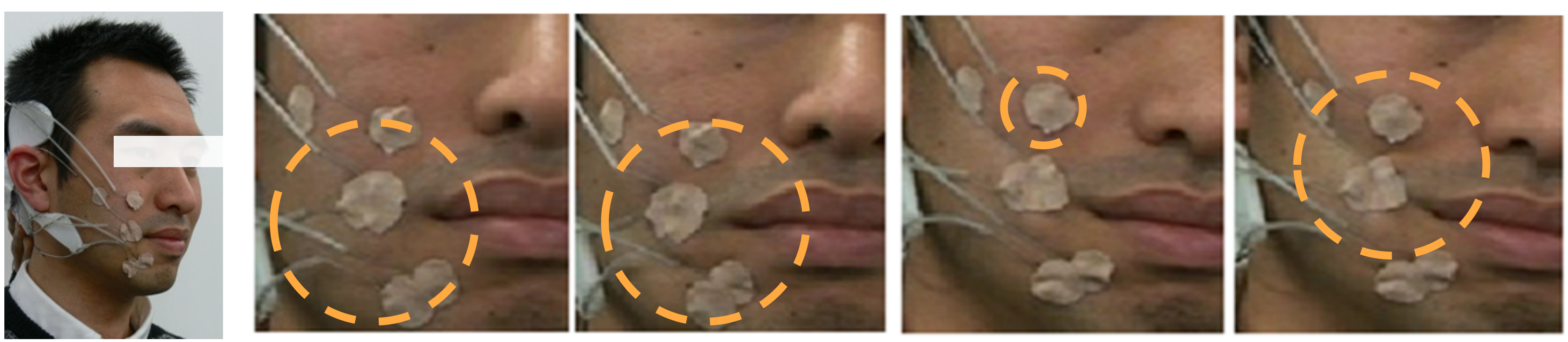
Assisting facial actions

Bioelectrical Signal extraction

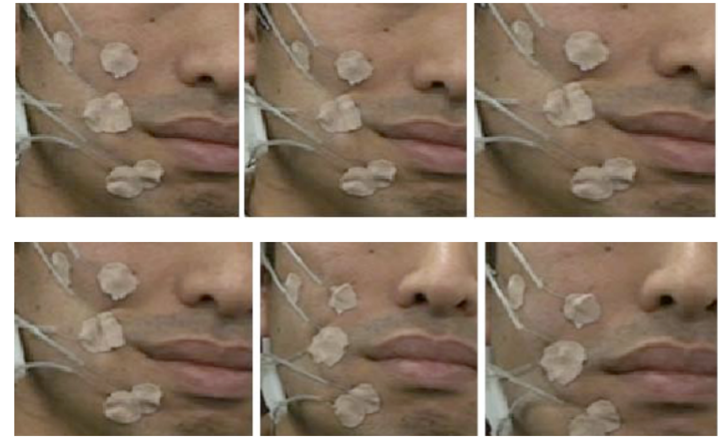
Parotid Gland



© Elsevier. Drake et al: Gray's Anatomy for Students - www.studentconsult.com
Facial nerve [VII] on the face. **A.** Terminal branches. **B.** Branches before entering the parotid gland.

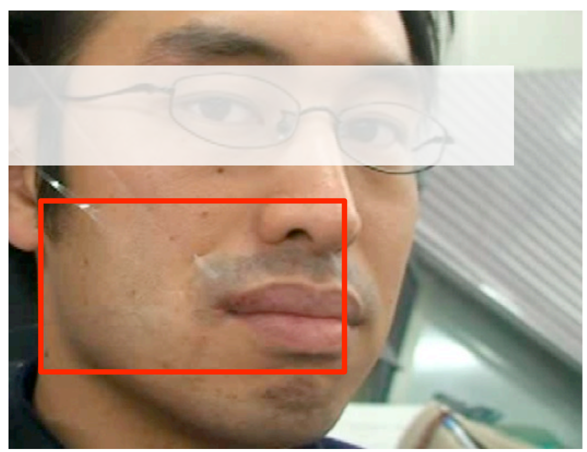


Basic Experiment: Artificial Wrinkles



4 locations & 5 DoF

- Actuation (transfer): polyester sheet (low stretch)
- Fixation: 10-20μm mat-polyurethane film (force < 4.0[N])

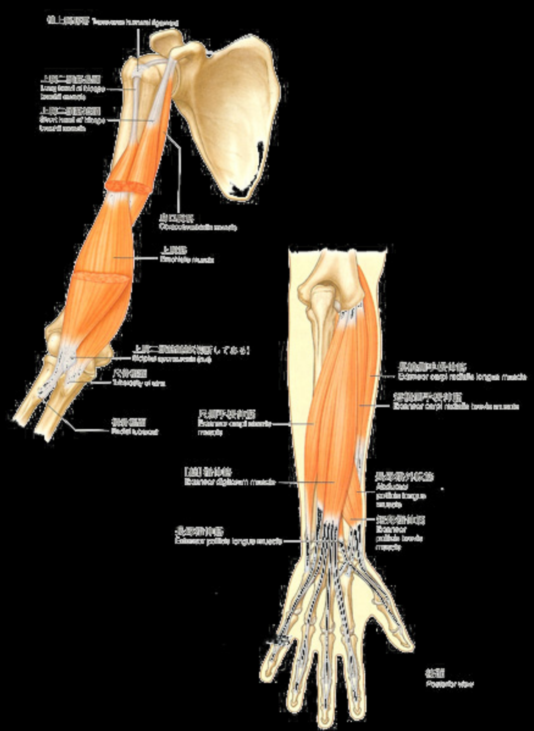


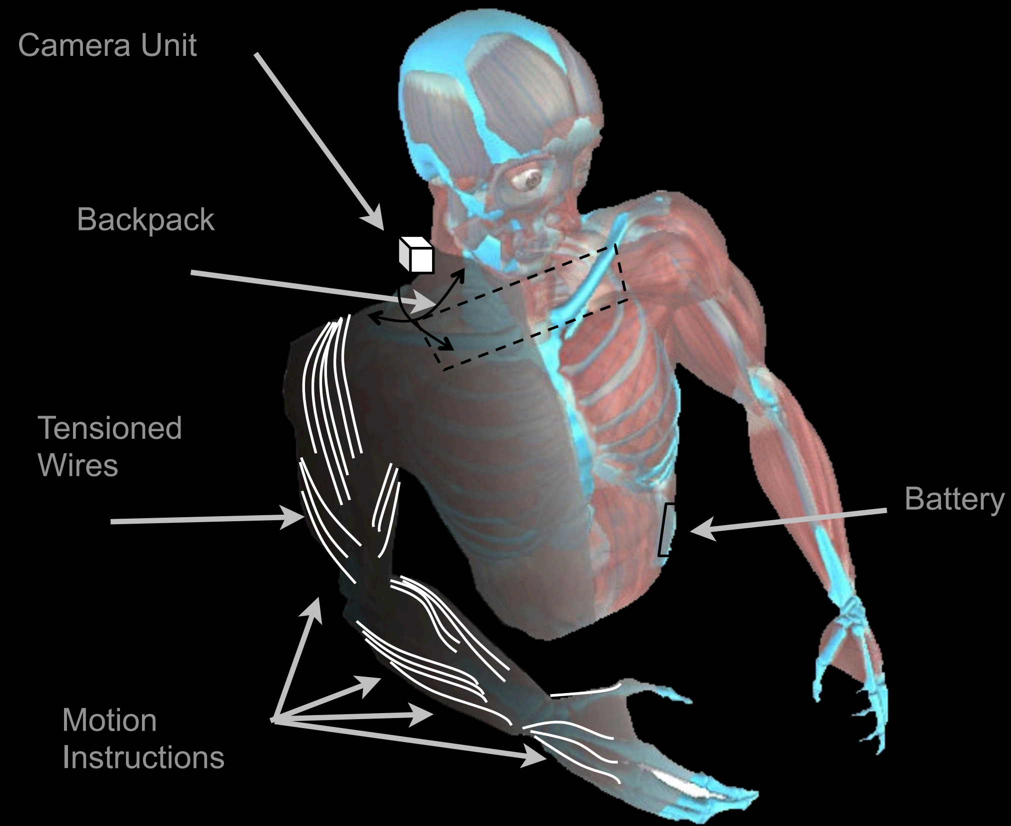
Dushyantha & Suzuki, IEEE IROS, France, 2008

Robot Skin

Cybernics
fusion of human, machine and information systems

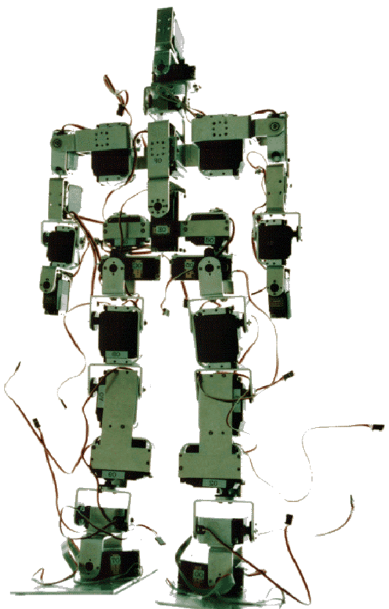
Members: Sho Yano, Kenji Suzuki

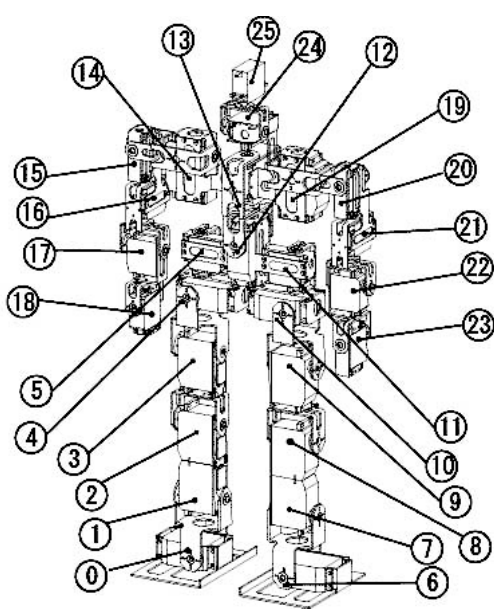




"Development of an Anatomical-based Robot Skin for Wearable Motion Instructions," **IEEE ICRA** (submitted)

A Case-study of Coaching
Experiments with a small humanoid



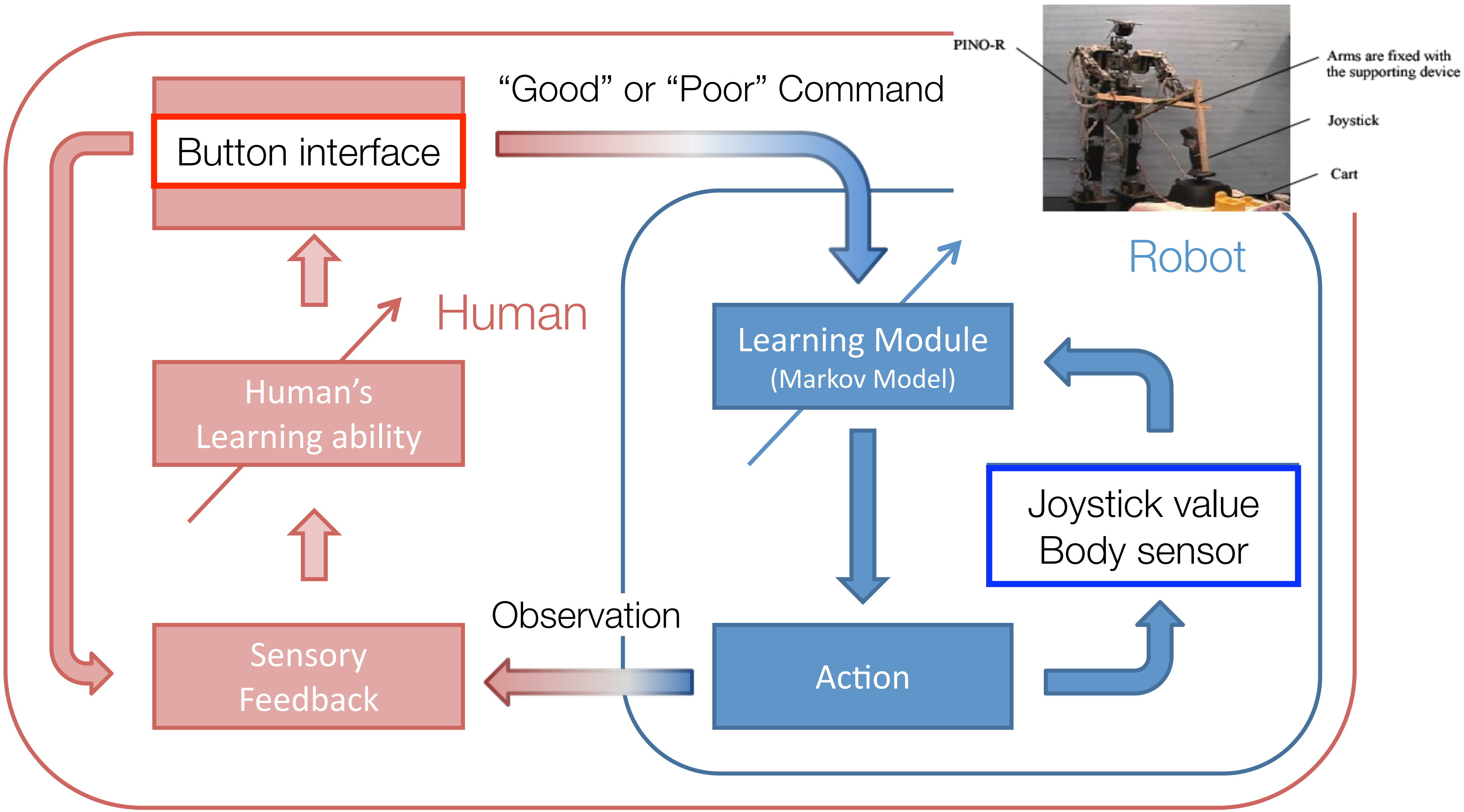


Height		70cm
Weight		4.5kg
DoF	Neck	2DOFs
	Waist	2DOFs
	Arm	each 5DOFs
	Leg	each 6DOFs
	Total	26 DOFs

Angular Sensors (x26): All joints

FSR Sensors (x8): bottom of feet

Coaching a small humanoid

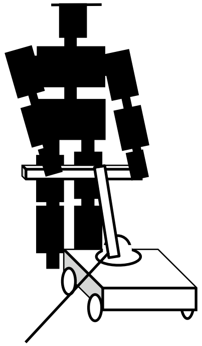


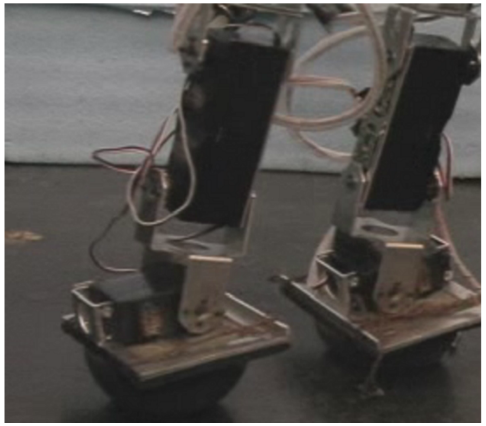
Balancing Experiment

$$\boldsymbol{y}(t) \quad = \quad j(t)\boldsymbol{k}(t)$$

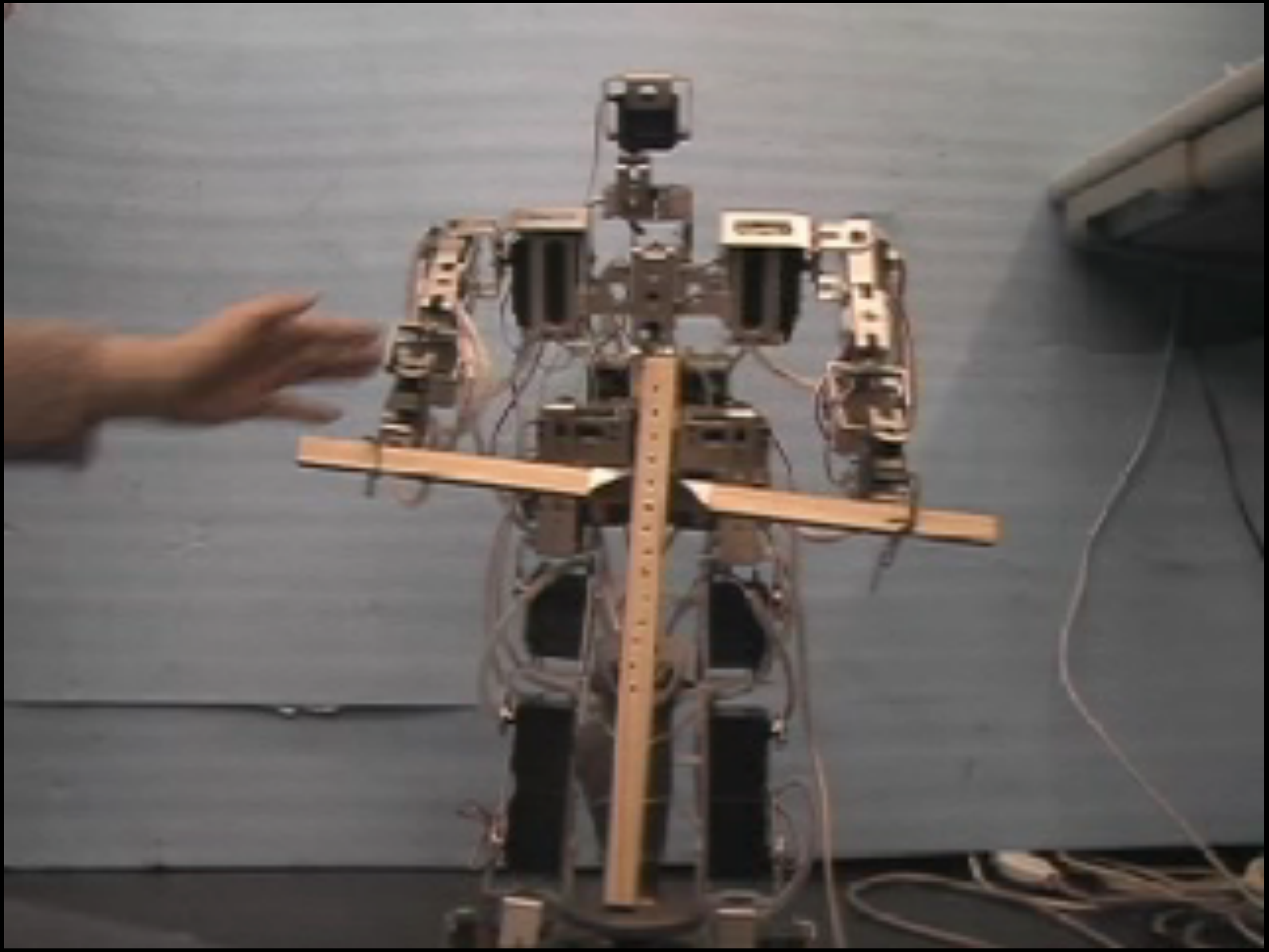
Gain parameter for
2 Joints (Ankle or hip)

$$\boldsymbol{k}(t) \quad = \quad \begin{bmatrix} k_{ankle}(t) \\ \\ k_{hip}(t) \end{bmatrix}, \qquad \boldsymbol{y}(t) = \begin{bmatrix} y_{ankle}(t) \\ \\ y_{hip}(t) \end{bmatrix}$$



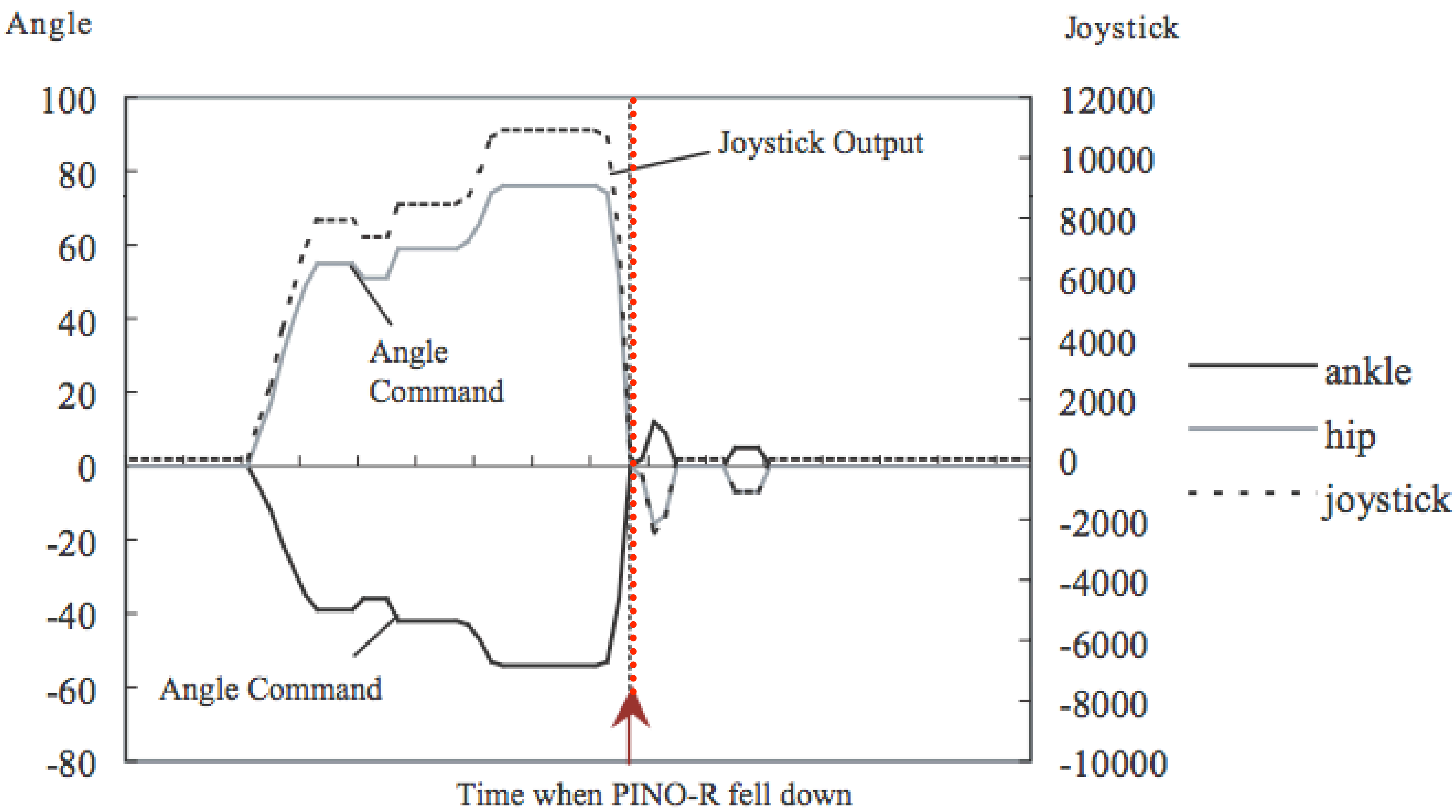


Parameter update: Each Episode (One sequence and observation)

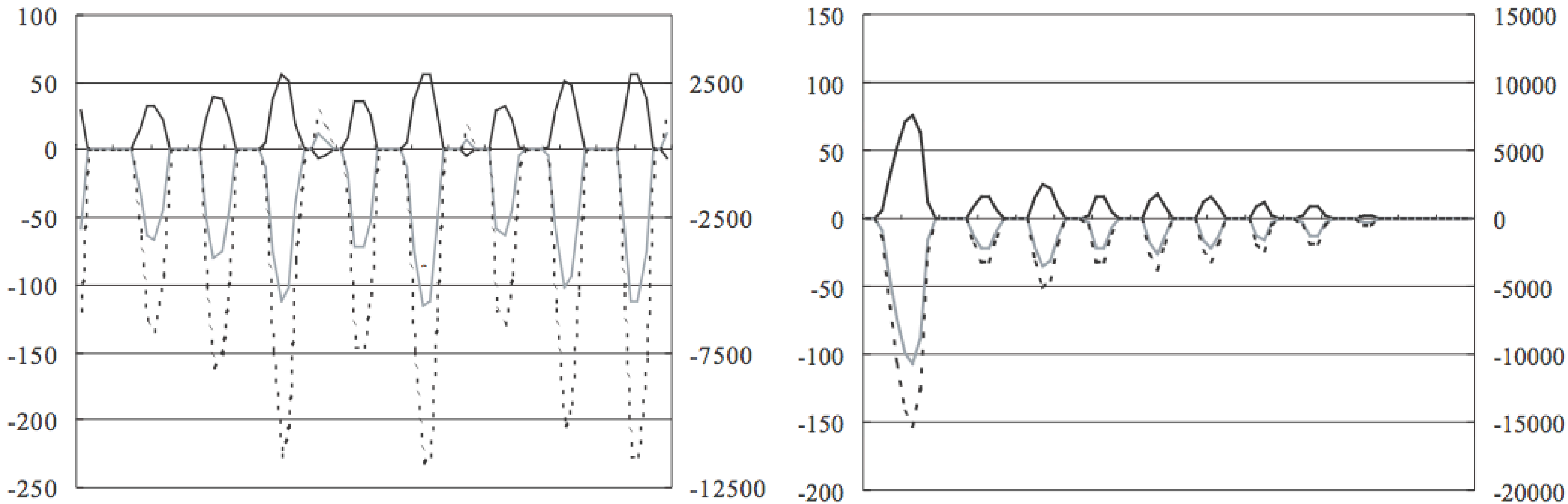


A small humanoid to keep balancing

Balancing: before training



Balancing: before/after training



Before training

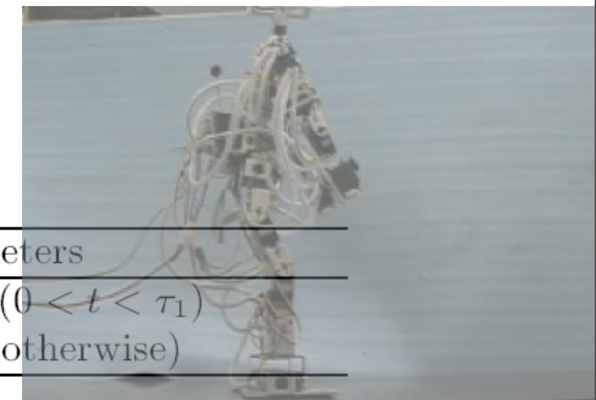
After training

Ankle joint

Hip joint

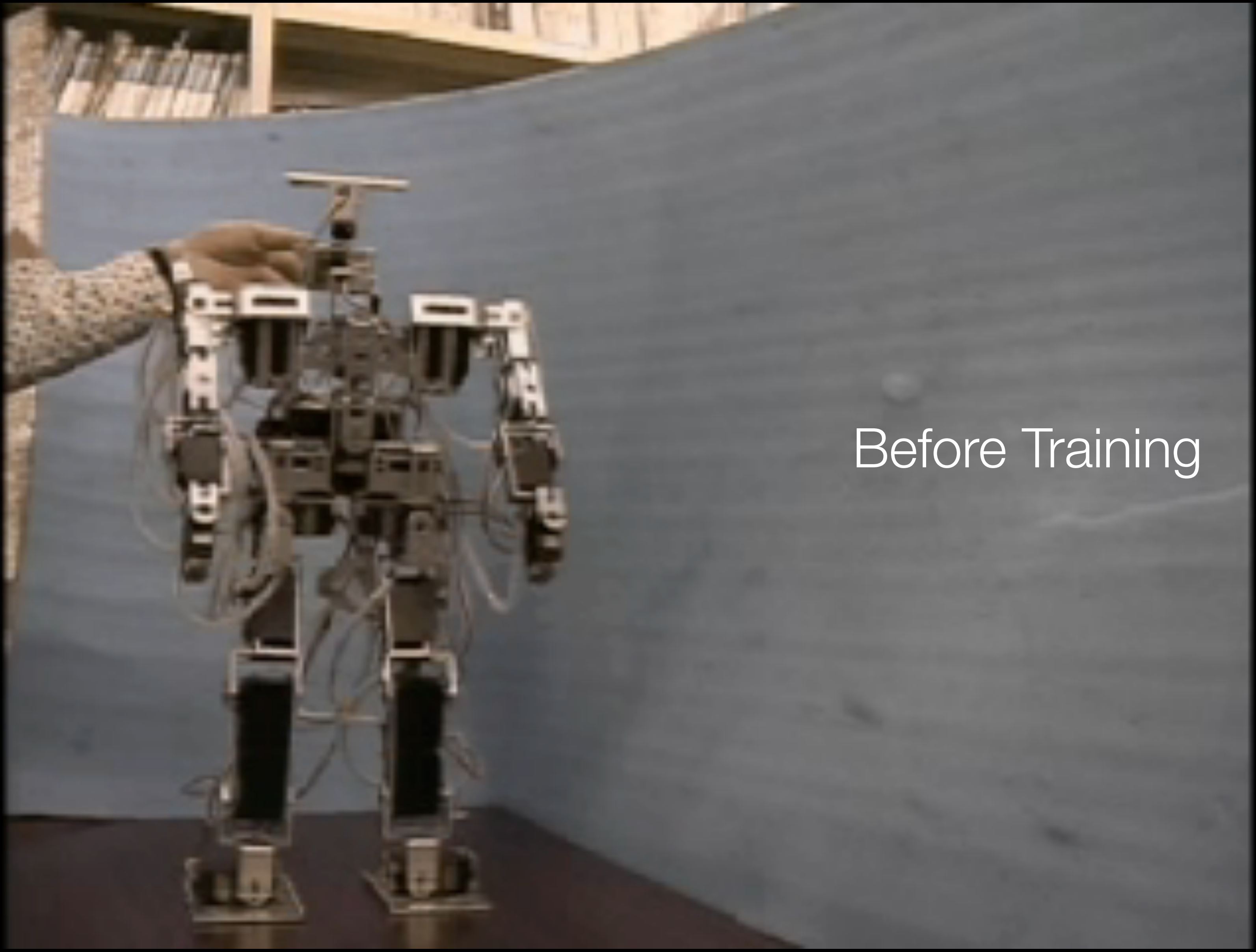
Joystick (feedback) value

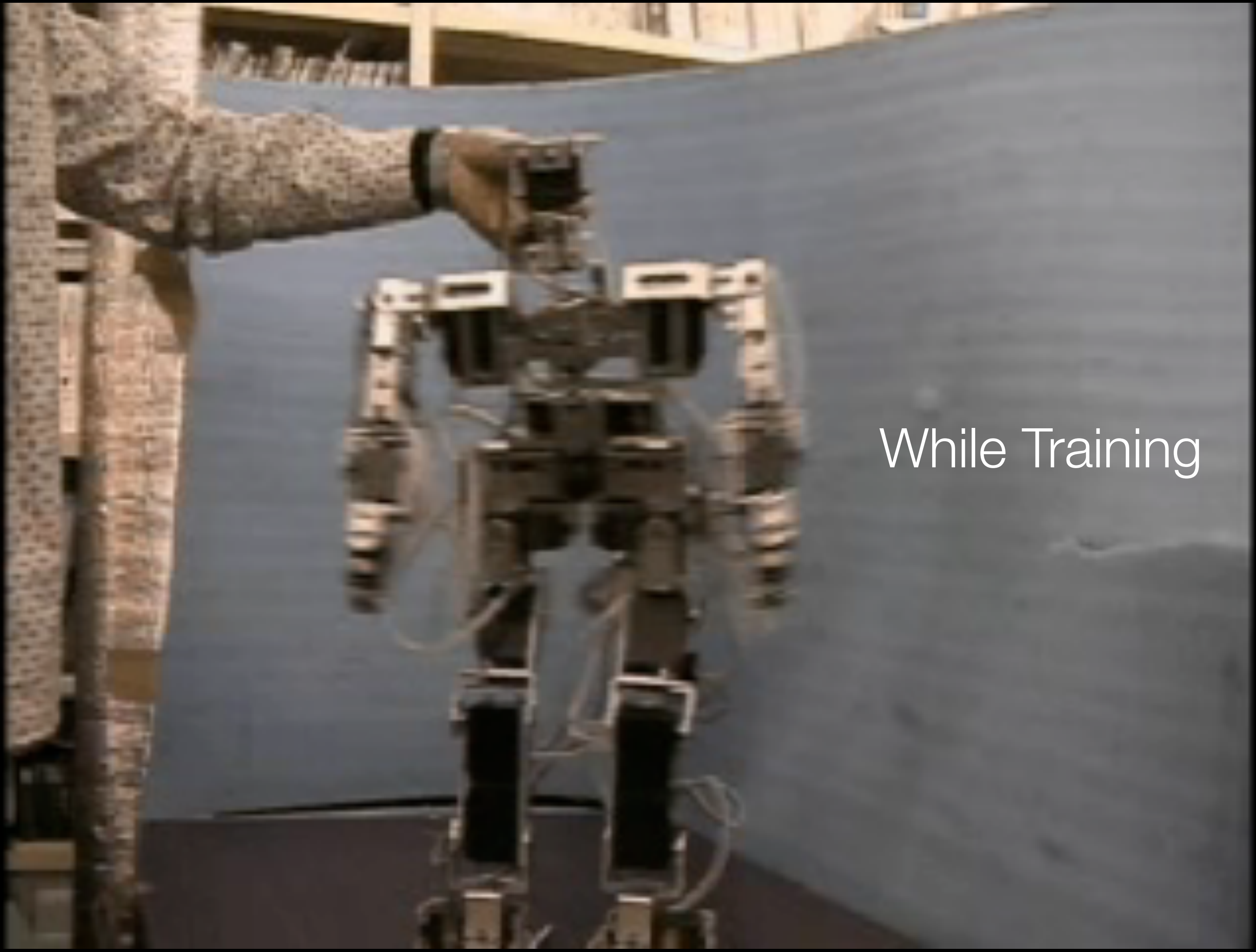
Walking Experiment

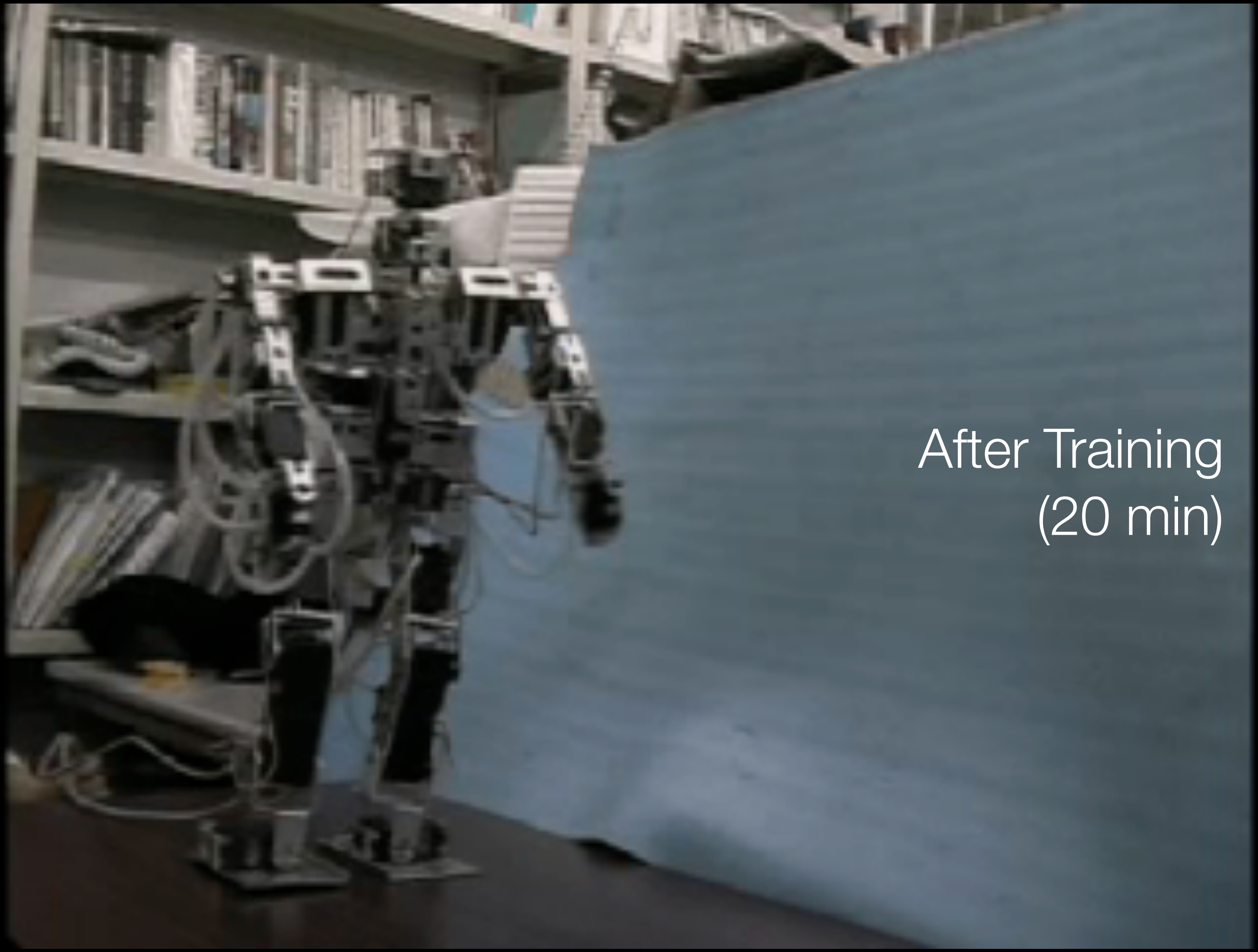


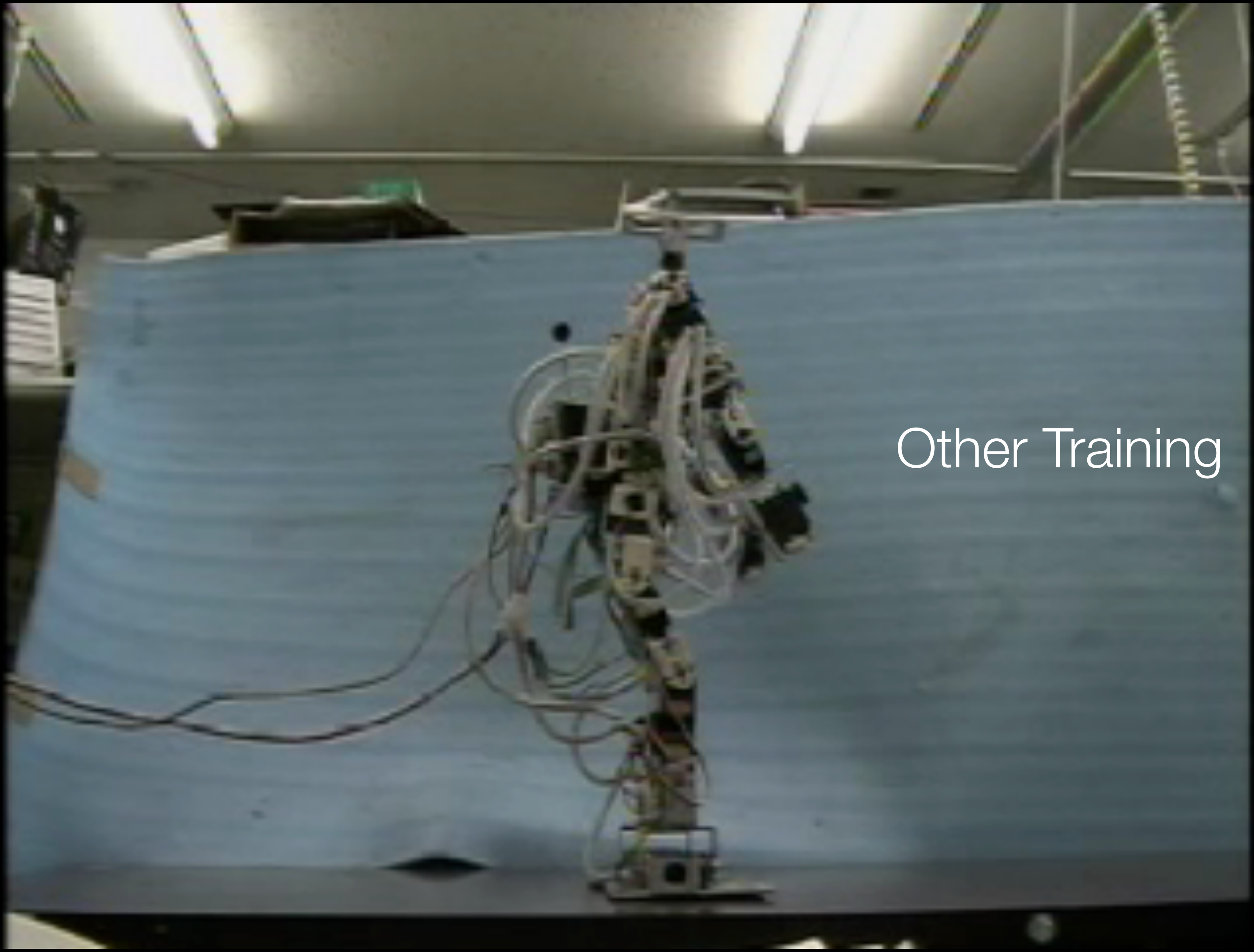
Body parts	Basic motion function	Parameters	
Knee (longitudinal)	$a_k(t) sin(\omega t + \varphi)$ $\omega = \pi \ / \ \tau_1$	left	$a_k(t) = p_{kl}(0 < t < \tau_1)$ $a_k(\mathrm{t}) = 0$ (otherwise)
			$\varphi = \tau_1, \tau_2$
		right	$a_k(t) = p_{kl}(0 < t < \tau_1)$ $a_k(\mathrm{t}) = 0$ (otherwise)
			$\varphi = \tau_1, \tau_2$
Thigh (longitudinal)	$\lvert 2a_t t - \tau_1 \rvert - a_t$	left	$a_t(t) = p_{tl}$
		right	$a_t(t) = p_{tr}$
Thigh (latitude)	$a_{st} = sin \ \omega t$ $\omega = \pi \ / \ \tau_1$	left	$a_{st}(t) = p_{stl}$
		right	$a_{st}(t) = p_{str}$
Ankle (longitudinal)	$a_a = sin(\omega t + \varphi)$ $\omega = \pi \ / \ \tau_1$	left	$a_a(t) = p_{al}$
		right	$a_a(t) = p_{ar}$
Ankle (latitude)	$a_{sa} = sin(\omega t + \varphi)$ $\omega = \pi \ / \ \tau_1$ $\varphi = 0, \pi$	left	$a_{sa}(t) = p_{sal}$
		right	$a_{sa}(t) = p_{sar}$
Hip (longitudinal)	$a_h = sin(\omega t + \varphi)$ $\omega = \pi \ / \ \tau_1$ $\varphi = 0, \pi$		$a_h(t) = p_h$

13 Parameters (Gain and phase)


Before Training


While Training


After Training
(20 min)

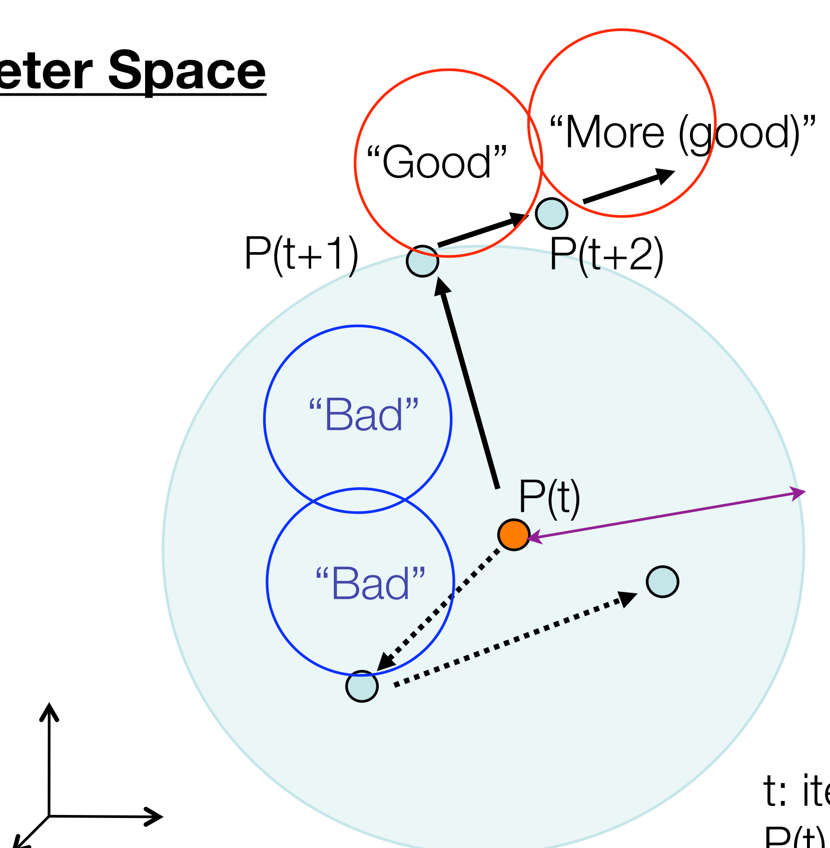

Other Training

Results: Balancing and walking

- No specific simulation model / No kinematic calculation
  - But simple primitives based on the kinematic model
- The **Quality** of balancing and walking is **dependent** upon the coach's performance
- **Analysis** of human's instruction
  - Revealed a simple update rule based on Markov process

Search Process in Parameter Space

Parameter Space

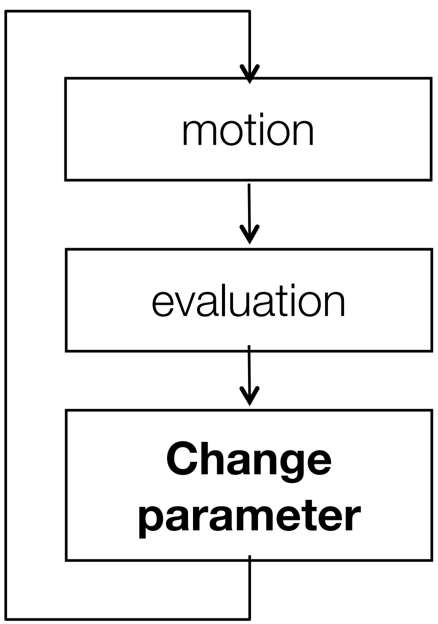


n- Dimensions
(number of search parameters)

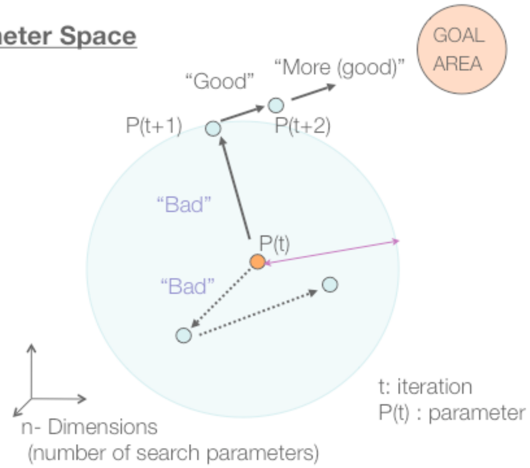
t: iteration
P(t) : parameter

"Bad", "Good" express degree to the parts of the robot

Each Iteration



Issues in the search process



"Good" Command

- Reinforcement learning
- Not coarse but fine search
- Learning parameter with regard to time is used for the parameter convergence

"Bad" Command

- "Bad" is not "No Good"
- Few information to show **the good direction**
- From **Random search** strategy to probabilistic approach

Learning from success and failure

- A modified reinforcement learning model
- Not only maximizing **the likelihood of success** but also minimizing **the likelihood of failure**
  - Introducing the likelihood of failure

- How the agent/robot performs when **the “bad” is given**?
  - Introducing the **“Temperature”** factors in the state space

Formulations

Likelihood : Success (Good)

$$P(g) = \sum_{s \in S} \sum_{j=1}^{n} P(g|a_j, s) \boxed{P(a_j|s)P(s)}$$

Choice probability

Likelihood : Failure (Bad)

$$P(u) = \sum_{s \in S} \sum_{j=1}^{n} P(u|a_j, s) \boxed{P(a_j|s)P(s)}$$

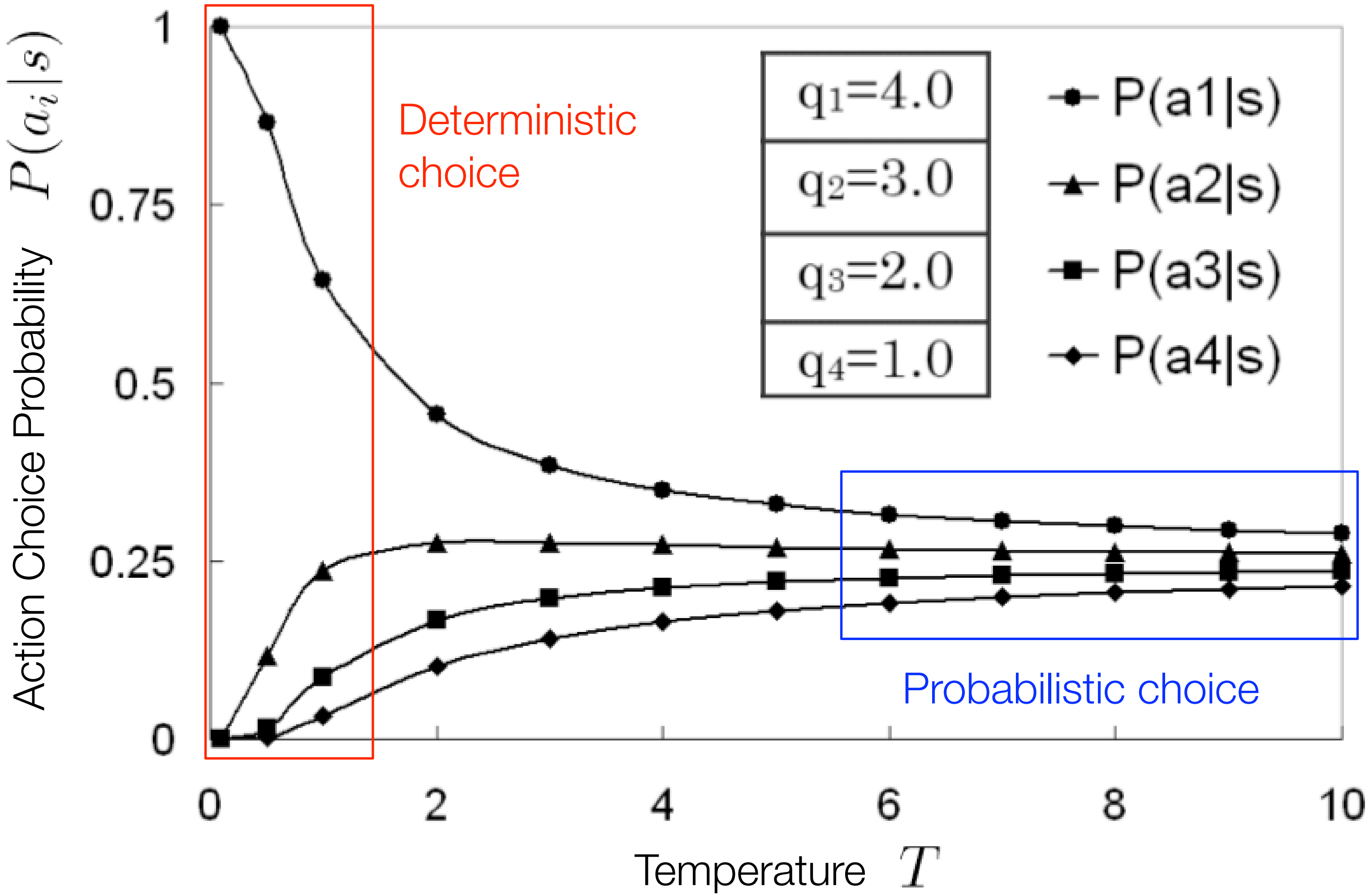
Success-Failure & Temperature Model

Choice probability with state-temperature

$$P(a_i|s) = \frac{\exp(q(s,a_i)/T)}{\sum_{j=1}^{n} \exp(q(s,a_j)/T)}$$

$q(s,a_i)$	Action priority in the set of a_i on the state s
$P(a_i\|s)$	Choice probability of action a_i on the state s
$T(s)$	**Temperature**

Choice Probability and Temperature



Combination of success/failure

$$L = P(g)(1 - P(u))$$

$P(g)$ Likelihood : Success (Good)

$P(u)$ Likelihood : Failure (Bad)

Maximizing the likelihood of success
and minimizing the likelihood of failure

$$\ell = \ln P(g) + \ln(1 - P(u))$$

$$T(s) = T(s) + \eta \frac{\partial \ell}{\partial T(s)}$$

* Temperature updating is theoretically guaranteed

$$\eta = \frac{1}{\sum_{j=1}^{n} \left| \frac{\partial P(a_j|s)}{\partial T_s} \right|}$$

Temperature learning

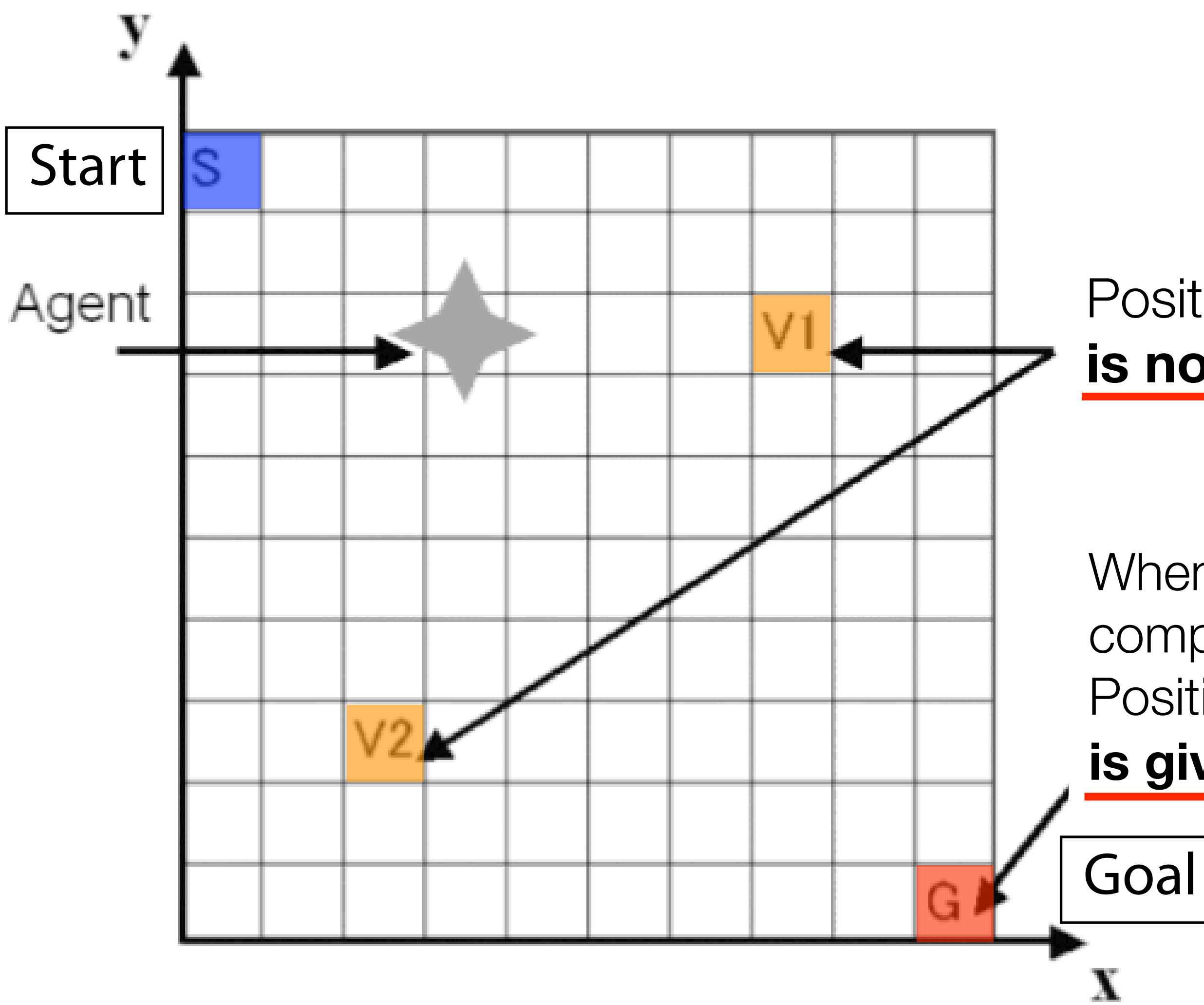
$$\frac{\partial \ell}{\partial T(s)} = \sum_{i=1}^{n} \Bigg[\left(\frac{P(a_i, s|g)}{P(a_i|s)} - \frac{P(a_i, s|u)}{P(a_i|s)} \times \frac{P(u)}{1 - P(u)} \right) \times \frac{1}{T(s)^2} \left(\frac{\sum_{j=1}^{n} q(s, a_j) \exp(q(s, a_j)/T(s))}{\sum_{j=1}^{n} \exp(q(s, a_j)/T(s))} - q(s, a_i) \right) P(a_i|s) \Bigg]$$

$$P(a_i, s|g) = \frac{\sum_{o=1}^{M} N_o(a_i, s, g)}{\sum_{o=1}^{M} N_o(g)}$$

$$P(a_i, s|u) = \frac{\sum_{o=1}^{M} N_o(a_i, s, u)}{\sum_{o=1}^{M} N_o(u)}$$

$$P(u) = \frac{\sum_{o=1}^{M} N_o(u)}{\sum_{o=1}^{M} N_o}$$

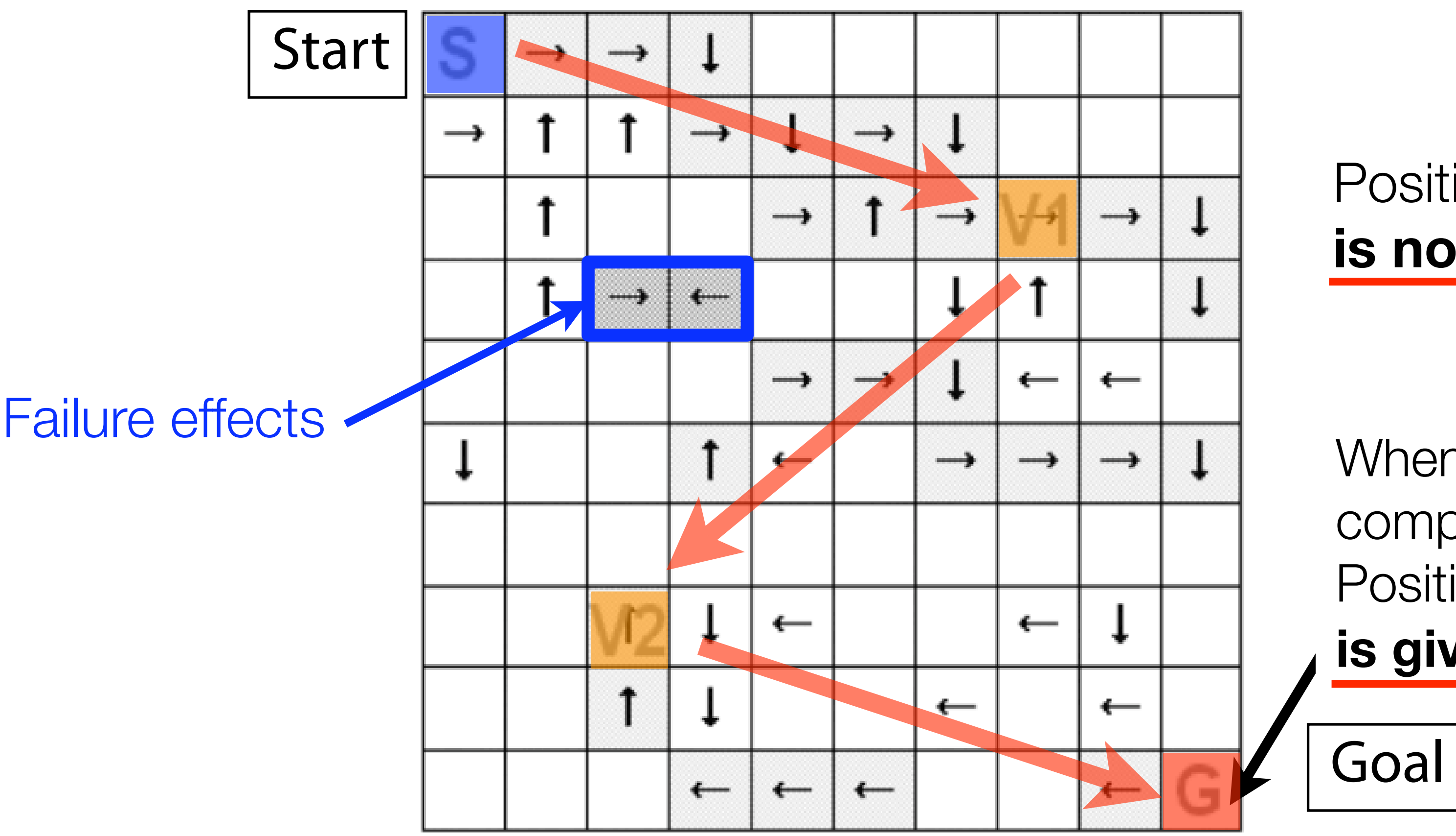
2D Grid : Task with Hidden Sub-goals



Positive reward
is not given

When only the agent completes the task, Positive reward
is given

2D Grid : Experimental Result

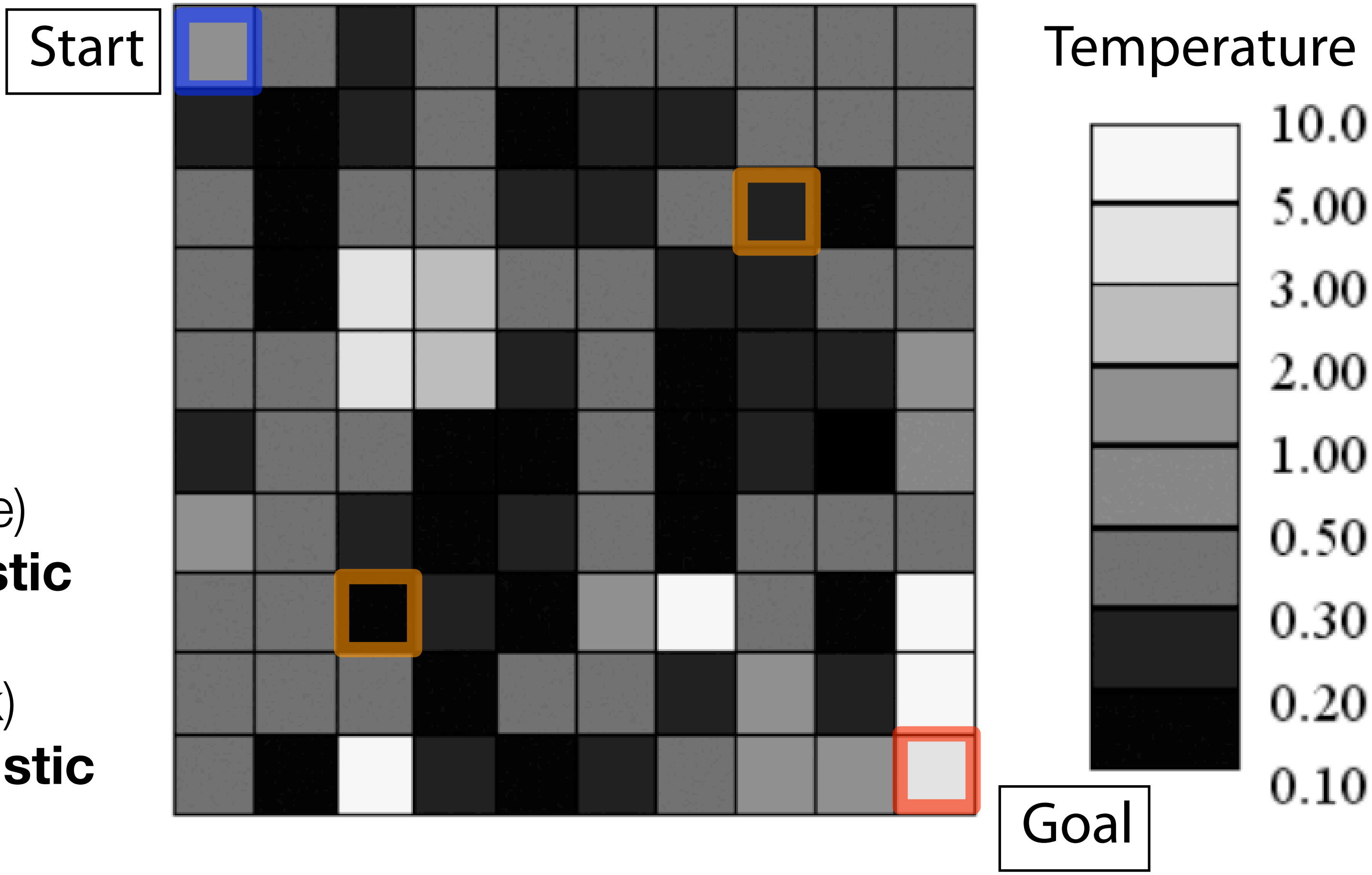


Positive reward
is not given

When only the agent completes the task, Positive reward
is given

2D Grid : Learnt Temperature Distribution

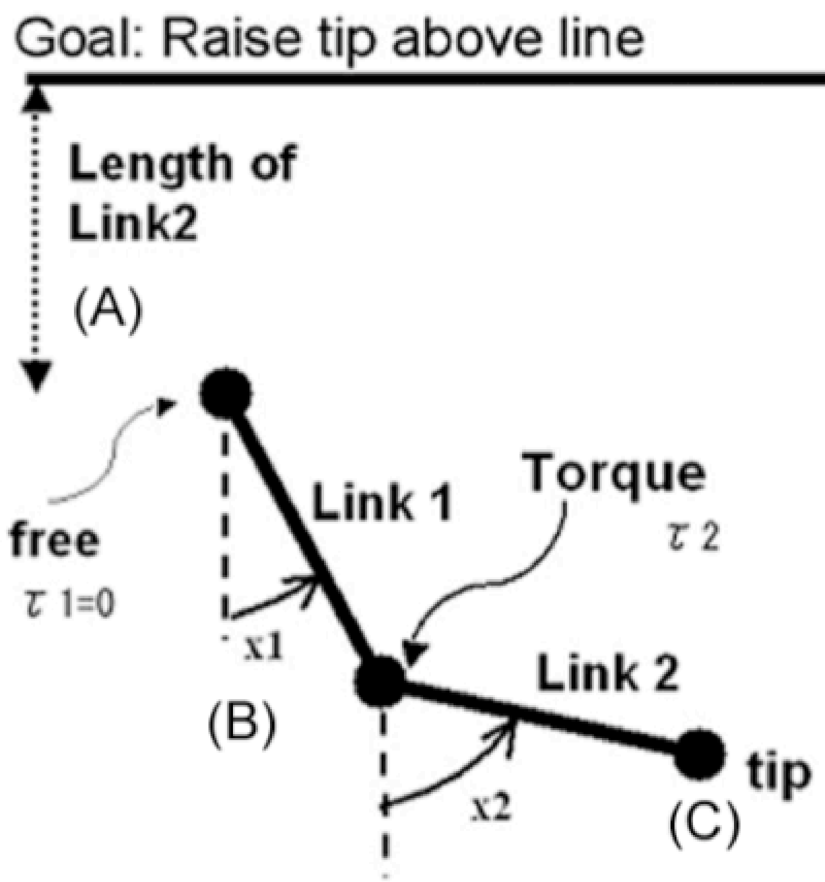
$$T(s) = T(s) + \eta \frac{\partial \ell}{\partial T(s)}$$



High T (White)
- **Probabilistic**

Low T (Black)
- **Deterministic**

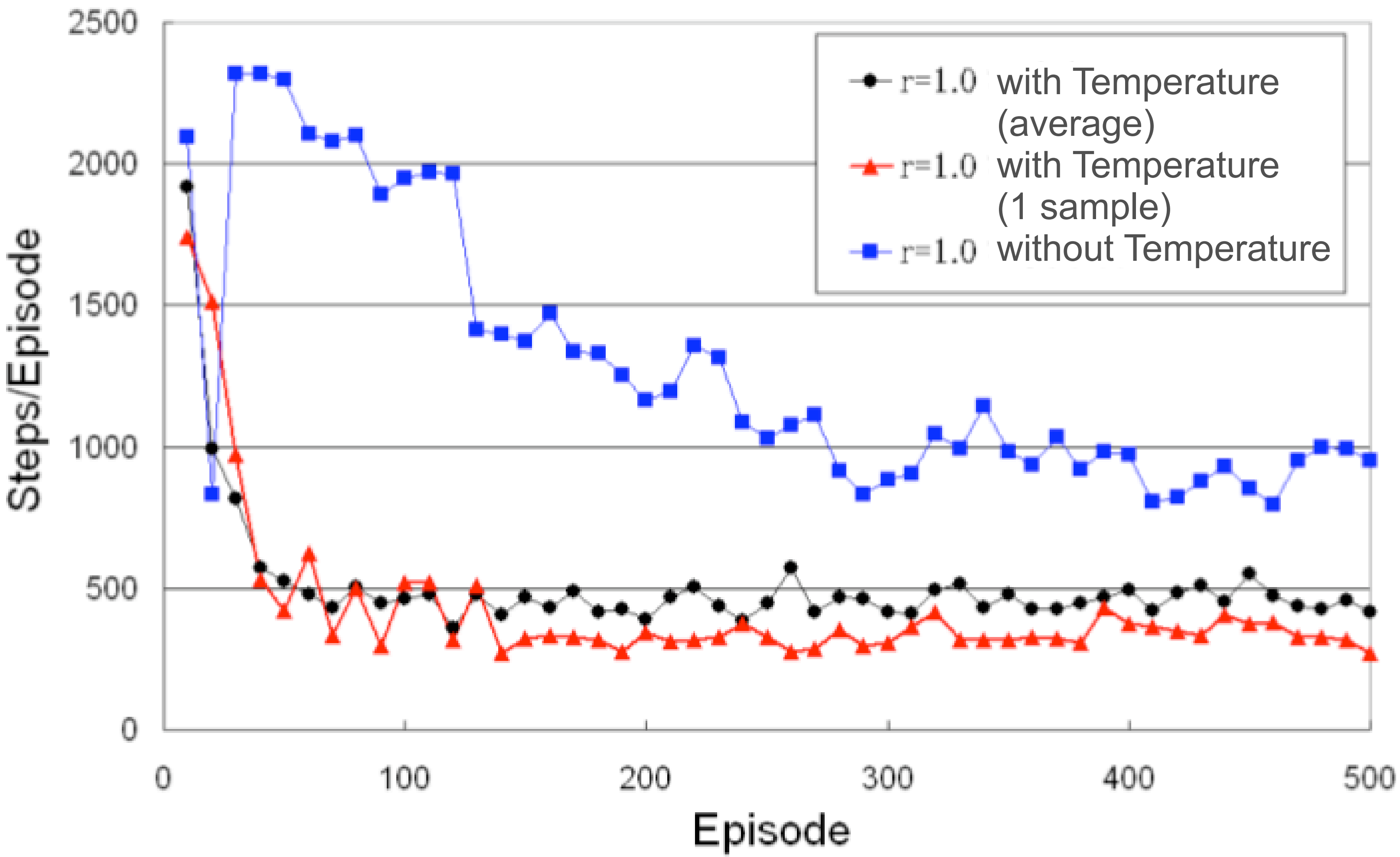
Acrobot* Model



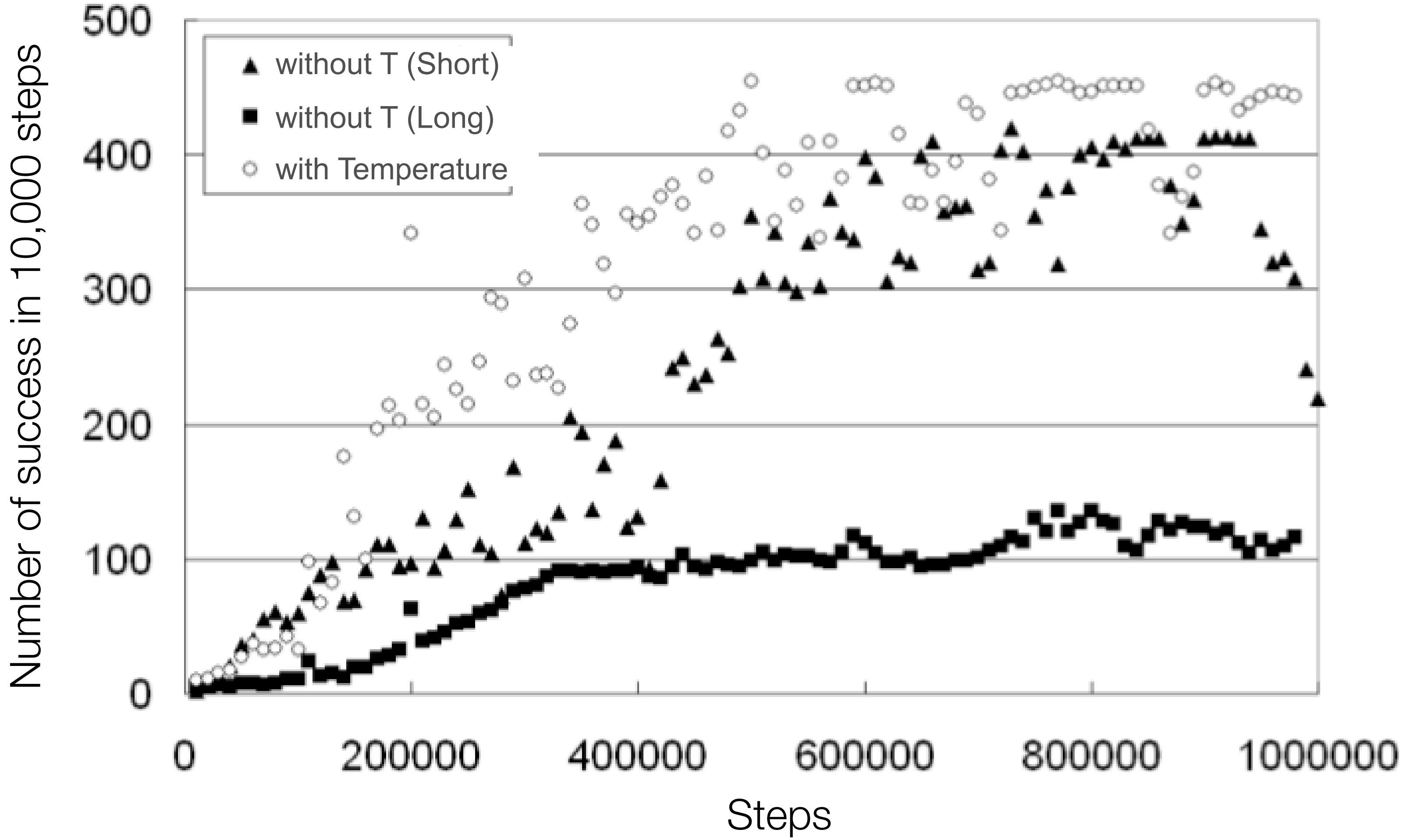
x_1, x_2 (rad) Angle from vertical direction	
m_1, m_2 (kg) Link mass	1.0 1.0
l_1, l_2 (m) Link length	1.0, 1.0
S_1, S_2 Center of mass	0.5 0.5
I_1, I_2 (kg m^2) Moment of inertia	1.0, 1.0
τ_1, τ_2 (Nm) Torque	0 [-1,0,1]

* R. S. Sutton and A. G. Barto: Reinforcement Learning, MIT Press, 1998

Comparison to conventional RL

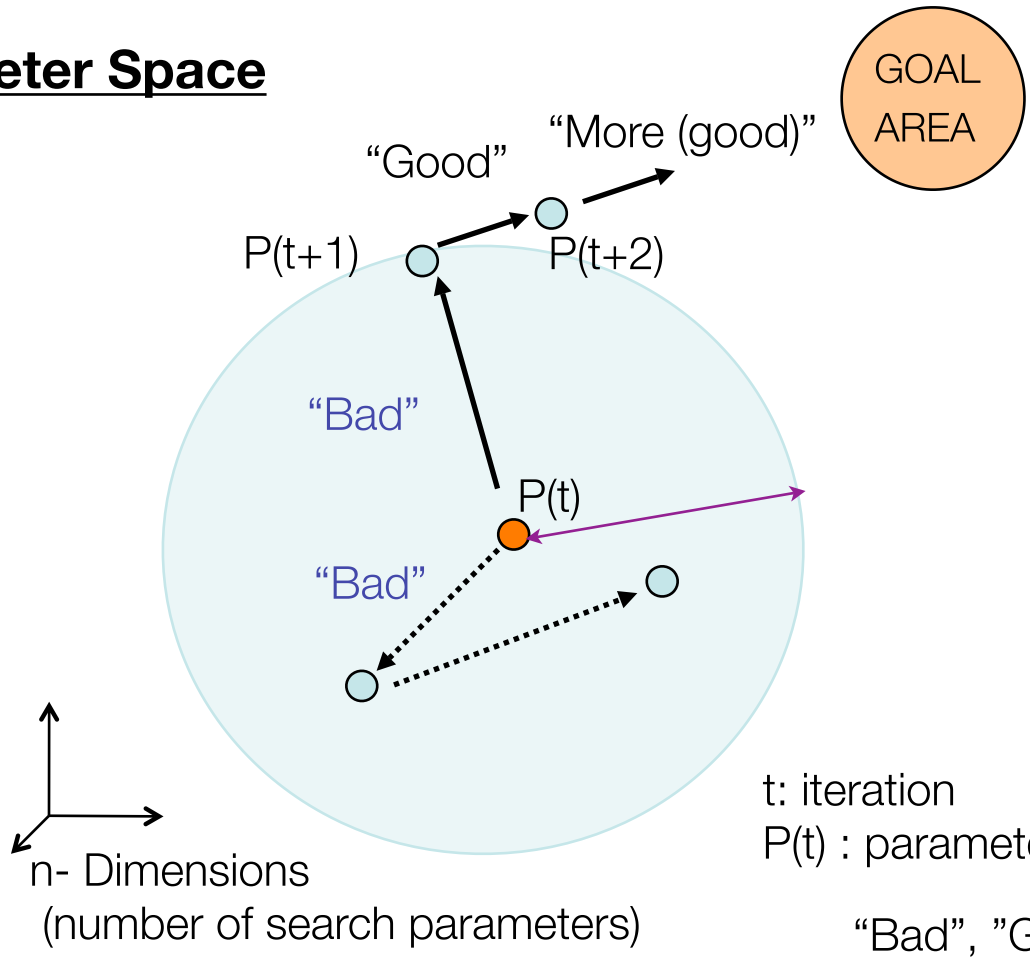


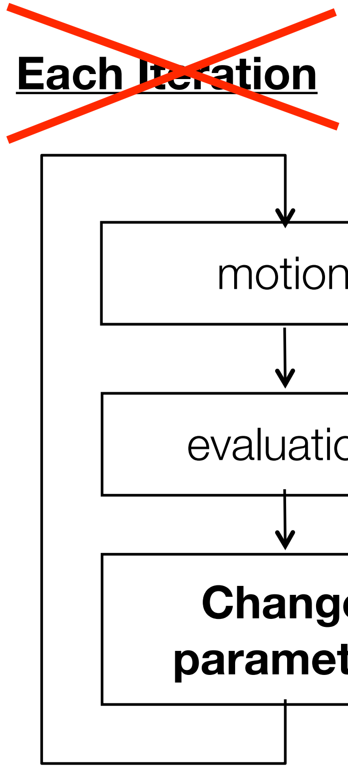
Performance with Acrobat Model



Discussion
Search Process in Coaching

Parameter Space

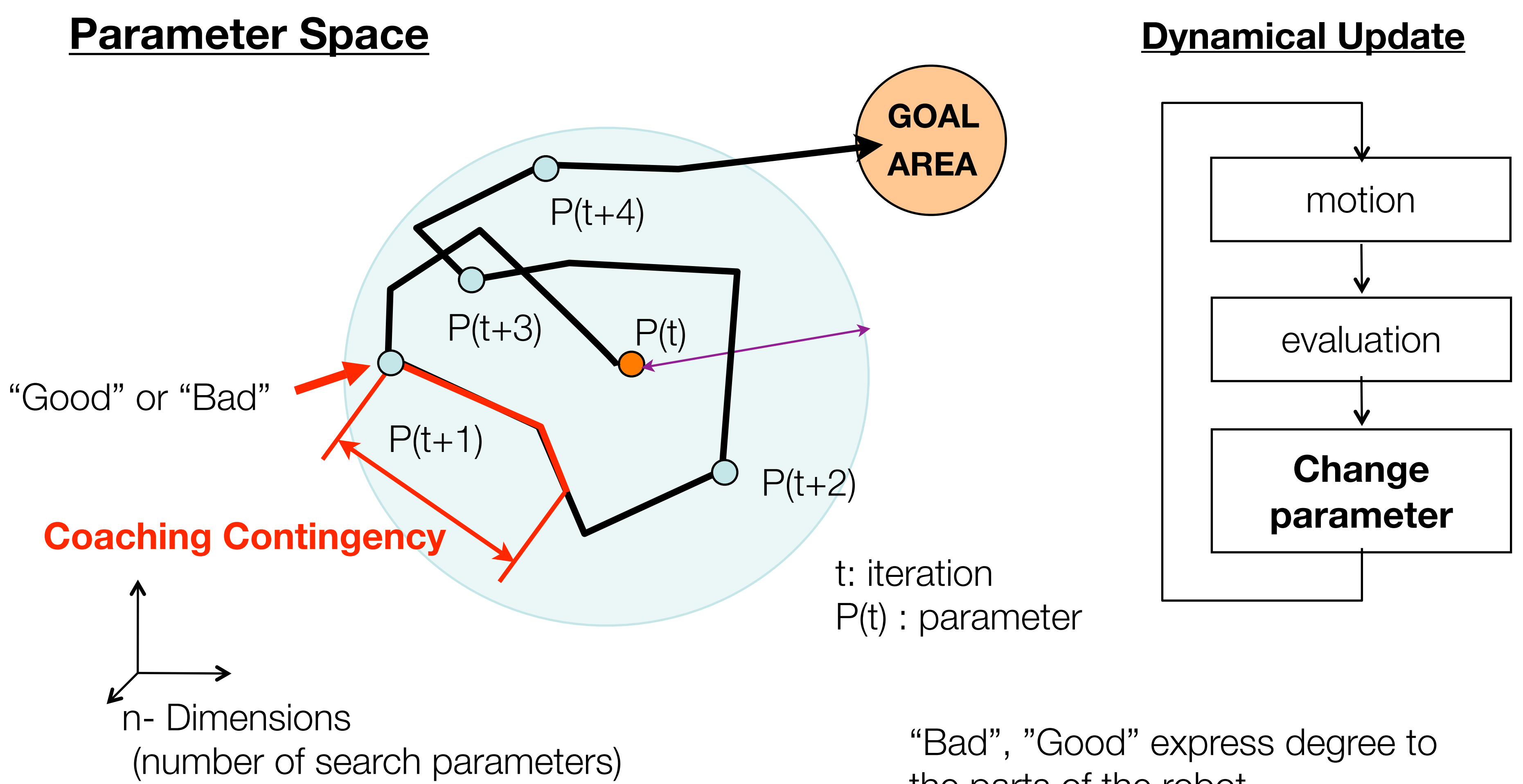




t: iteration
P(t) : parameter

"Bad", "Good" express degree to the parts of the robot

Discussion
Search Process in Coaching

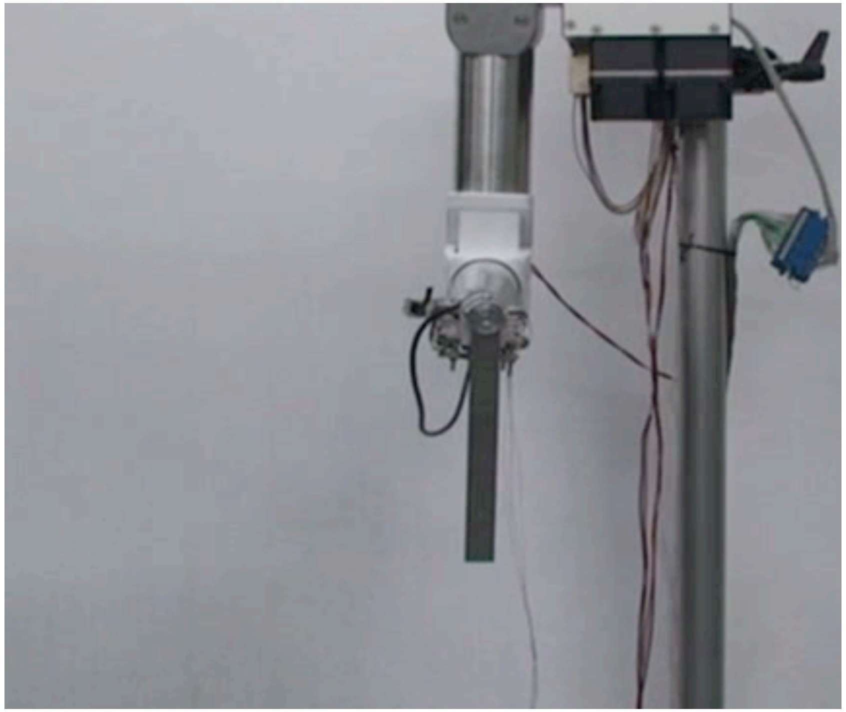


"Bad", "Good" express degree to the parts of the robot

Discussion
Coaching a robot and contingency

Direct continuous control by dial-shape interface (1ch)

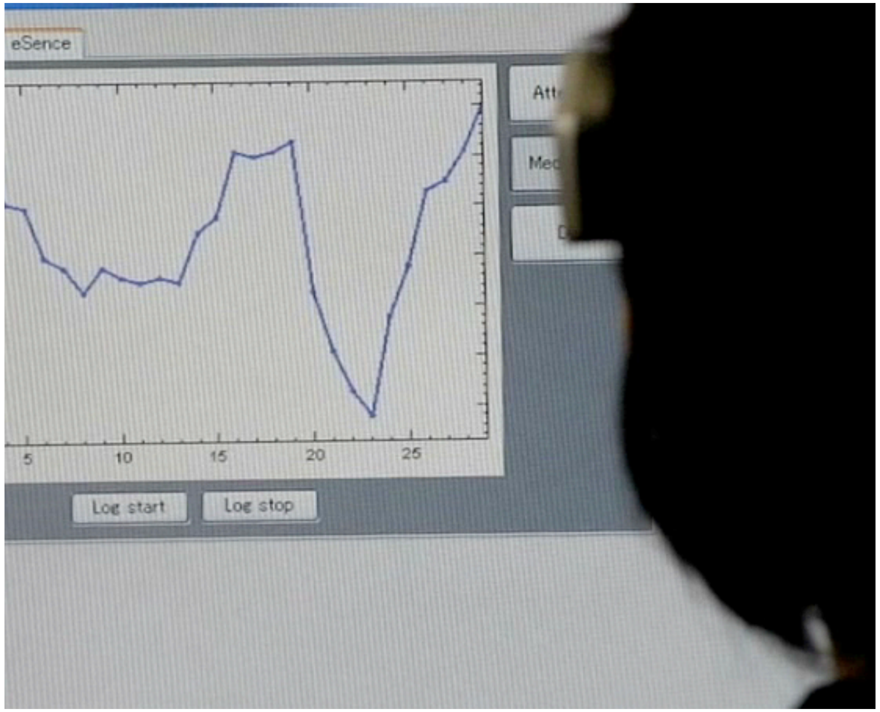
1 ch dial-shape interface



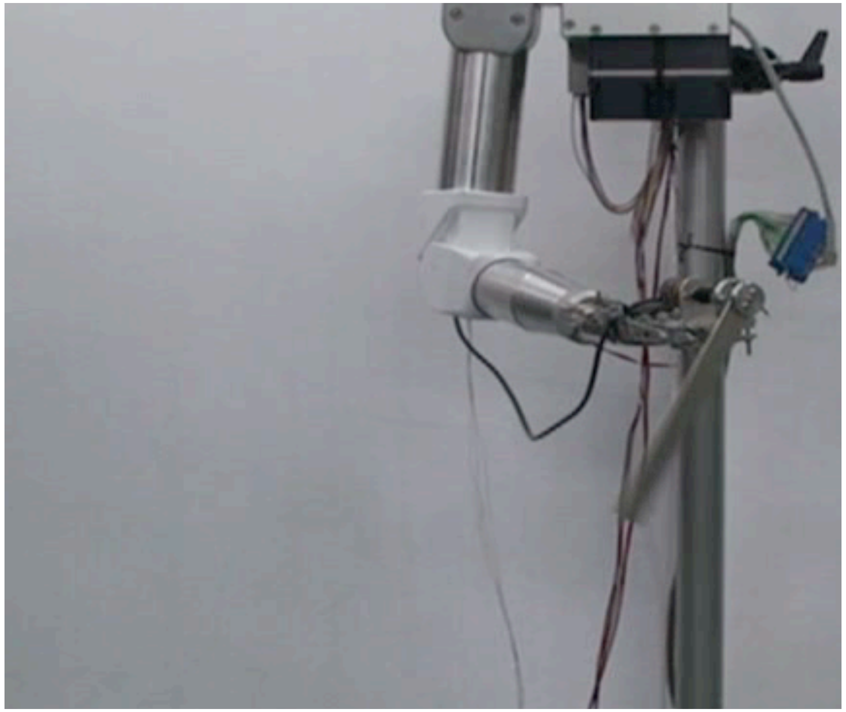
Manipulator + robot hand 6+17 DoF

Discussion
Coaching a robot and contingency

Direct continuous control by bioelectrical signal (1ch)



1 channel ECG-based signals



Manipulator + robot hand 6+17 DoF

Summary

- Coaching by a very simple instruction through continuos interaction
- Showing potential applications of coaching by using bioelectrical signals
- Balancing and walking task by a small humanoid robot
- A model of learning from success and failure
- Coaching a robot with timing control