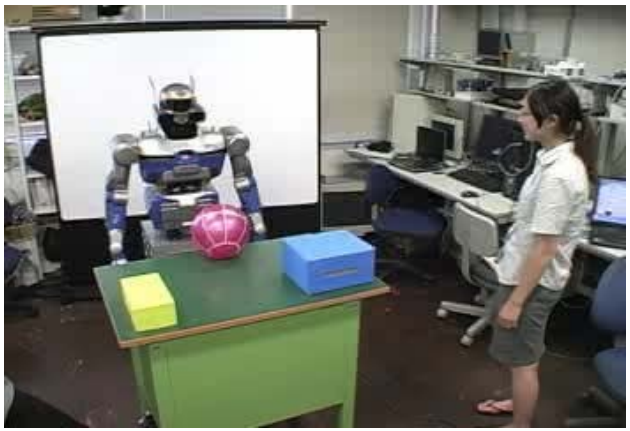
Robotic Coach: how to revise humans' motions by Emphatic Demonstration

Tetsunari Inamura
National Institute of Informatics
The Graduate University for Advanced Studies

National Institute of Informatics

The Graduate University for Advanced Studies (SOKENDAI)
国立大学法人
総合研究大学院大学

My previous work: daily life robots



The University of Tokyo, prof. Inaba's lab.

Main interest

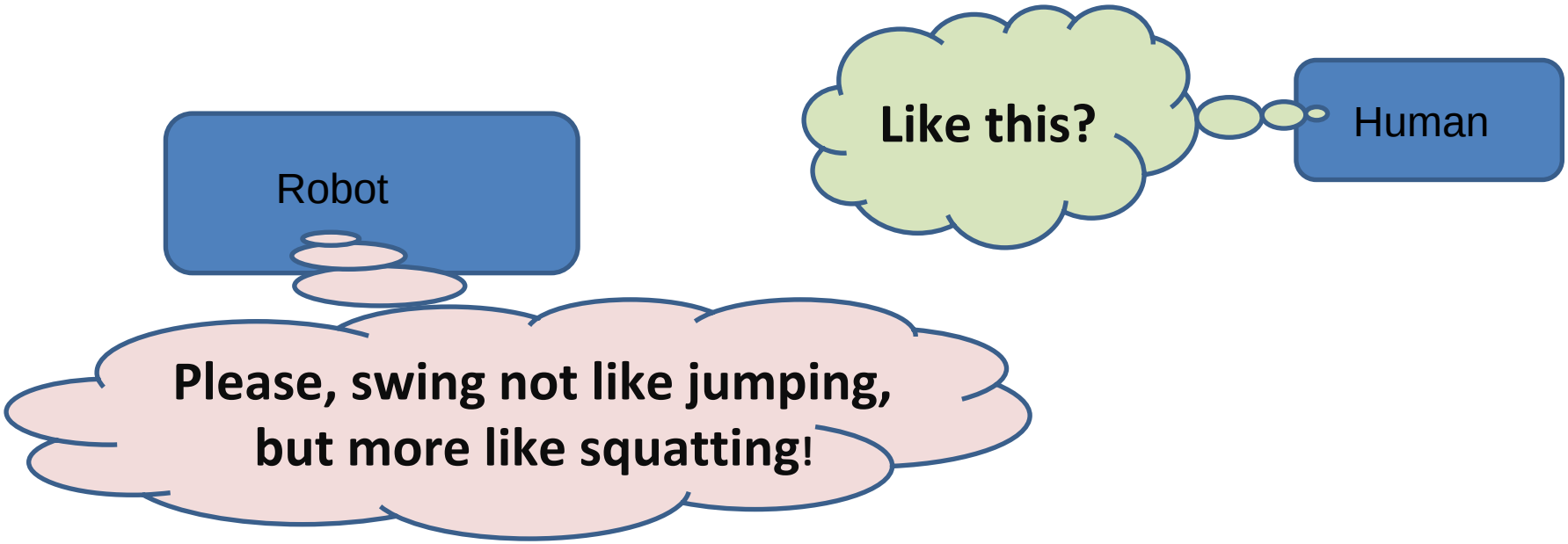
- How to integrate symbolic expression and motion pattern of whole body
  - For easy interaction between human and robots

"Pouring water in a white cup"

"a white cup"

Latest topic

- Robotic Coach that teaches human beings



- Realization of not imitative robots, but robots that can let human beings imitate
  - One of the most useful and complex tasks which require integration of symbols and motion performance

Background

- Standard coaching methods in sports / dancing
  - Coaching by demonstration (or video material)
    - Imitate whole body motion is often difficult
    - So many attention points
  - Direct coaching with physical interaction
    - Effective but expensive
  - Coaching by verbal explanation
    - Low cost, effective in various situations
    - Conversion from verbal expression into motion is unstable

Purpose of this project

- Realization of robotic coach system that is used for training of human beings
- Integration of verbal explanation and physical demonstration with emphasis
- Design of common representation among “emphasis of motion” and “explanation by verbal expression”

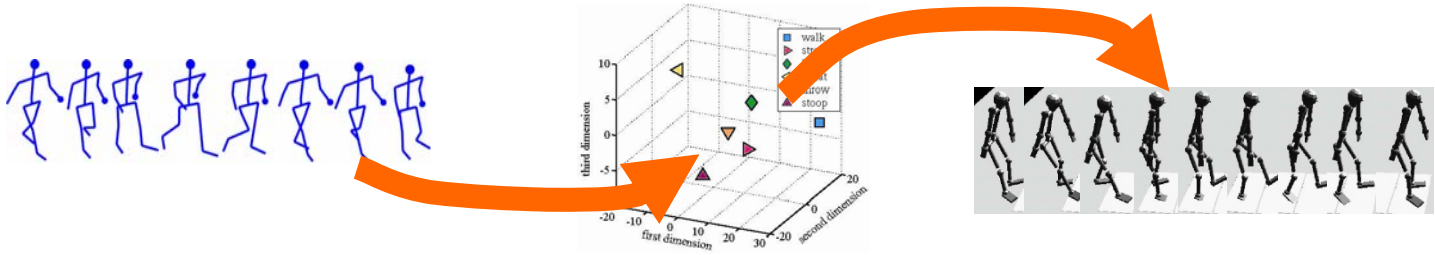
Related works

- Motion emphasis (modification, edit)
  - Interpolation / extrapolation (SIGGRAPH) [Bruderlin95][Rose98][Glardon04][Hoshino04]
  - Parameterization of motion[Matubara]
    - No relationship between symbol
- Symbolization of motion
  - RNNPB (A kind of Recurrent Neural Network) [Tani][Ogata]
    - Generation of arbitrary motions is difficult
  - Self organization map for motion[Okada]
    - Only periodic motions are discussed

Approach

稲邑 03～

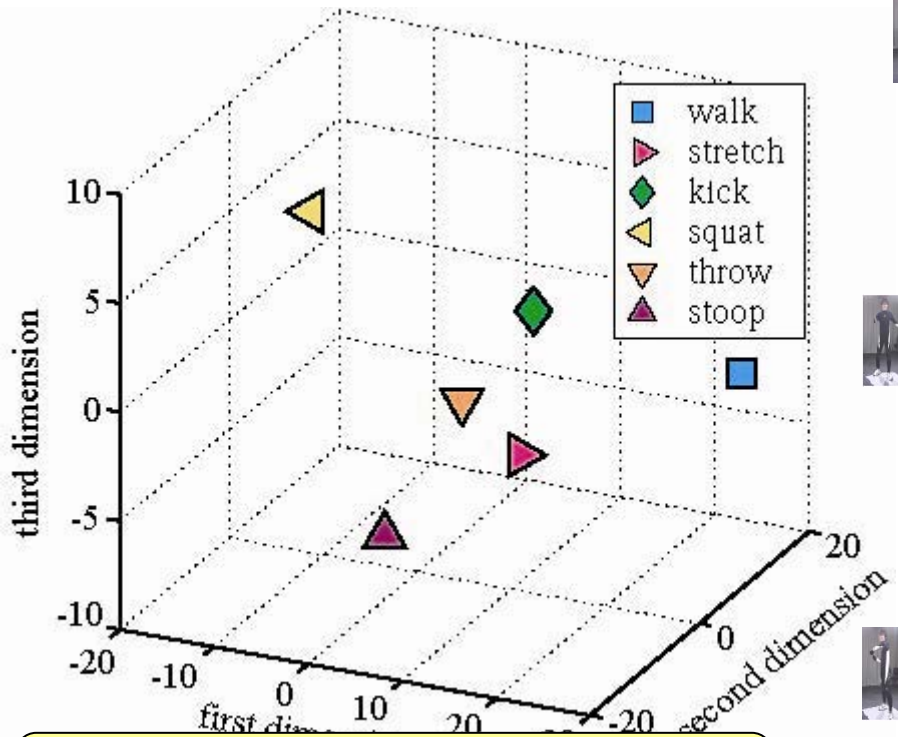
Mutual conversion model between sensorimotor patterns and symbols by proto-symbol space

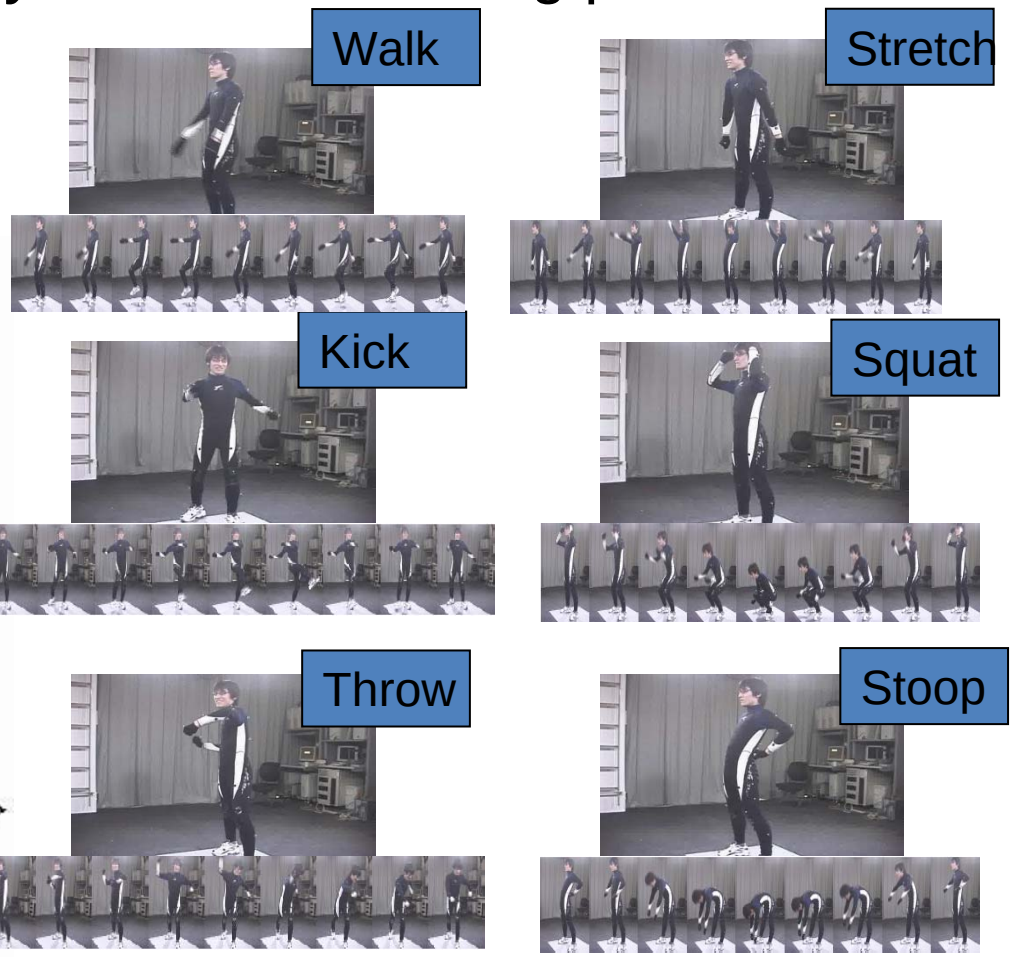


- Recognition, generation and abstract of patterns
  - Bi-directional model of recognition and generation
  - Imitation learning system for humanoid robots
- Motion primitive: Decomposition and composition
  - Association of sensory pattern from motion pattern
  - Imitation of unknown motion
- Conversion of patterns and symbols
  - Assignment of primitives using state point in phase space

Geometric representation of sensorimotor patterns

- Sensorimotor patterns are assigned as static points
- Configuration is defined by similarities among patterns
- Internal/External division generate various patterns

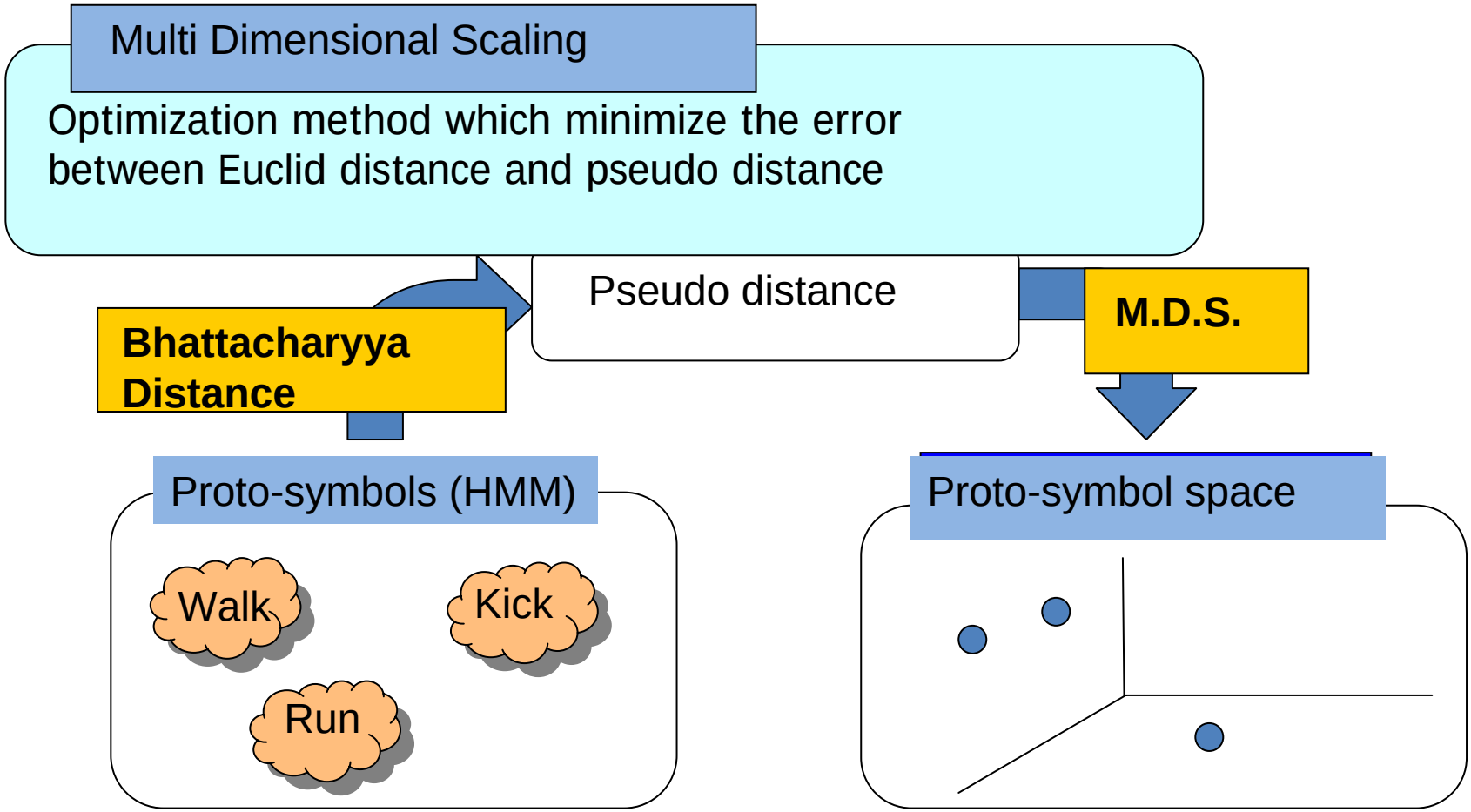




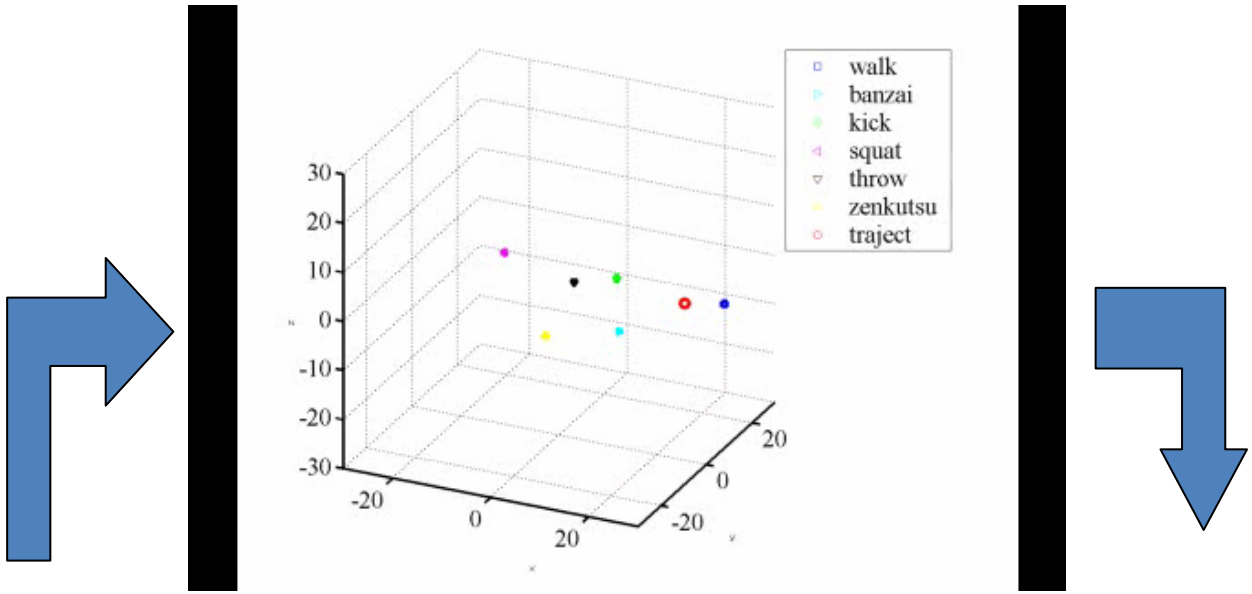
Symbol, discrete world ⟷ Physical, continuous world

Construction of proto-symbol space

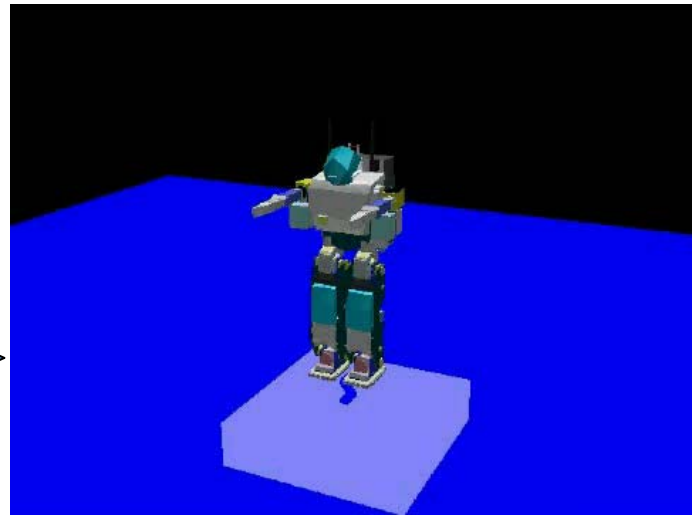
Placement in the Euclid space based on the pseudo distance between proto-symbols [Inamura ICHR03, IROS2006]



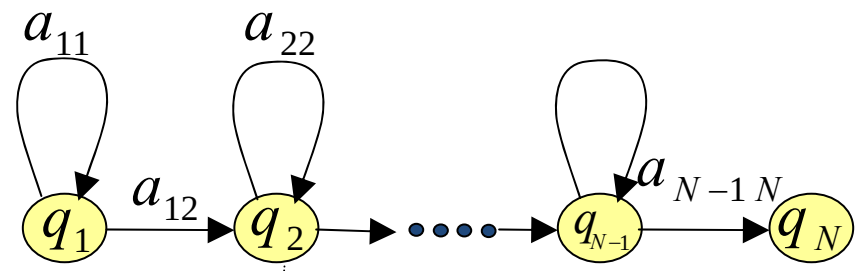
Realtime behavior imitation via symbol space representation



Not simple copy

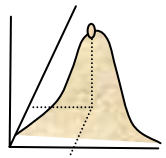


Motion abstract/recognition by Hidden Markov Models



$b_1(o)$ $b_2(o)$

$$b_i(o) = \sum_{j=1}^{M} c_{ij} \mathcal{N}_{ij}(o; \mu_j, \Sigma_j)$$



$O = \{o_1\ o_2 \cdots o_M\}$

abstraction

Joint angle vector θ

Parameter of HMM

$\lambda = \{a_{ij}, b_i(o)\}$

a_{ij} : state transition probabilities

$b_i(o)$: output probabilities

Proto-symbol

recognition

Using likelihood $P(O \mid \lambda)$ to recognize motion pattern O among the candidates of categories (proto-symbols)

Motion synthesis by proto-symbol synthesis

Inamura [IROS'08]

- Time-domain synthesis by Expected duration

$$s_i = \sum_{n=1}^{\infty} n(1-a_{ii})a_{ii}^{(n-1)} = \frac{1}{1-a_{ii}}$$

Calculation of the expected duration at node i

$$\hat{s}_i = \sum_{j=1}^{m} c_j s_i^{(j)}$$

Expected duration at node i of the synthesized HMM with the ratio of $c_1, \cdots, c_m$ using m HMMs

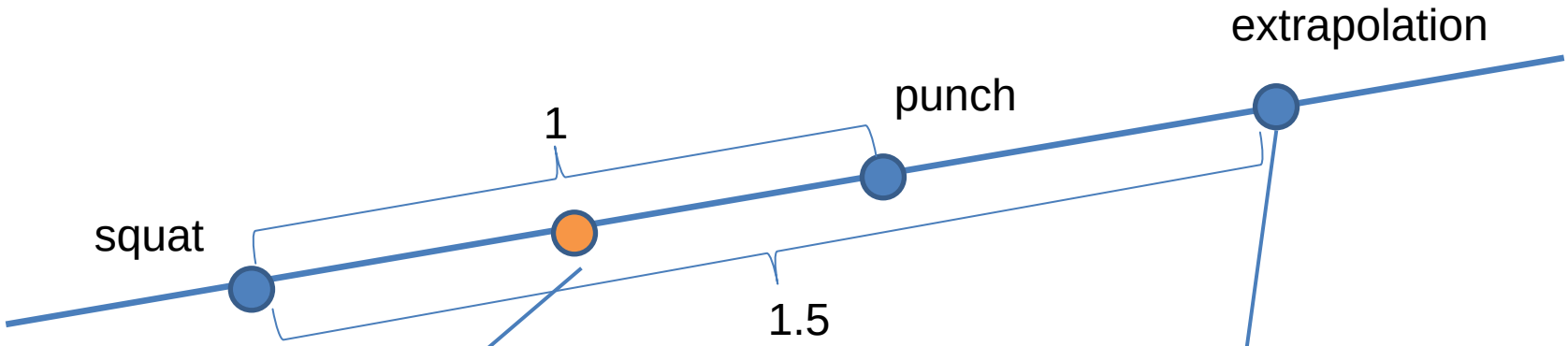
- Space-domain synthesis by Gaussian

$$b_i = N(\mu_i, \sigma_i)$$

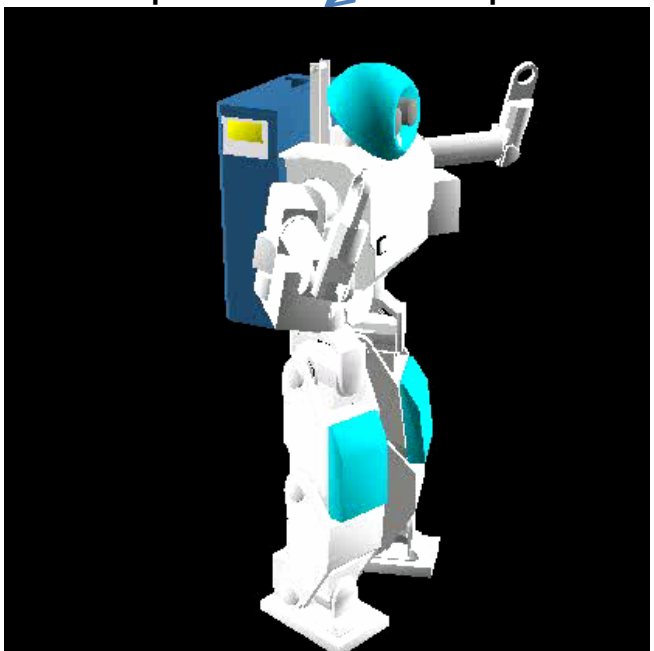
Output probability is modeled by single Gaussian

$$\hat{\mu}_i = \sum_{j}^{m} c_j \mu_i^{(j)} \quad \hat{\sigma}_i^2 = \sum_{j}^{m} c^2{}_j \sigma_i^{(j)2}$$

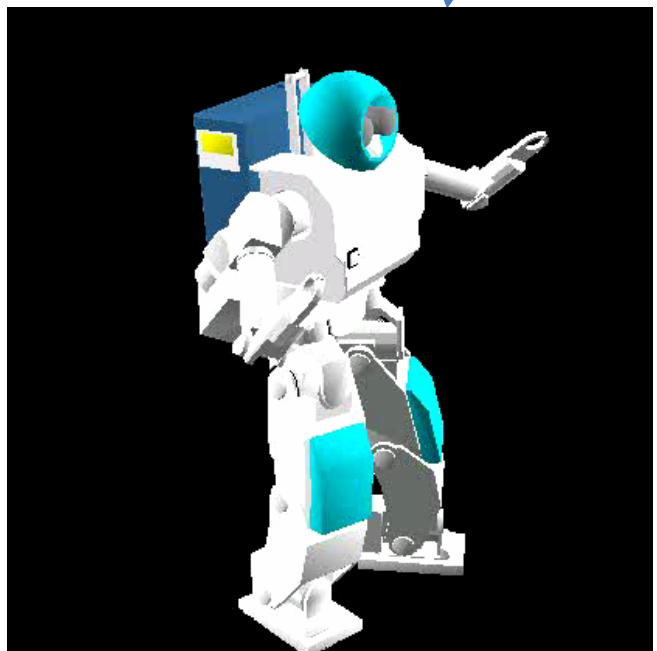
Mean vector and covariane matrix should be the target of interpolation/extrapolation



Interpolation
0.5 * punch + 0.5 * squat



Extrapolation
From Squat to punch

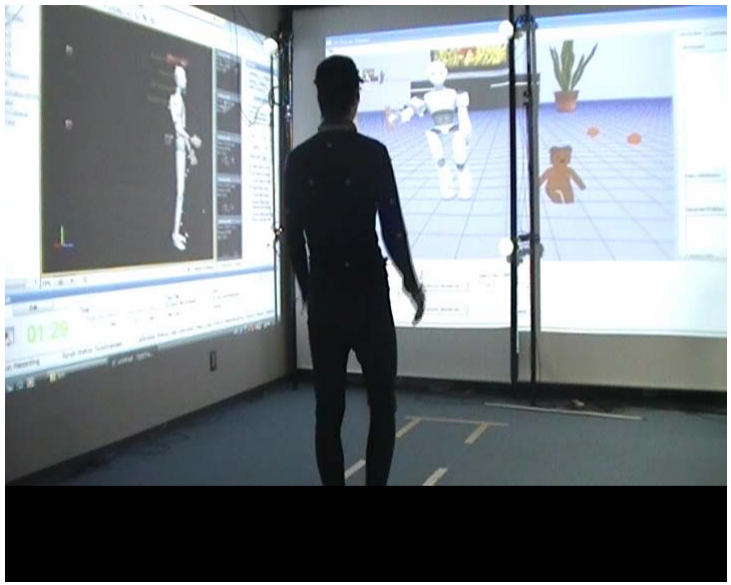


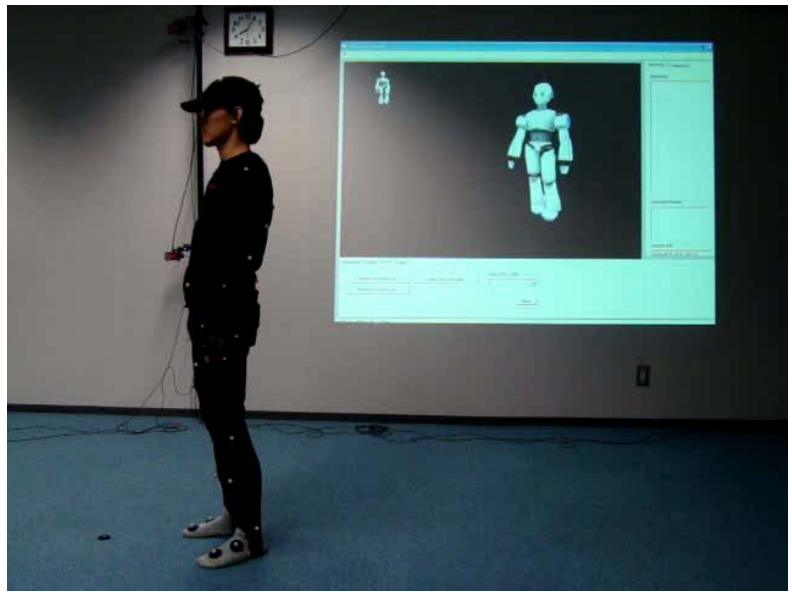
Inamura
IROS 2008

Experiment environment

Combination of immersive VR (surrounding display) and motion capturing system

Interaction system between virtual agent with dynamic whole body motion

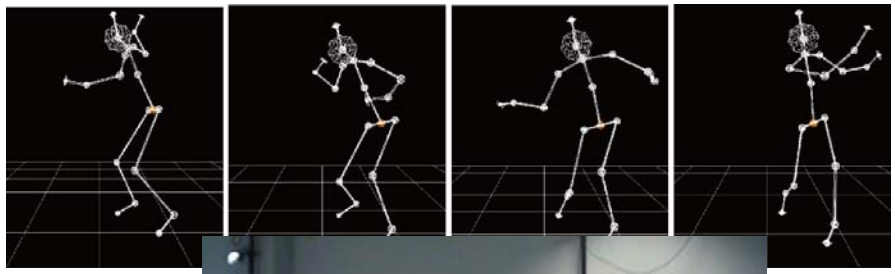




Apply to the coaching system

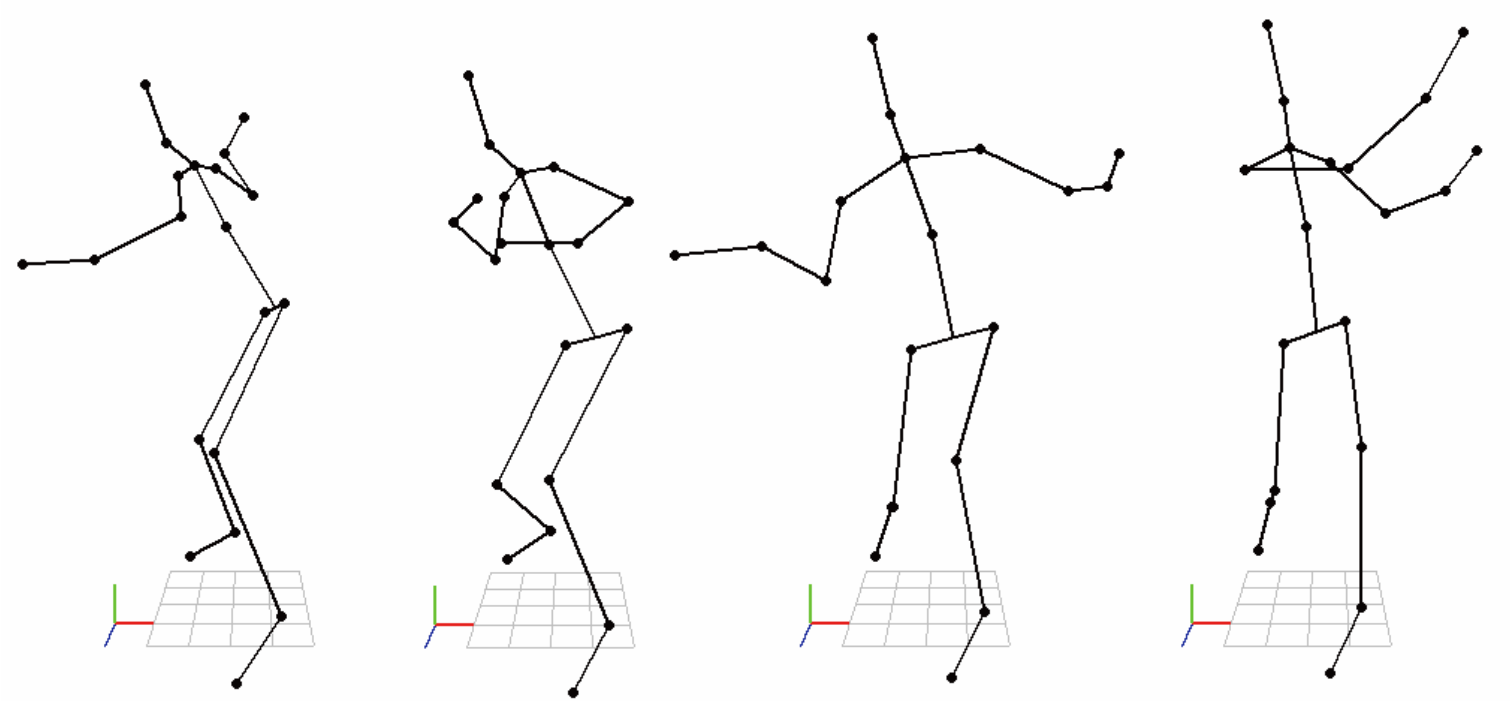
Experiment conditions

- Target motion: Swing motion of tennis
- 5 subjects (beginner of tennis)
- Output of HMM : joint angle of all joints
- Proto-symbol space is constructed from two motions:
  1) beginner's motion
  2) Target motion by expert



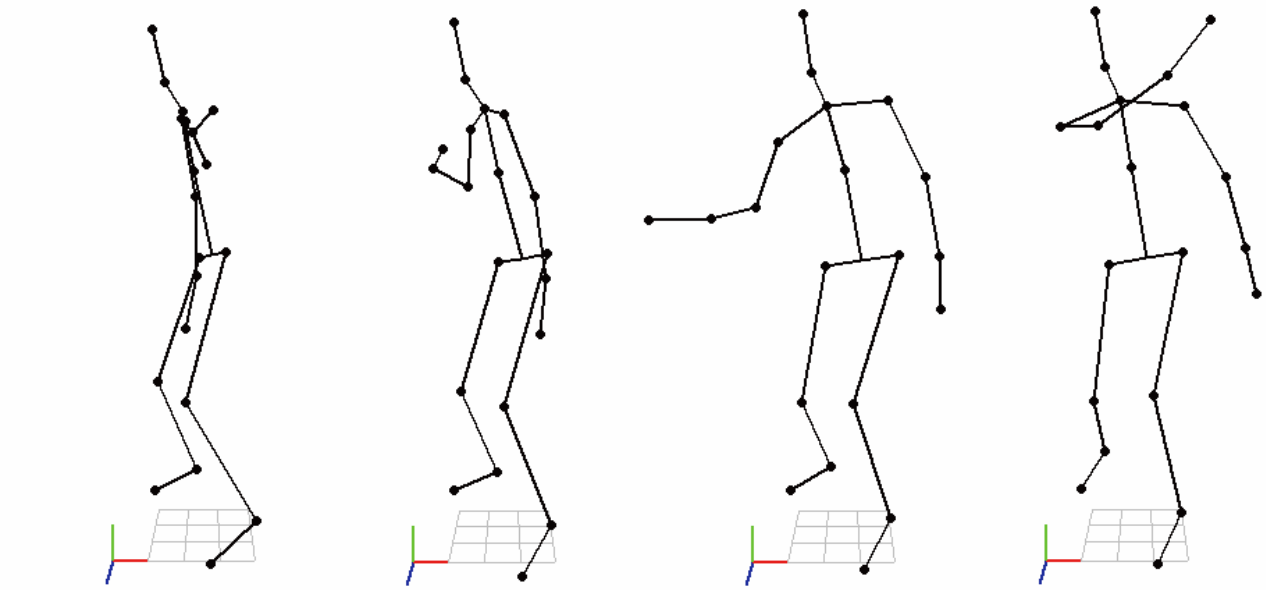
- 3 coaching strategies
  - Coefficient of emphasis
  - Verbal expression [on/off]

Target motion shown to the beginner



Performed motion by the player

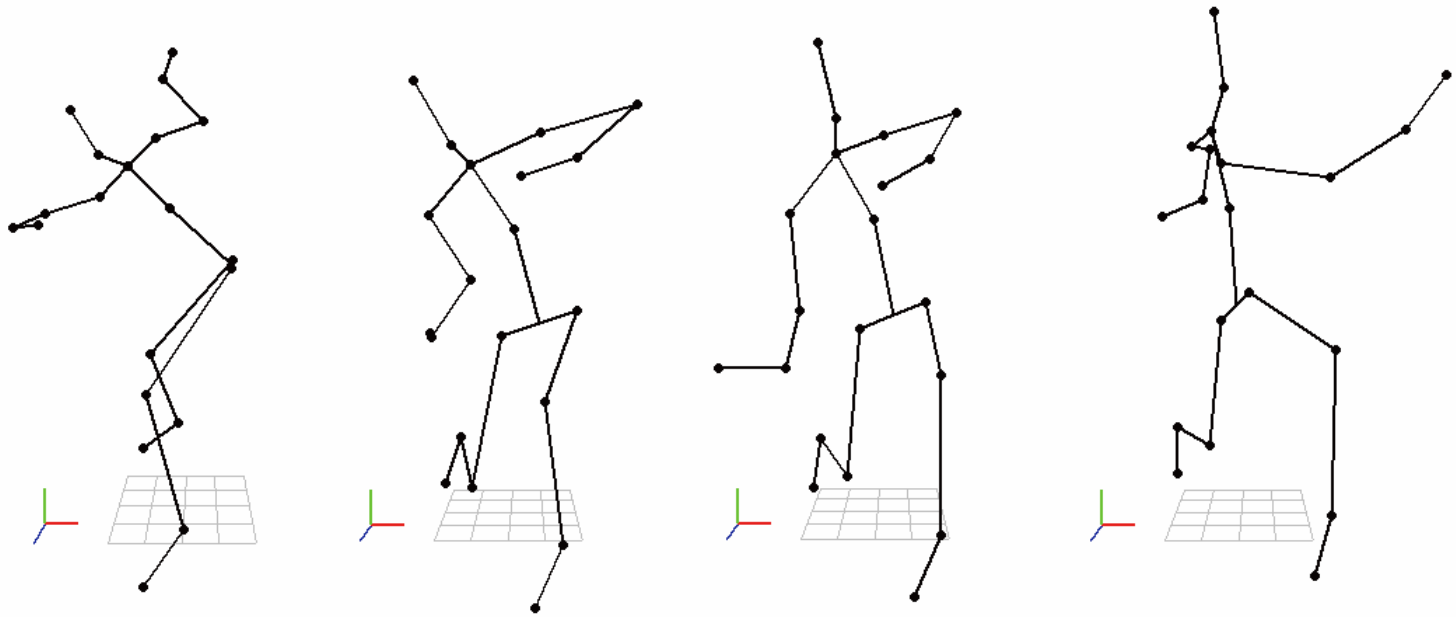
- Not good motion: knee is not bending, right elbow should be lower, and so on.

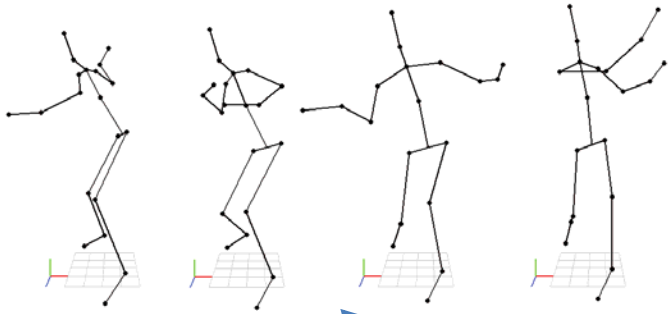


Generated emphasized motion by the coaching system

- -0.5 x [beginner motion] + 1.5 x [target motion]

“ not like the previous motion” “Please follow more like this motion”





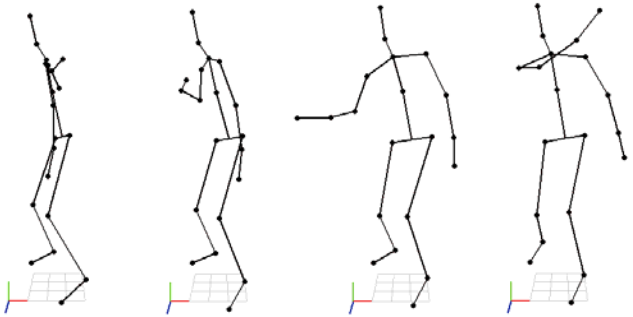
Emphasized motion

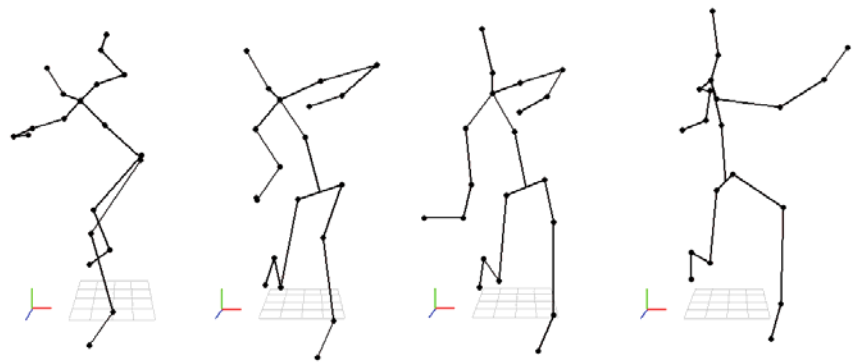
Target motion

1

Beginner motion

1.5





-0.5 x [bginner] + 1.5 x [target]

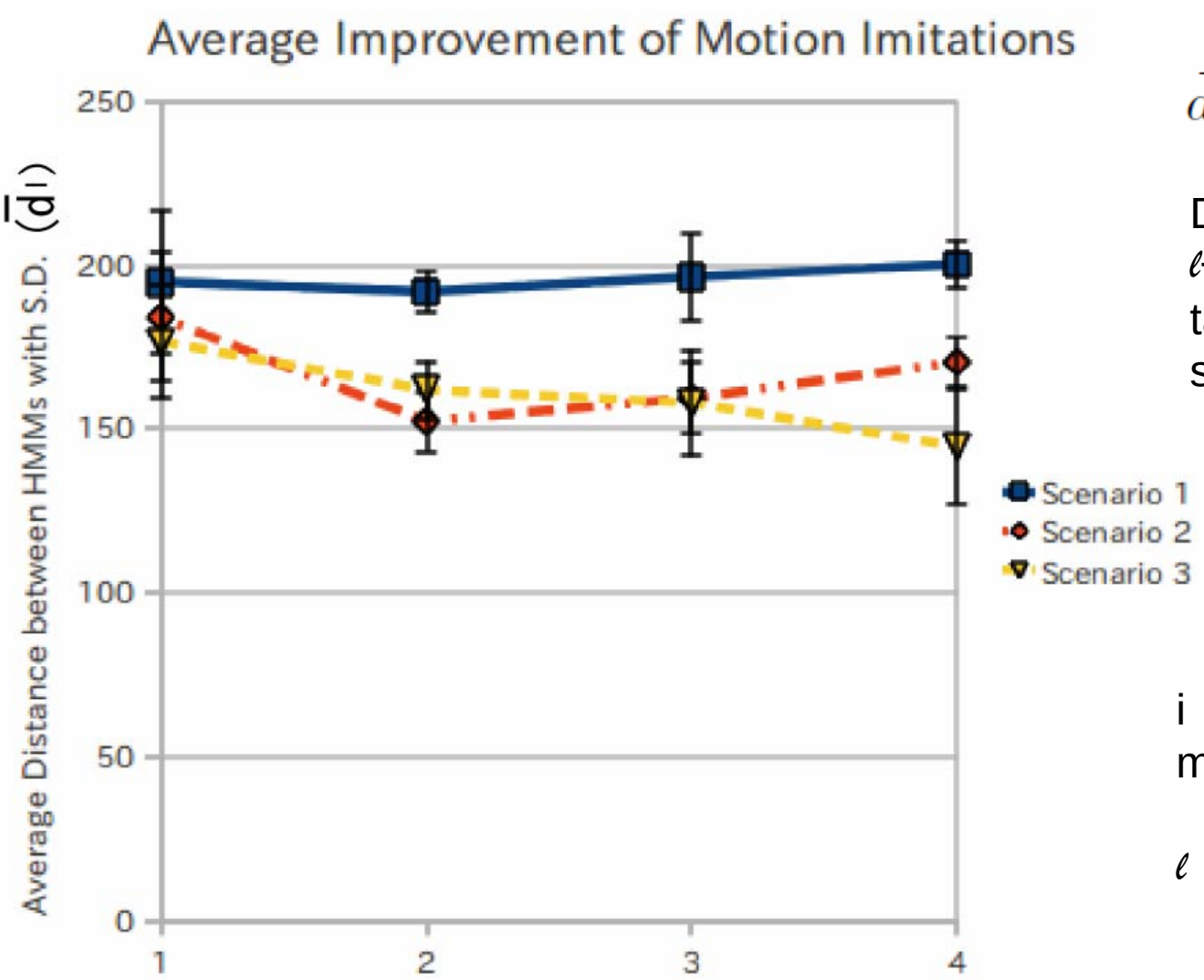
3 conditions for evaluation

1. **Only showing the target motion (without emphasis)**
   - Regardless of player's performance
   - α =1.0, no verbal expression

2. **Showing emphasized motion (without verbal exp.)**
   - Emphasized motion is shown to the player
   - α =2.0、without verbal expression

3. **Showing emphasized motion and using verbal expression**
   - If the error was bigger, verbal expression is added
   - α =2.0

Evaluation result (Ave. error of imitation)



$$\bar{d}_l = \frac{\sum_{i=1}^{m} d_{il}}{m}$$

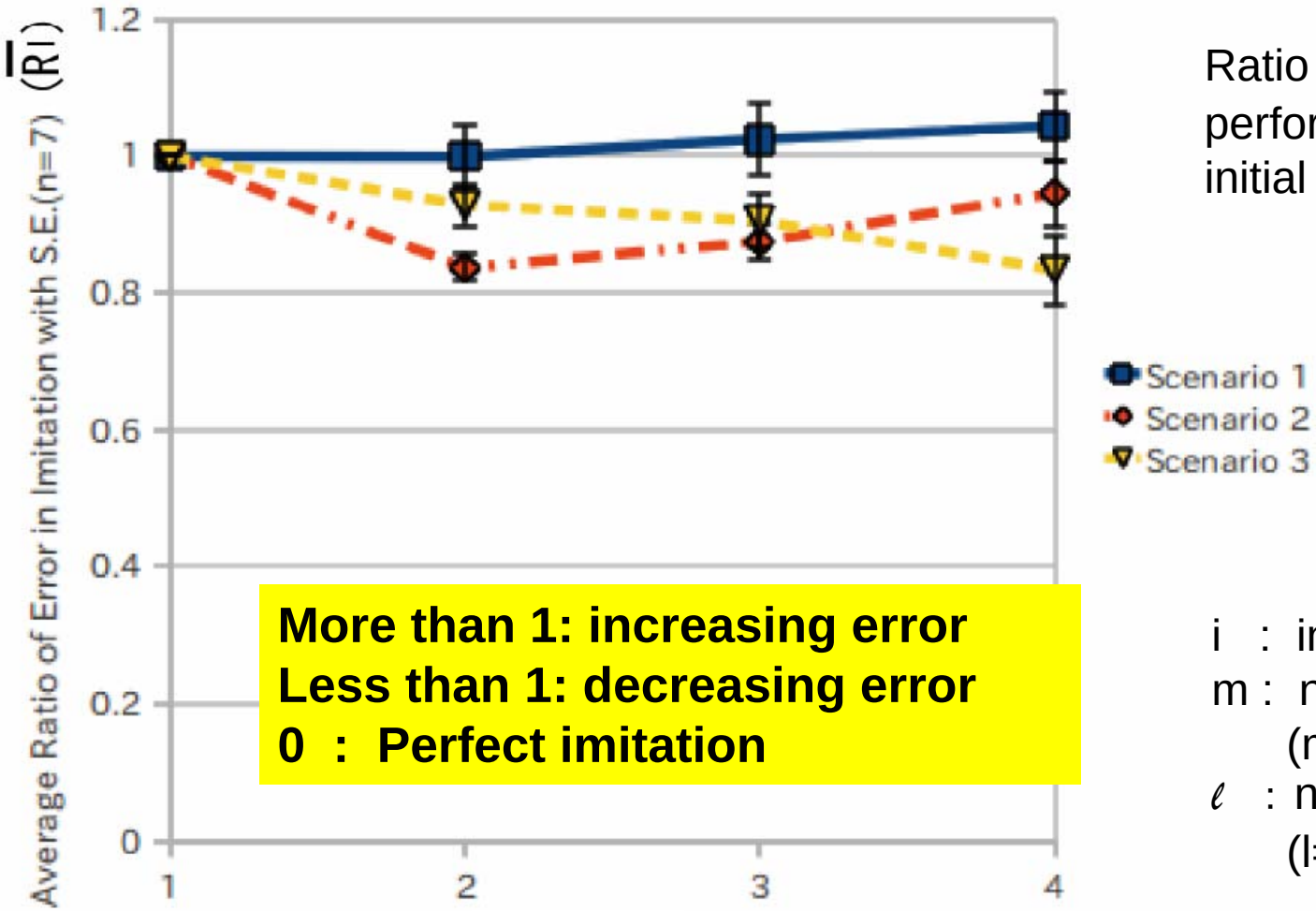
Distance (error) between ℓ-th performance and the target motion in proto-symbol space

i : index of subject
m : number of subjects (m=5)
ℓ : number of trials (l=1,2,3,4)

Evaluation (cont. error ratio)

$$\bar{R}_l = \frac{\sum_{i=1}^{m} \frac{d_{il}}{d_1}}{m}$$

Ratio of error of ℓ-th performance to the error of initial performance



More than 1: increasing error
Less than 1: decreasing error
0 : Perfect imitation

i : index of subject
m : number of subjects (m=5)
ℓ : number of trials (l=1,2,3,4)

Conclusion

- Proposal of coaching robot system that shows emphasized motion and uses verbal expression
- Motion emphasis and generation of verbal expression based on proto-symbol space
- Immersive VR system for coaching evaluation

Future works

- Mutual imitation learning between human and robot. Teach and learn in daily life env.