Presented at the 28th Annual Conference of the Japanese Neural Network Society (JNNS2018)
Oct 26th, 2018

# Visuomotor Associative Learning under the Predictive Coding Framework : a Neuro-robotics Experiment

Jungsik Hwang[1,2] and Jun Tani[2]

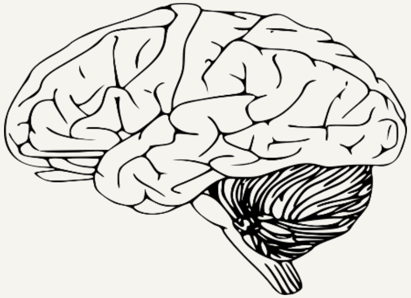
[1]Korea Advanced Institute of Science and Technology, Korea

[2]Cognitive Neurorobotics Research Unit, Okinawa Institute of Science and Technology, Japan

# Research Objectives

Build a **Cognitive Agent** which can

- Develop cognitive functions autonomously



**Embodiment**

"Learning from sensorimotor experience"

acquired from dynamic interaction with the world

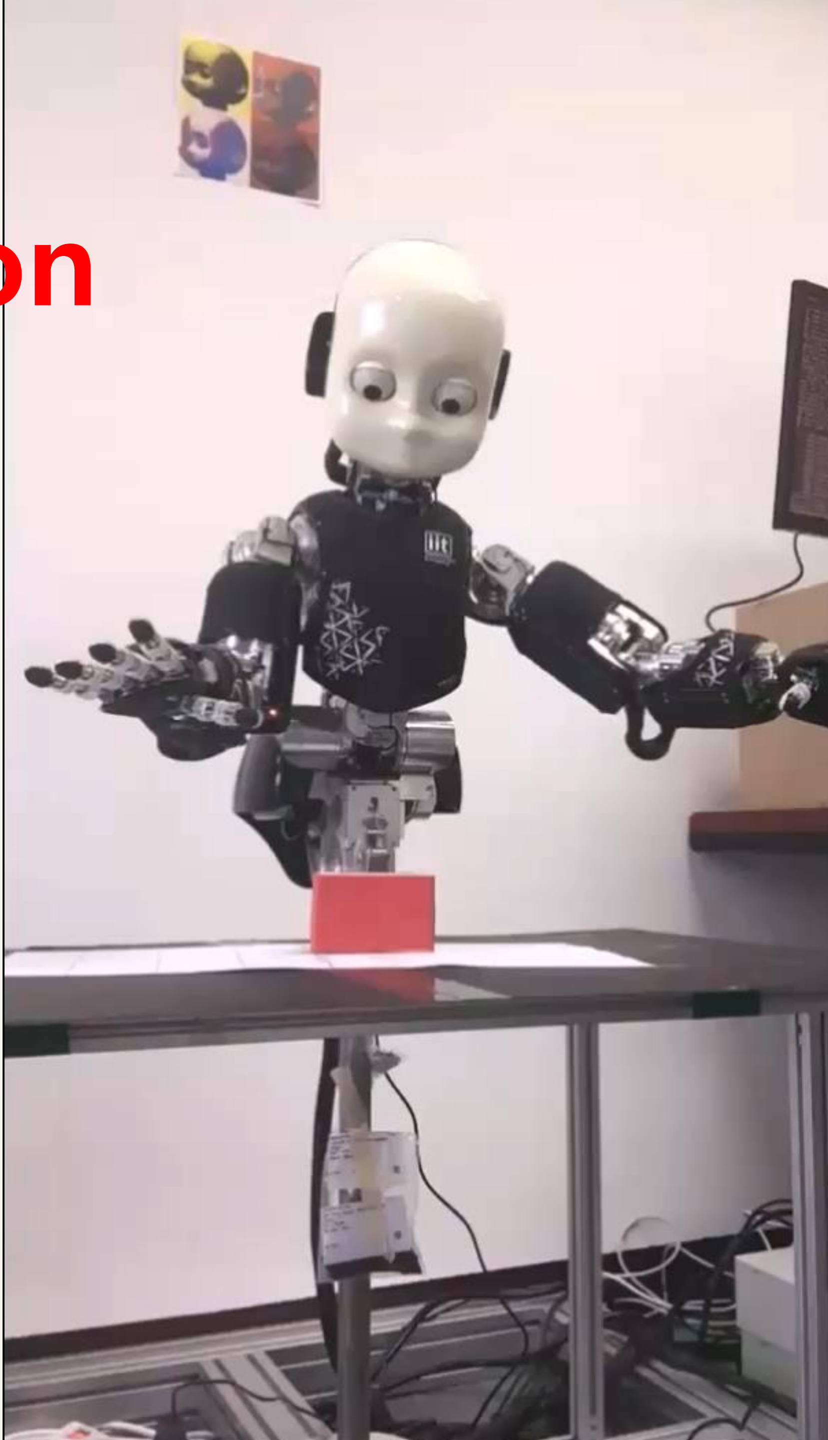
**Prediction**

"Brain = a Prediction Machine"

# Learning
# from Demonstration

Obtaining **sensorimotor experience**
By **showing a robot how to do it.**

Then, we make the robot **learn from** this **experience**.

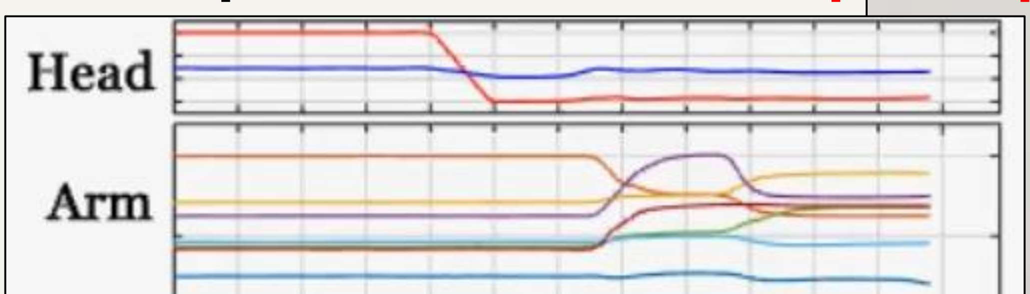


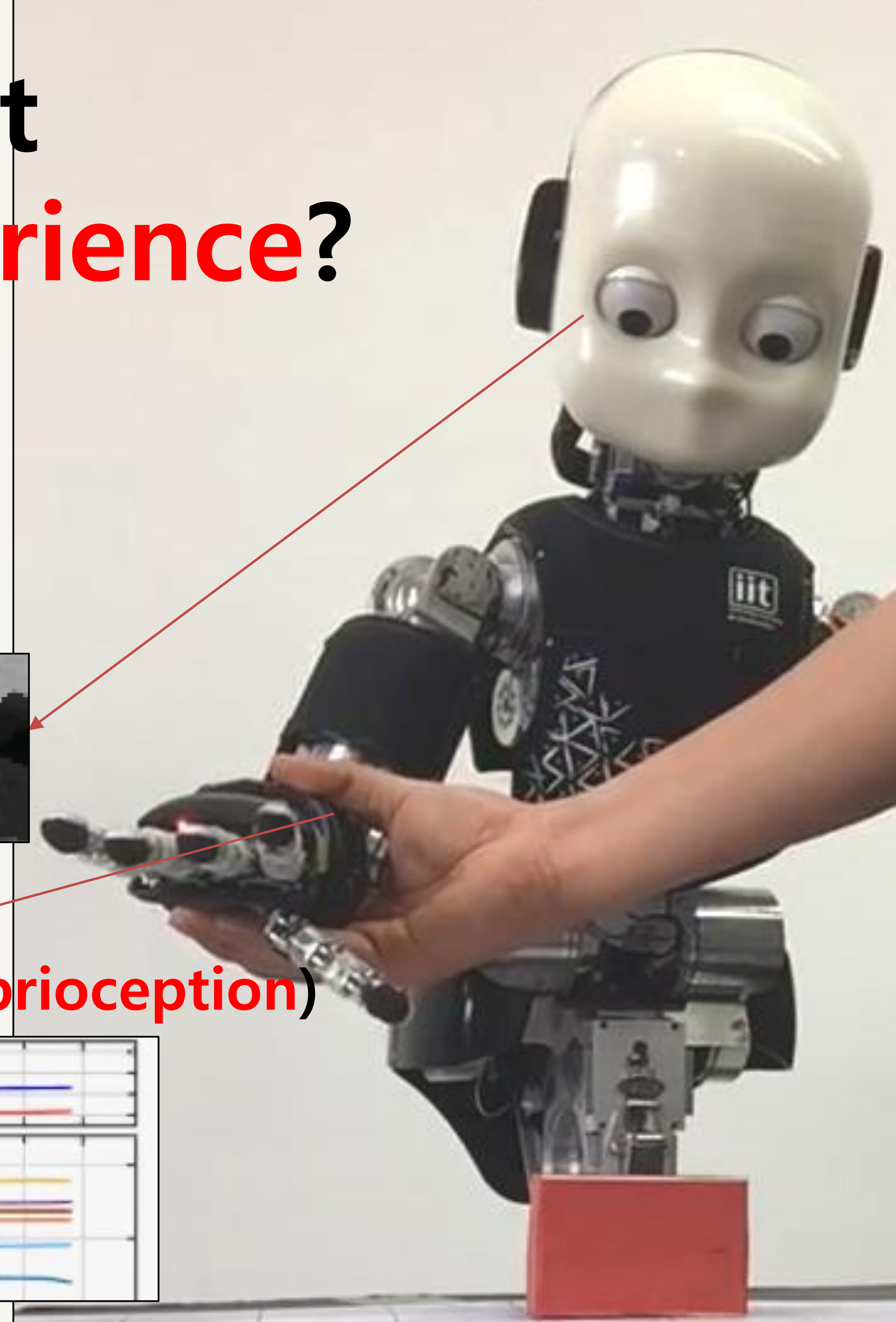
# How can a robot **learn from experience**?

Experience consists of

- **Visual images (Vision)**

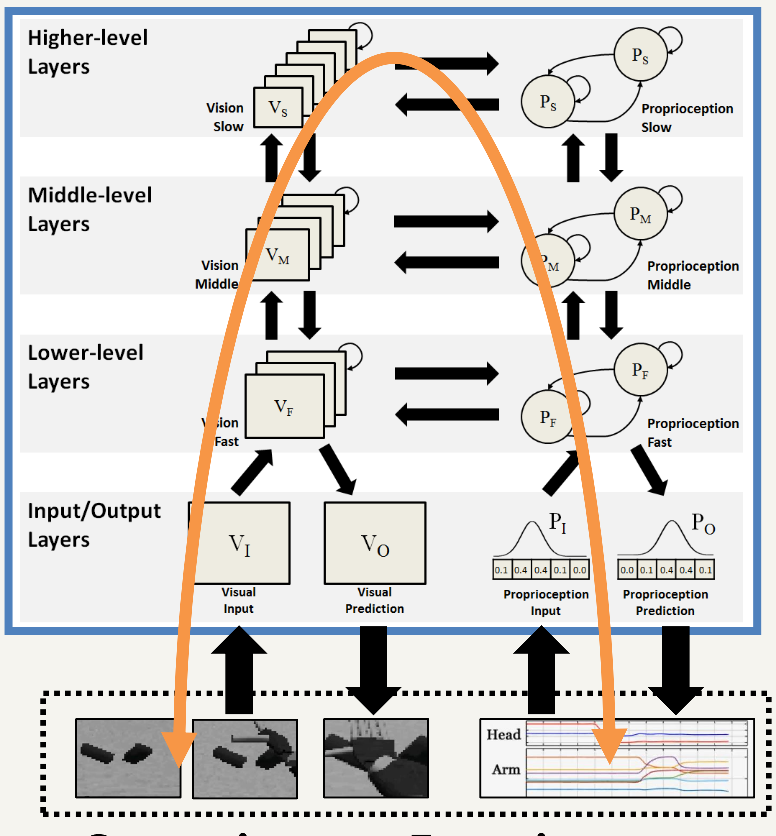
- **Joint position values (Proprioception)**

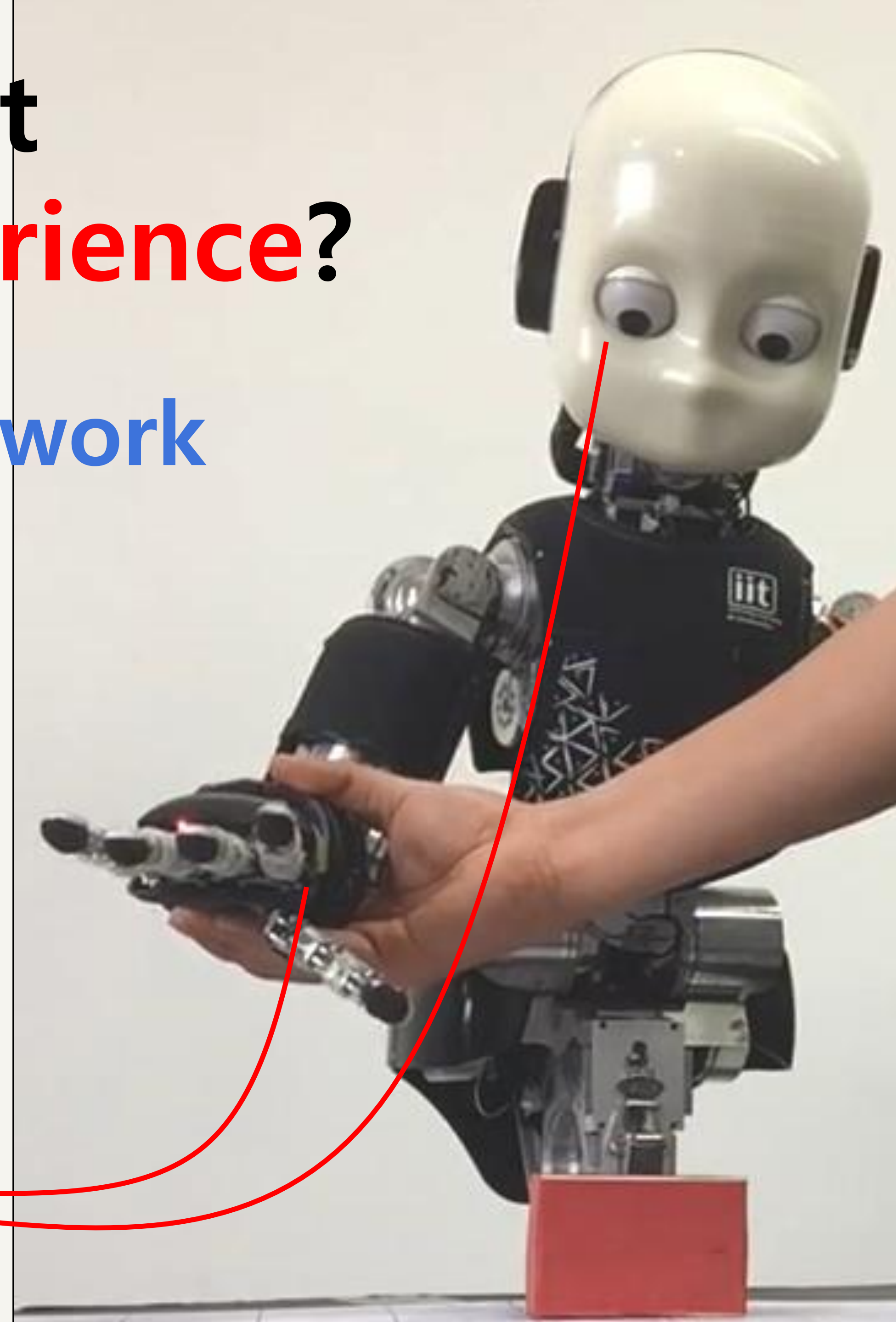


# How can a robot learn from experience?

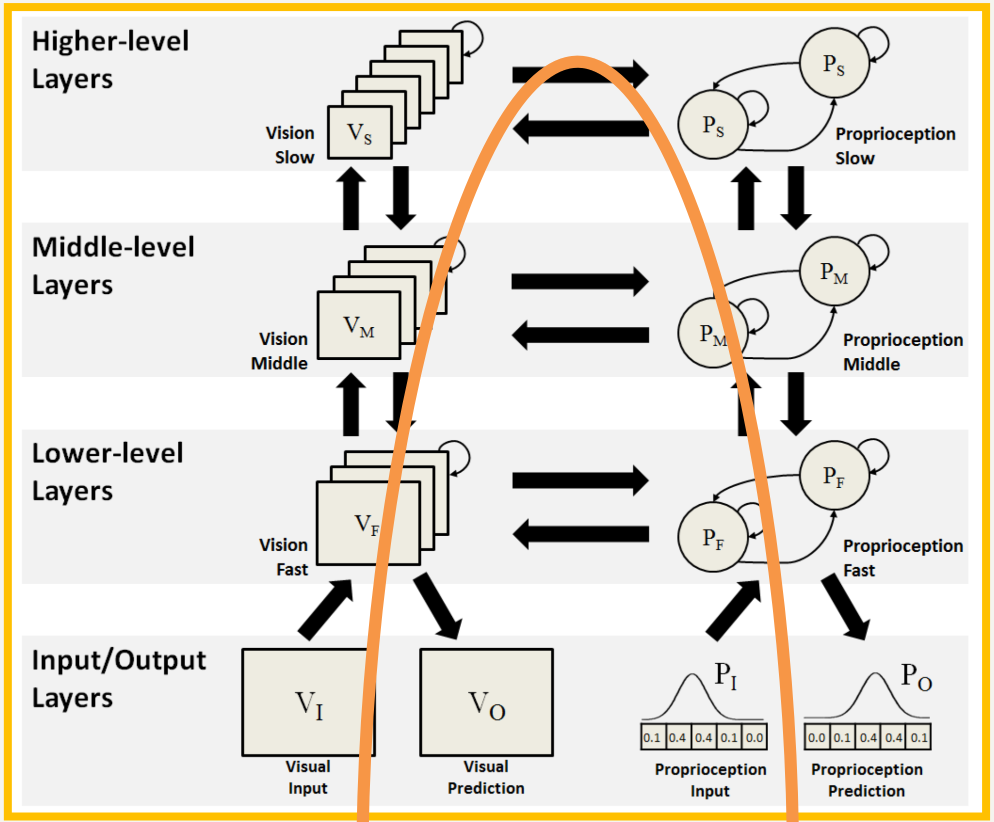
# Dynamic Neural Network



# Proposed Model

## Predictive Visuo-Motor Dynamic Neural Network (P-VMDNN)

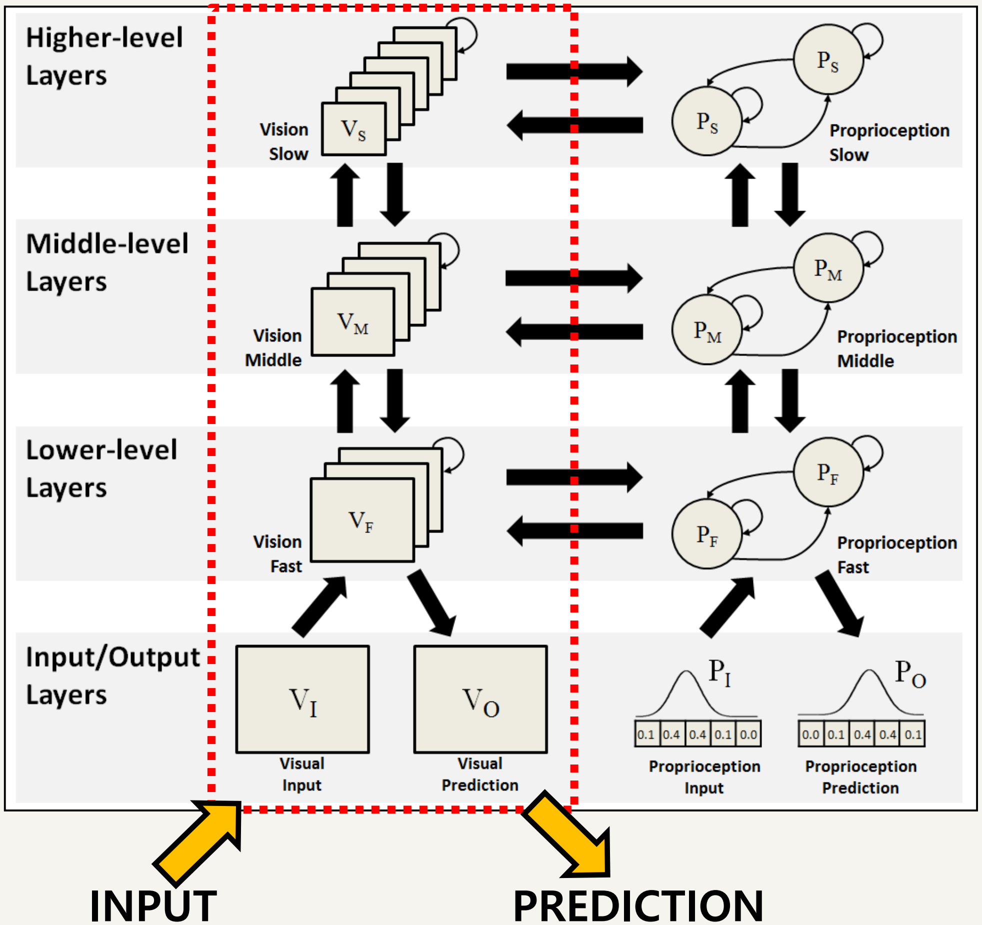




"Learning to **predict sensorimotor signals simultaneously** in an end-to-end manner"

# Visual Pathway

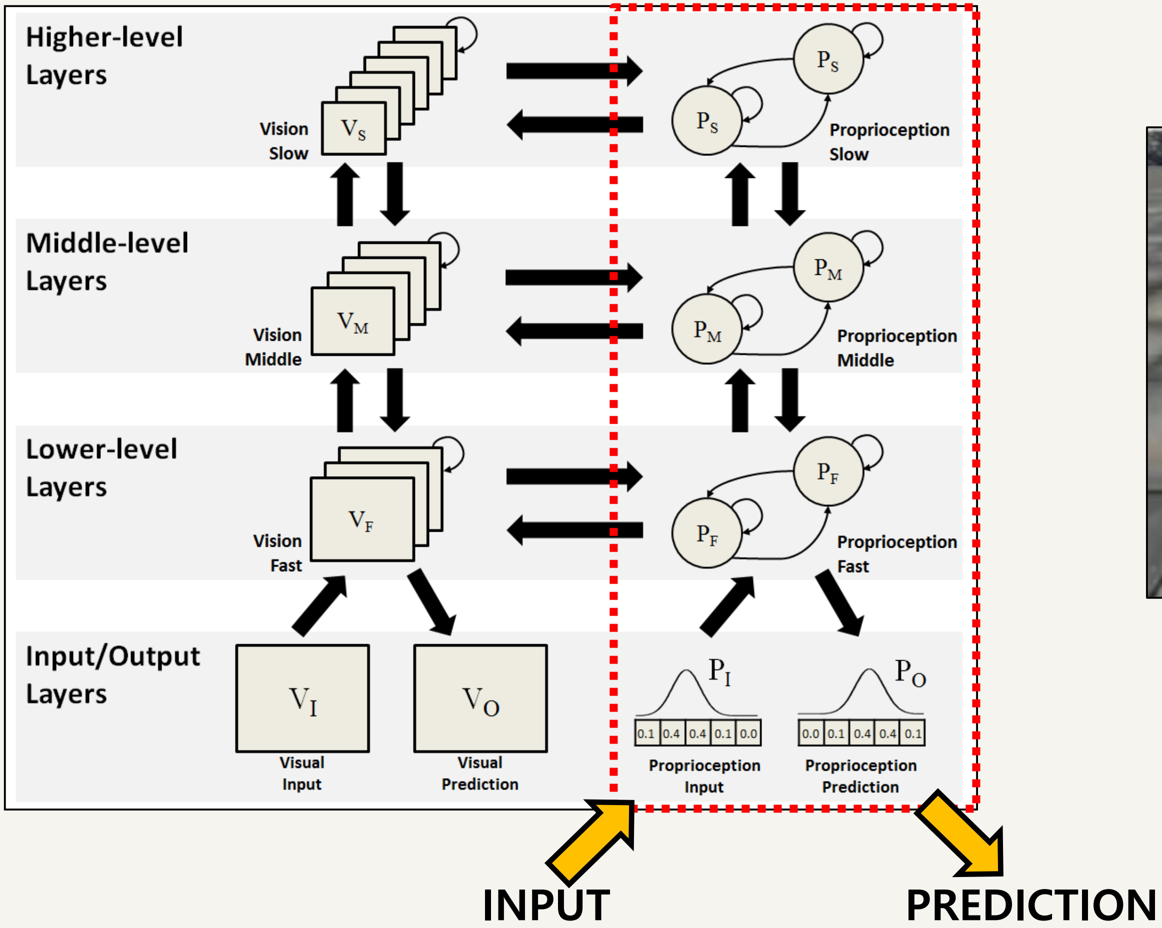
- Predicts pixel-level dynamic visual images
  - P-MSTRNN: Predictive - Multiple Spatio-Temporal Scales RNN




Example of Visual Prediction

# Proprioceptive Pathway

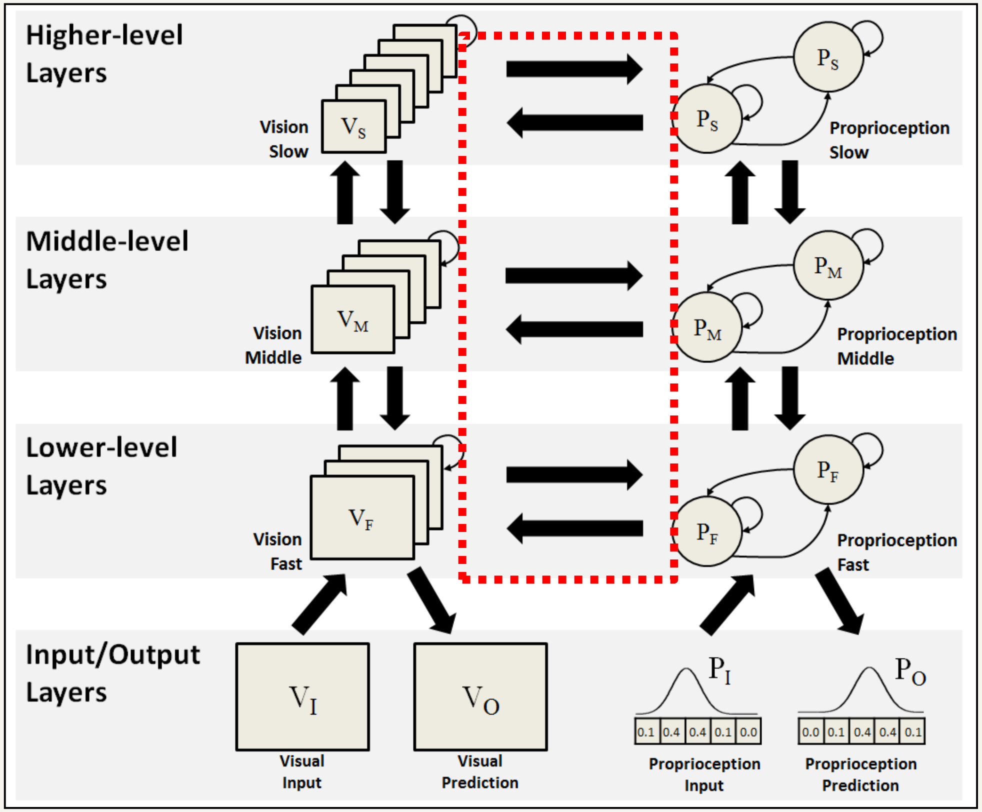
- Predicts robot's action (specified as joint positions)
  - MTRNN: Multiple Timescales RNN




Example of Action Generation

# Lateral Connections between 2 Pathways

- Bi-directional flow of visuomotor information

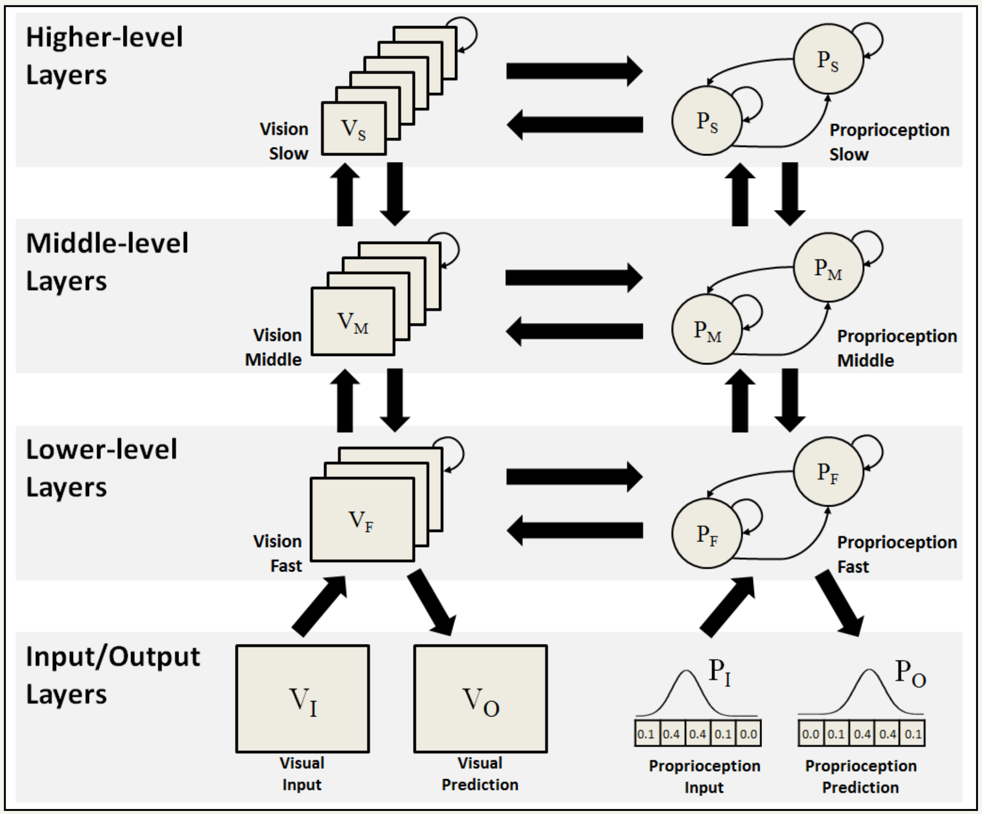



"Sensorimotor integration is a key part of the "intelligence algorithm" of the neocortex."

- Jeff Hawkins (2017)

# Key Characteristics

- **Temporal Hierarchy**
  - Imposing different constraints on neural activation




Larger Time Constants → Slowly-changing Neural Activity

$$u_i(\mathrm{t}) = \left(1 - \frac{1}{\tau}\right) u_i(t-1) + \frac{1}{\tau}\left\{\sum_j w_{ij} x_j(t)\right\}$$

"Emergence of Functional Hierarchy"

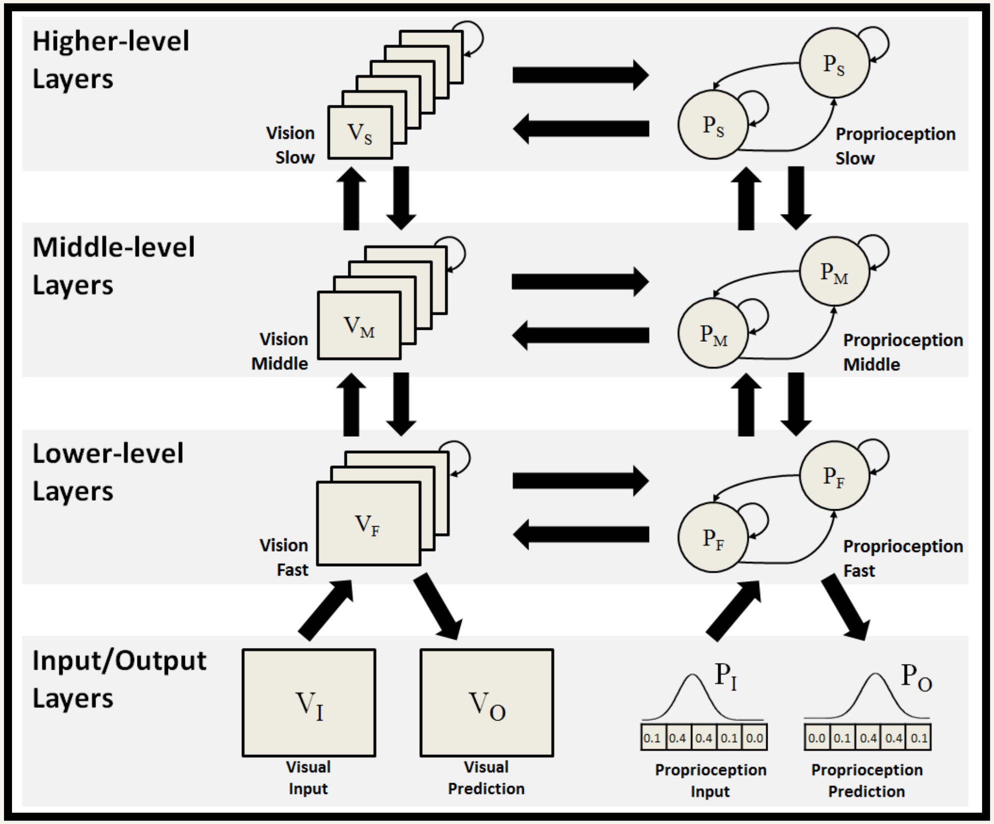
Smaller Time Constants → Fast-changing Neural Activity

# Experiment **Setting**

- Task: **Imitating** human gestures
  - 9 gestures x 3 human subjects
- Robot Platform
  - iCub simulator
  - Vision) 64 x 48 grayscale
  - Action) 10 DoFs

Simulator View

# Predictive Visuo-Motor Dynamic Neural Network




## Key Features

- Processing of Spatio-Temporal Patterns
- Coupling of Vision & Proprioception
- **Mental Simulation**
- **Prediction Error Minimization**

# 1 Mental Simulation

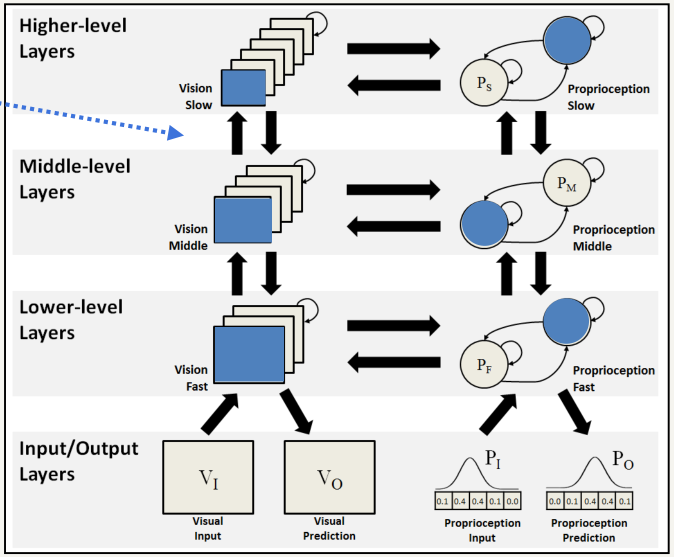
# Mental Simulation of Action

- **Mental Simulation**
  - Ability to imagine probable result of our actions
  - Important in social interaction
  - Needs “**What** to simulate”

- **In Our Experiment**
  - Ability to generate visuo-proprioceptive predictions with given intention
    - *Intention: specified as initial states
    - They are learnable parameters.

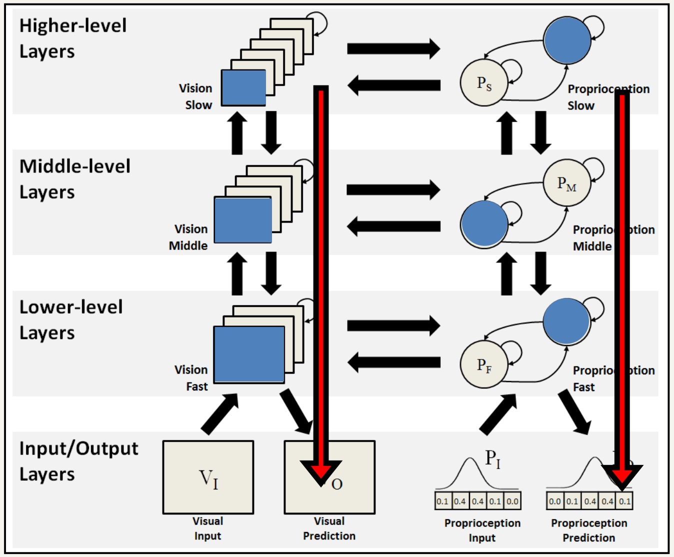
# Mental Simulation of Action

- **Mental Simulation in the Proposed Model**
  1. **Set the "Intention"**
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. Iterate (2) – (3)



# Mental Simulation of Action

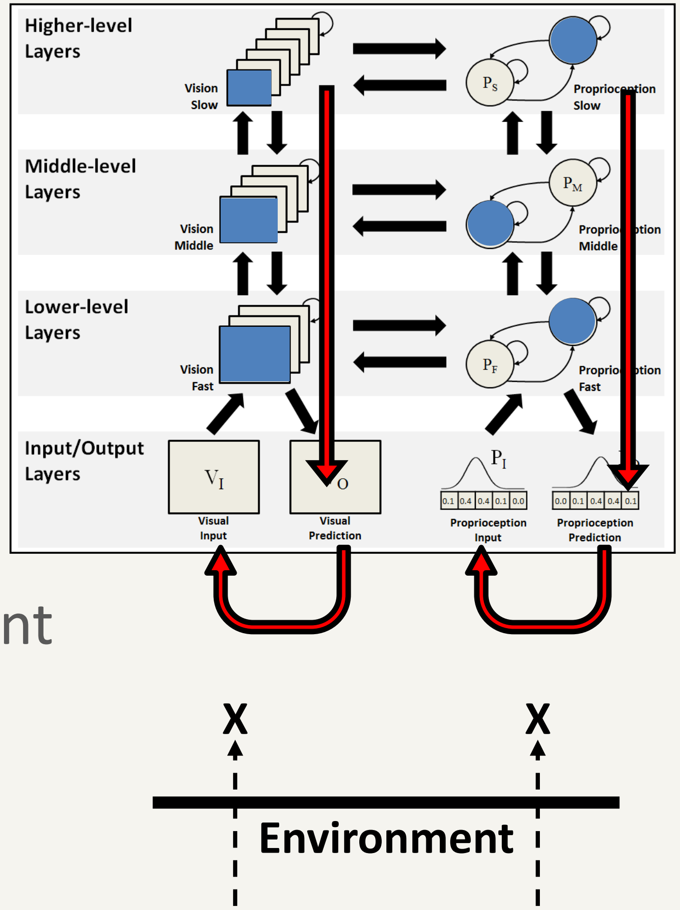
- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. **Generate Output**
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. Iterate (2) – (3)



# Mental Simulation of Action

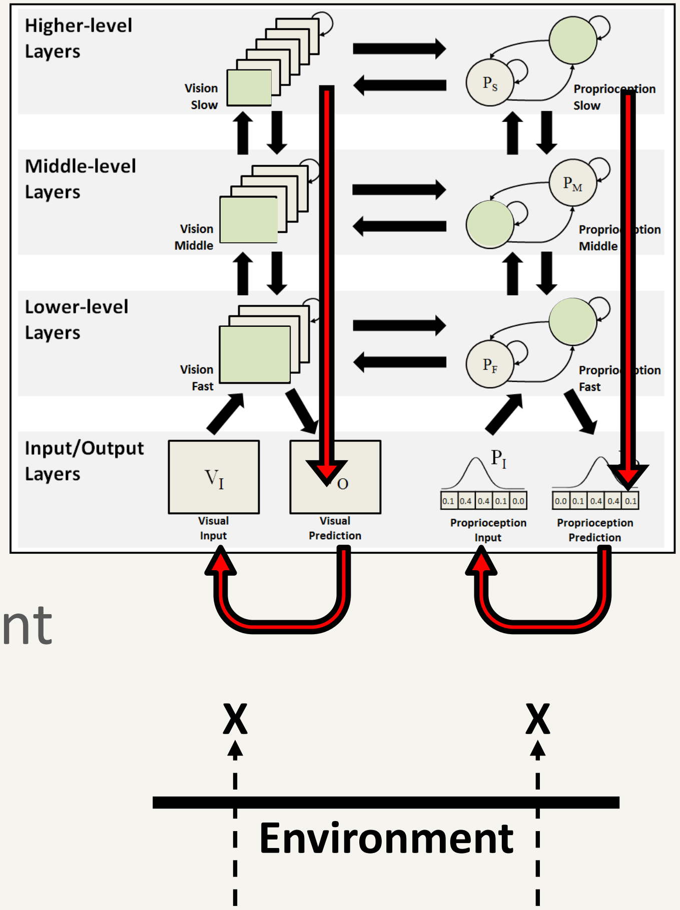
- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. **Feed Prediction Output into Input**
     - "Closed-loop Generation"
  4. Iterate (2) – (3)

- Without external input from environment
- **Only with given intention**



# Mental Simulation of Action

- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. **Iterate (2) – (3)**

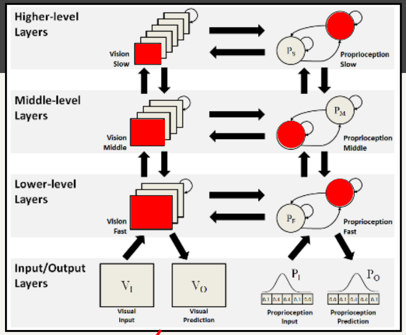


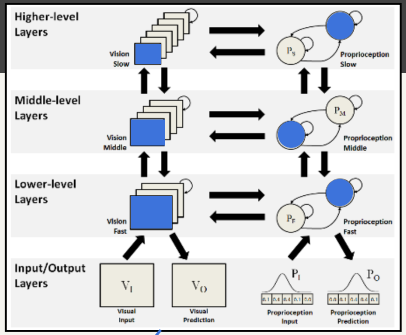

- Without external input from environment
- **Only with given intention**

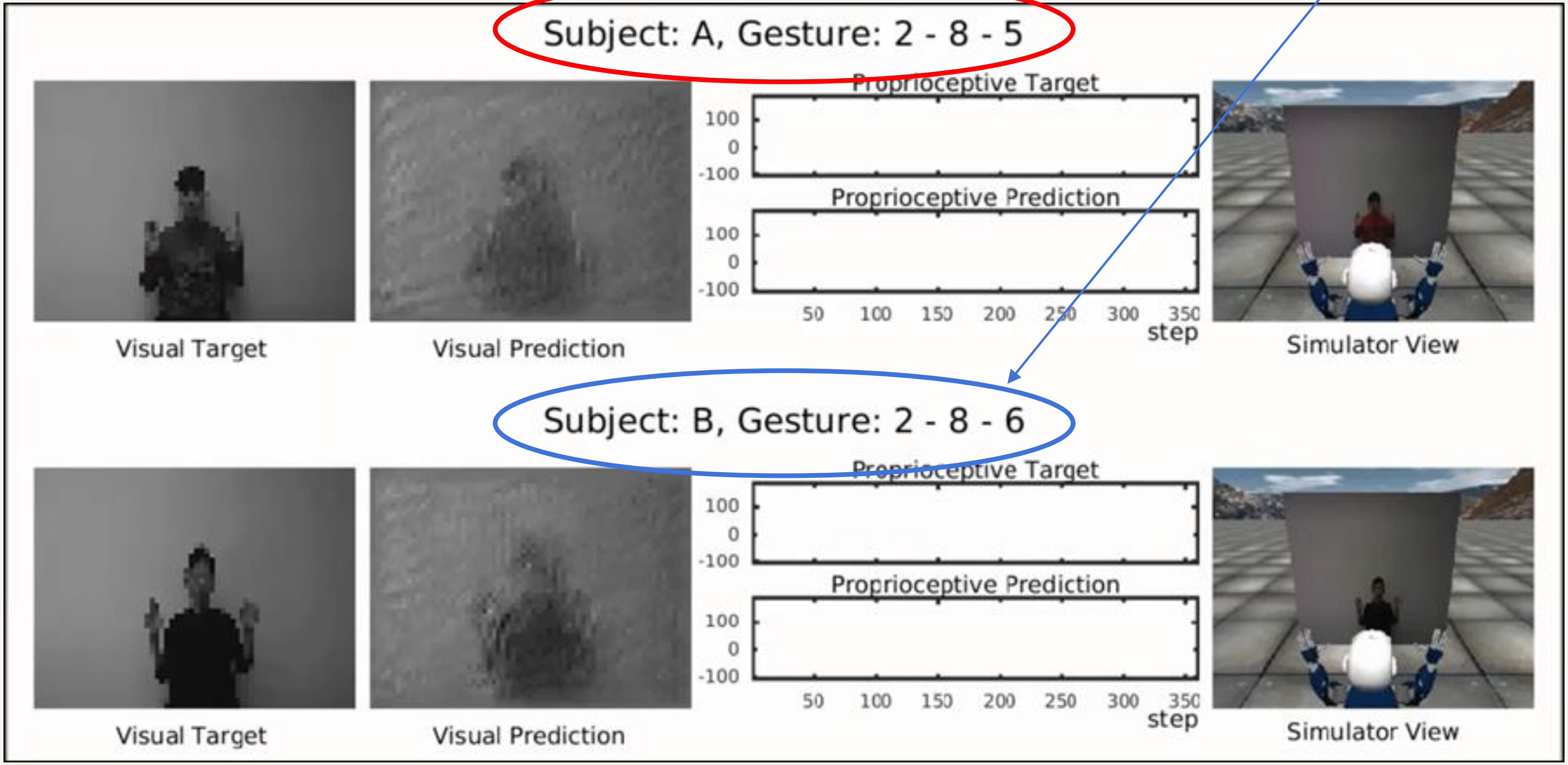
Result

# Mental Simulation of Action

- Setting **intention states**
  - at the onset of mental simulation
  - Obtained from training



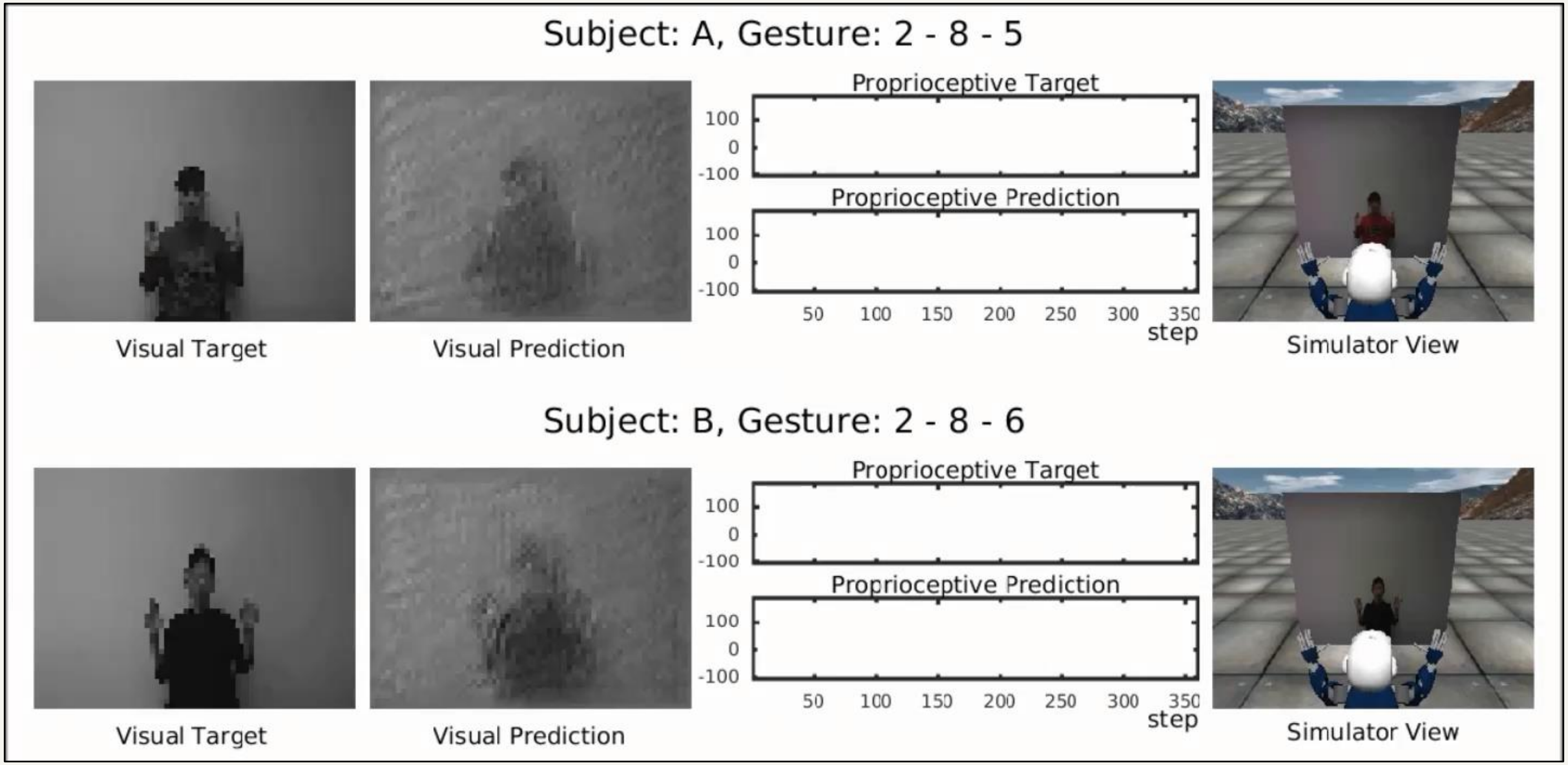



Result

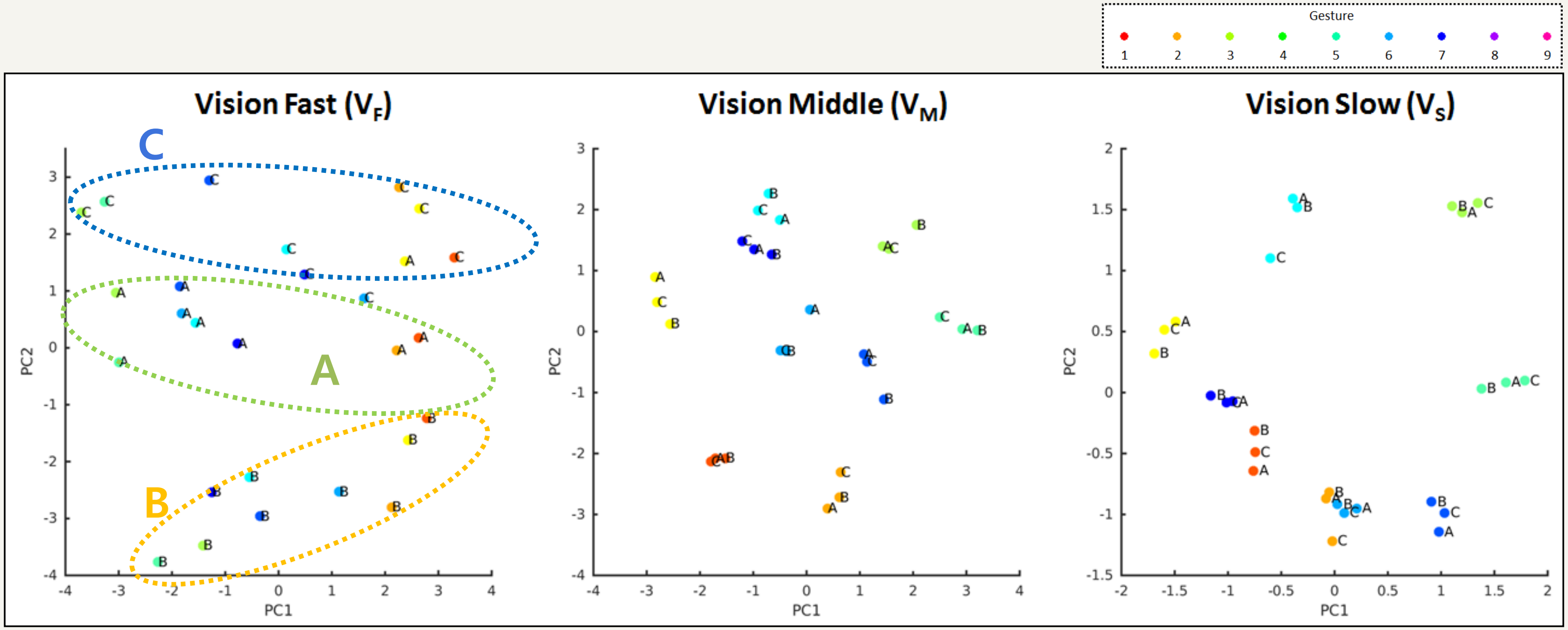
# Mental Simulation of Action

- With given 'intention', the model generated coherent visuo-proprioceptive patterns
  - Imagination without any input from the external world



# Self-organized Functional Hierarchy

- **Hierarchical representation** of visuo-proprioceptive patterns
  - **Abstract** information at **higher-level**: Type of gesture
  - **Specific** information at **lower-level**: Shape of specific human subject




Low-level Representation
(shape of a specific subject)

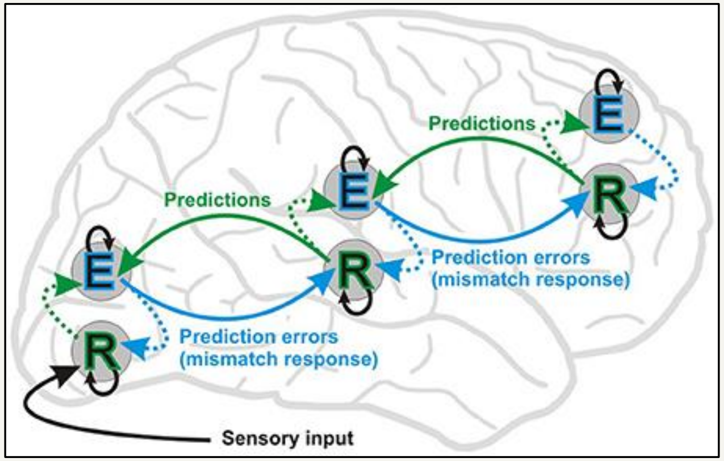
**Initial States obtained from Training**

High-level Representation
(type of the gesture)

# 2 Prediction Error Minimization

# Prediction Error Minimization

- Core of "Predictive Coding"
  - Recognizing intention from observation by minimizing prediction error
  - Account for MNS (Mirror Neuron Systems)
    - *Mirror Neurons: Activated while executing & observing an action

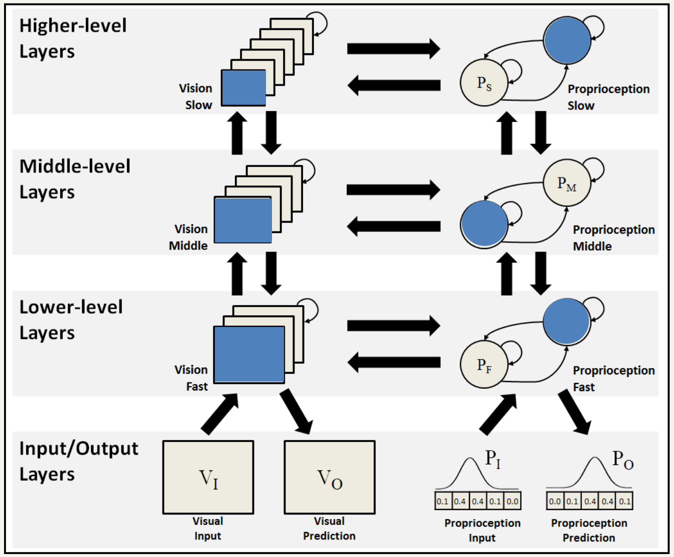


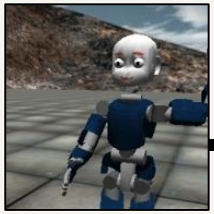

Predictive Coding Framework
Stefanics, et. al., (2014)

# Prediction Error Minimization

At each time step

1. Generate Visuo-Prop. Predictions
   - From Intention State (Top-Down Process)
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times



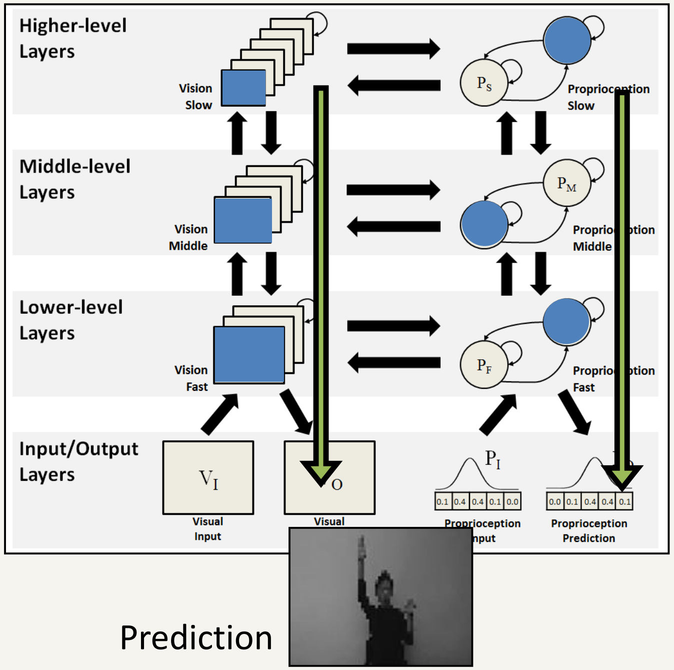



Environment

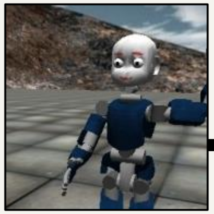
# Prediction Error Minimization

At each time step

1. **Generate Visuo-Prop. Predictions**
   - From Intention State (Top-Down Process)
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times




Prediction

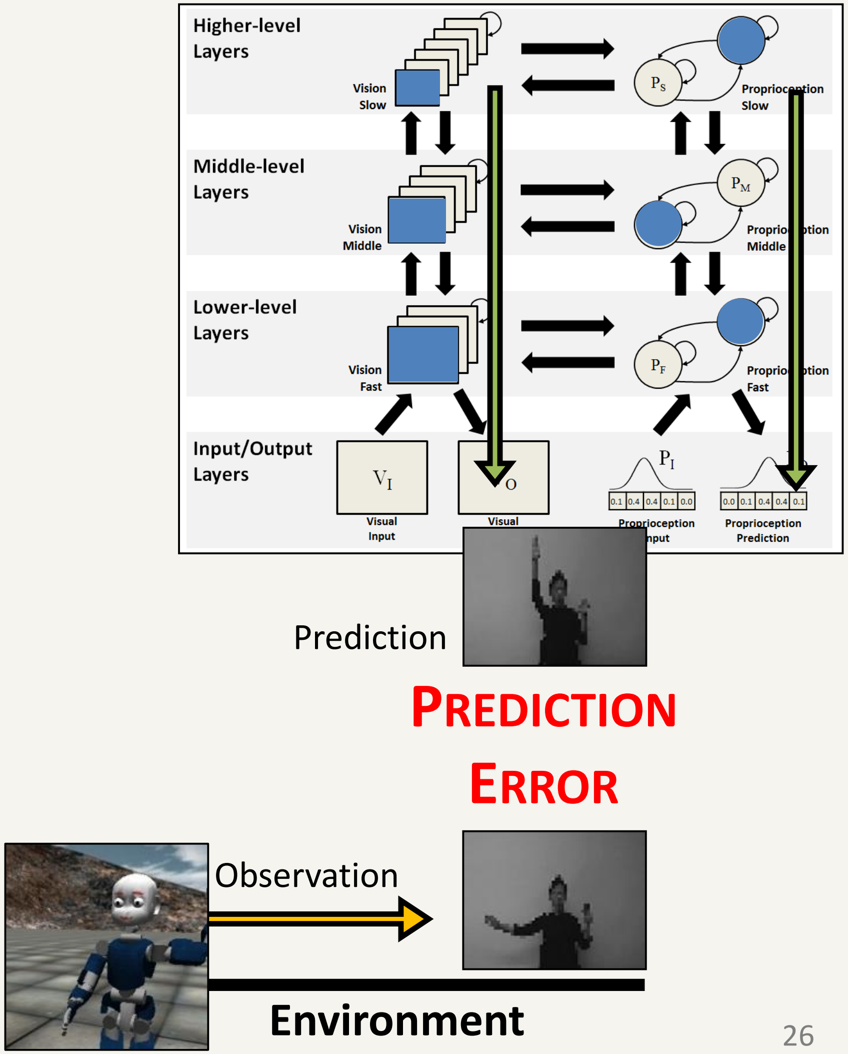


Environment

# Prediction Error Minimization

At each time step

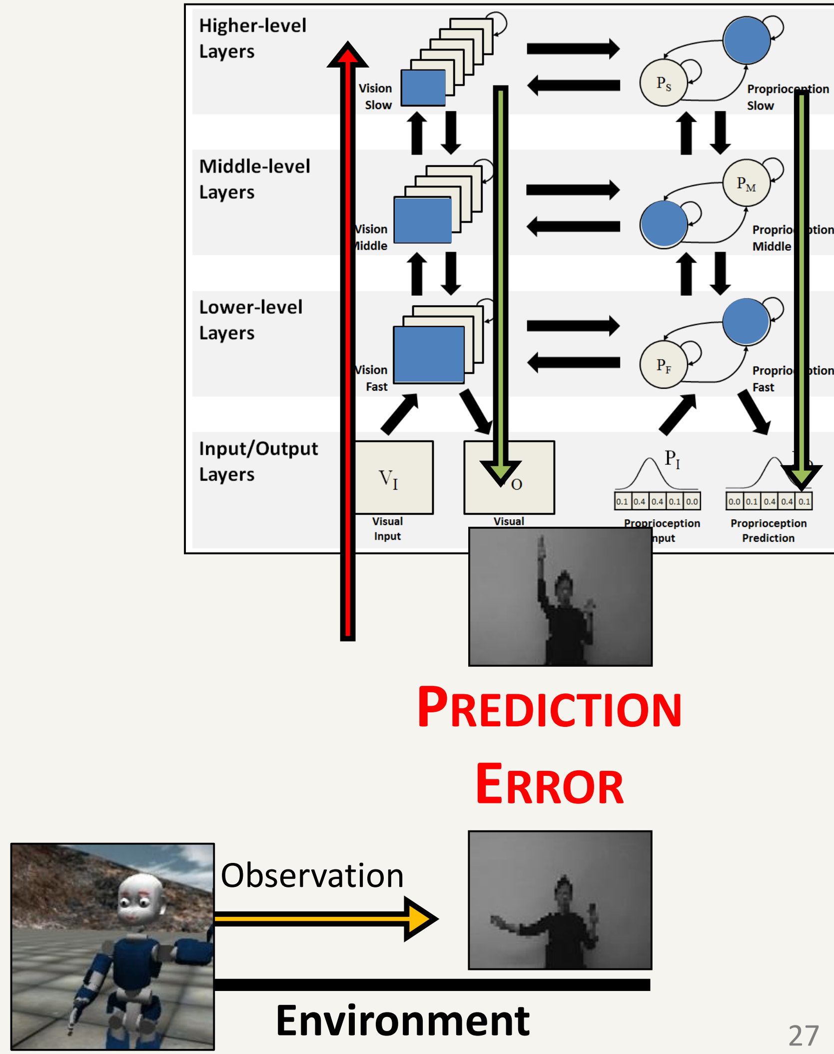
1. Generate Visuo-Prop. Predictions
   - From Intention State (Top-Down Process)
2. **Compute Prediction Error**
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times



# Prediction Error Minimization

At each time step

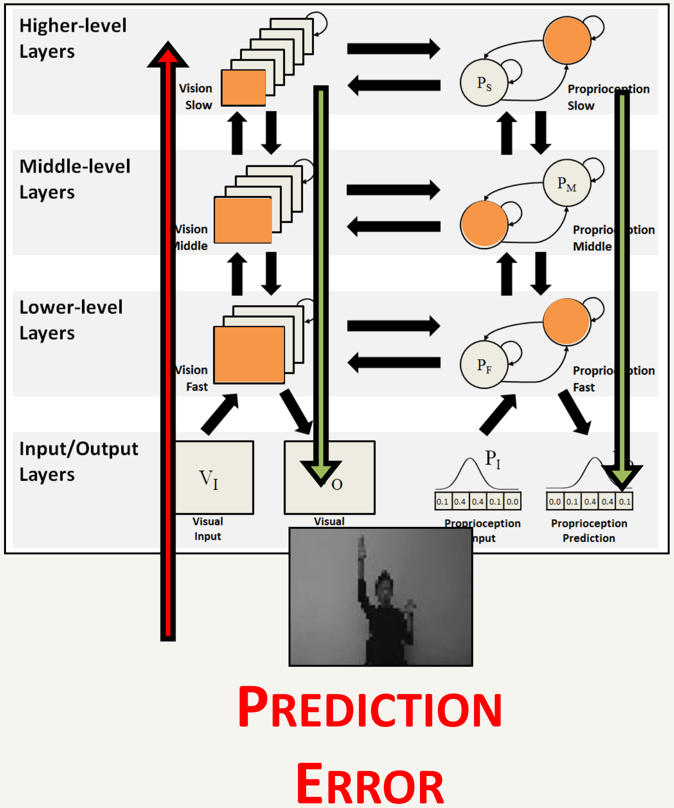
1. Generate Visuo-Prop. Predictions
   - From Intention State (Top-Down Process)
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. **Backpropagate Prediction Error** & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times

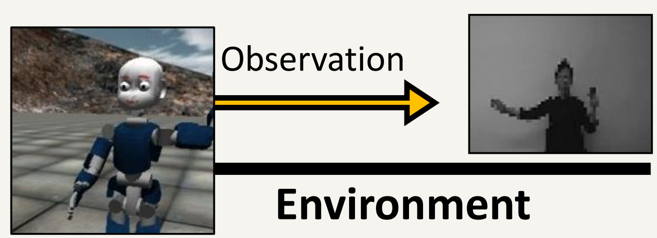


# Prediction Error Minimization

At each time step

1. Generate Visuo-Prop. Predictions
   - From Intention State (Top-Down Process)
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & **Update Intention State**
   - Bottom-Up Process
4. Iterates a Few Times

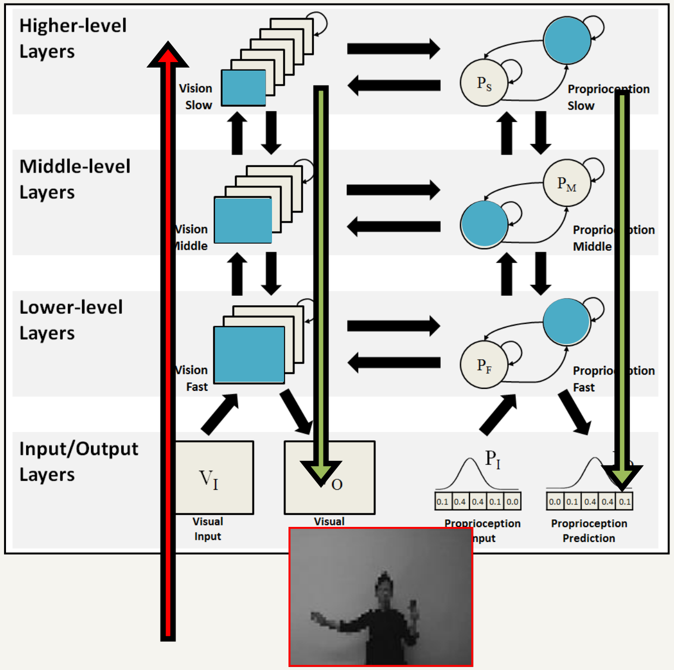


# Prediction Error Minimization

At each time step

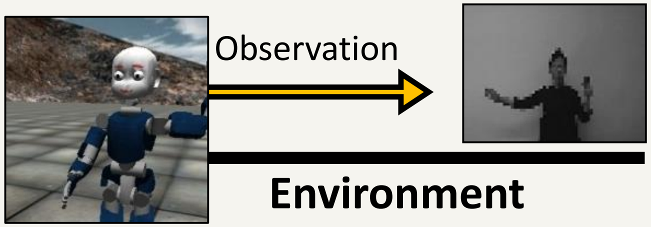
1. Generate Visuo-Prop. Predictions
   - From Intention State (Top-Down Process)
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. **Iterates a Few Times**



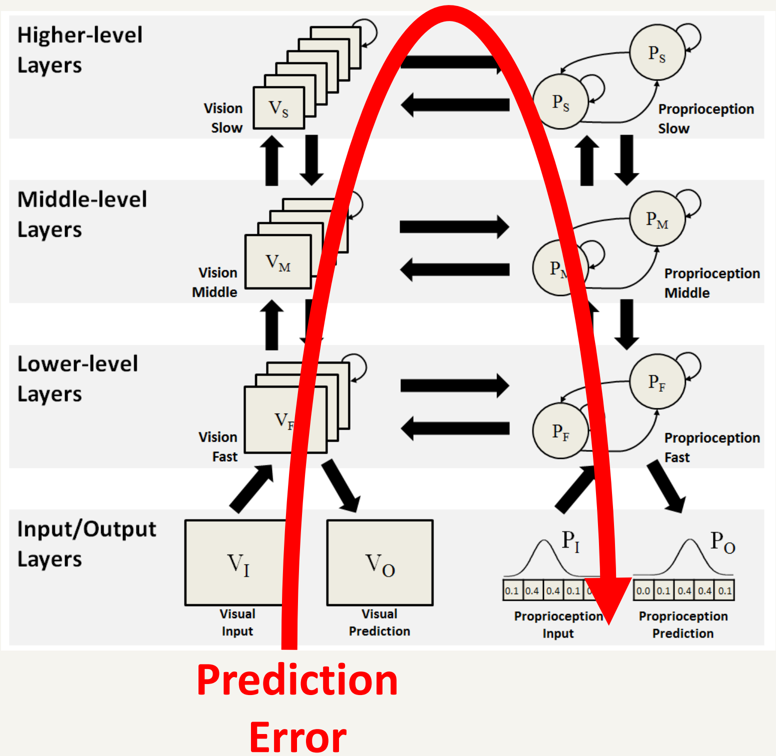

**MINIMIZED PREDICTION ERROR**

**“Perception as an Active Process”**

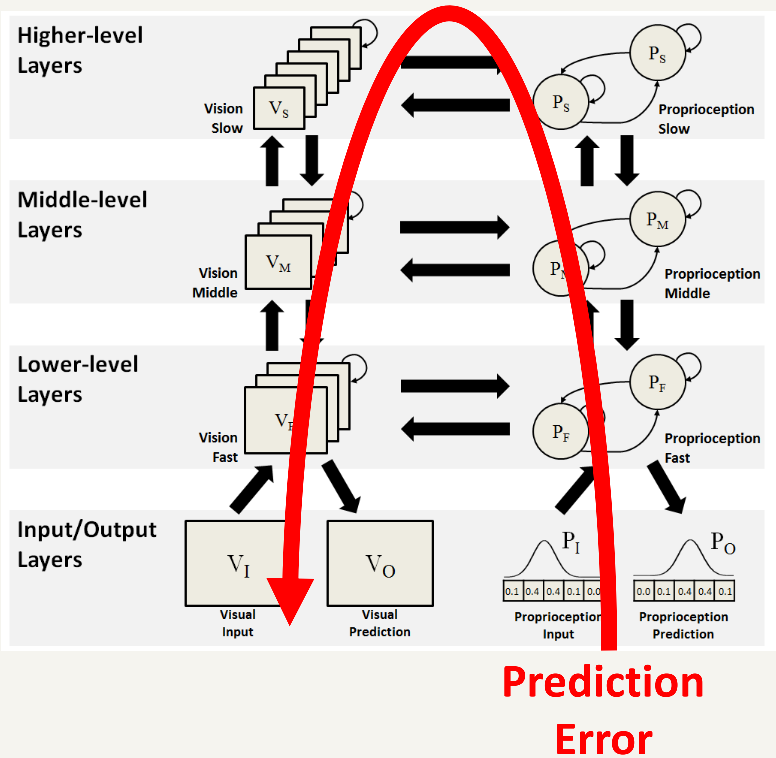
⇔ Solely determined by input



# Prediction Error Minimization

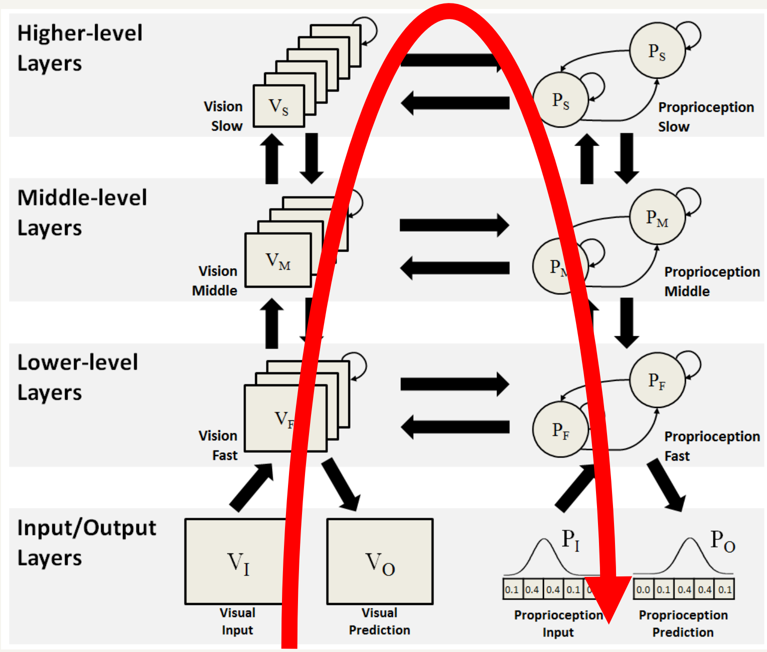



**Minimizing Visual PE**




**Minimizing Prop. PE**

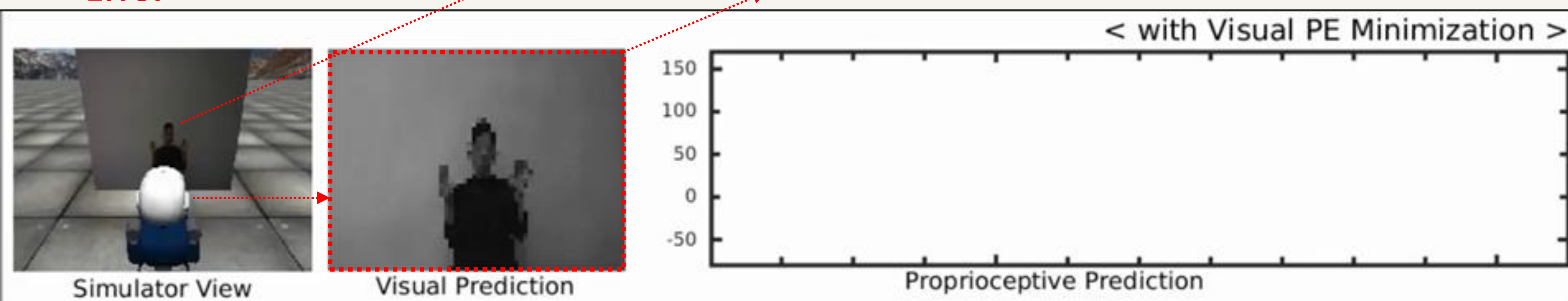
# Prediction Error Minimization



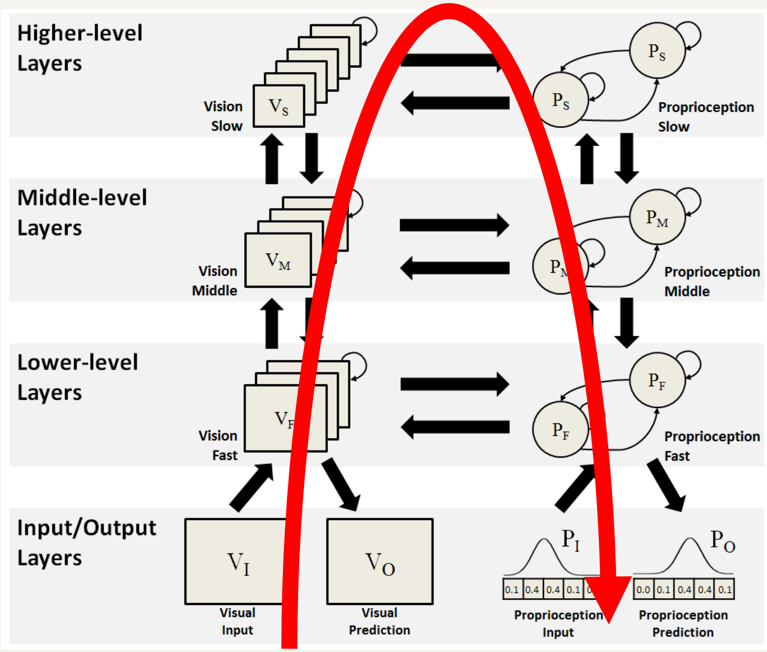

## Minimizing Visual PE

- Minimizing the difference b/w
  - Visual Prediction (predicted gesture)
  - Observation (observed gesture)

**Prediction Error** = Difference ( Observed Gesture, Predicted Gesture )



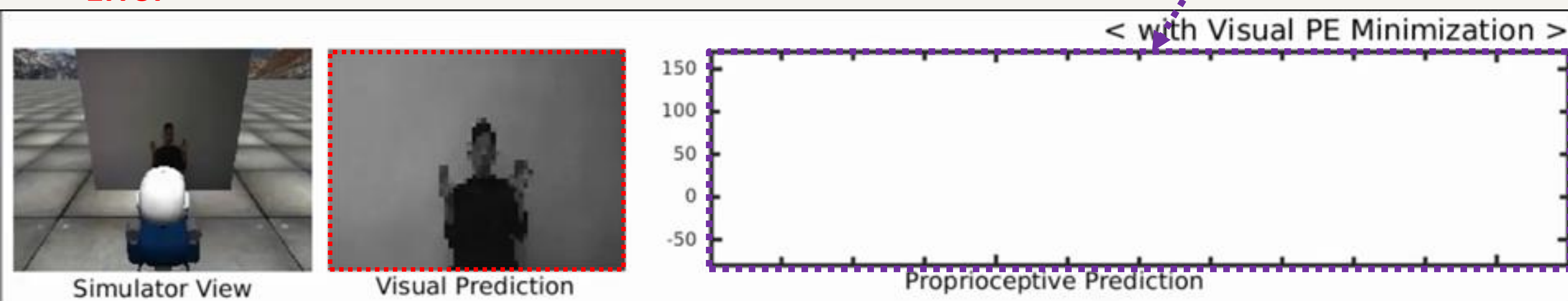
# Prediction Error Minimization



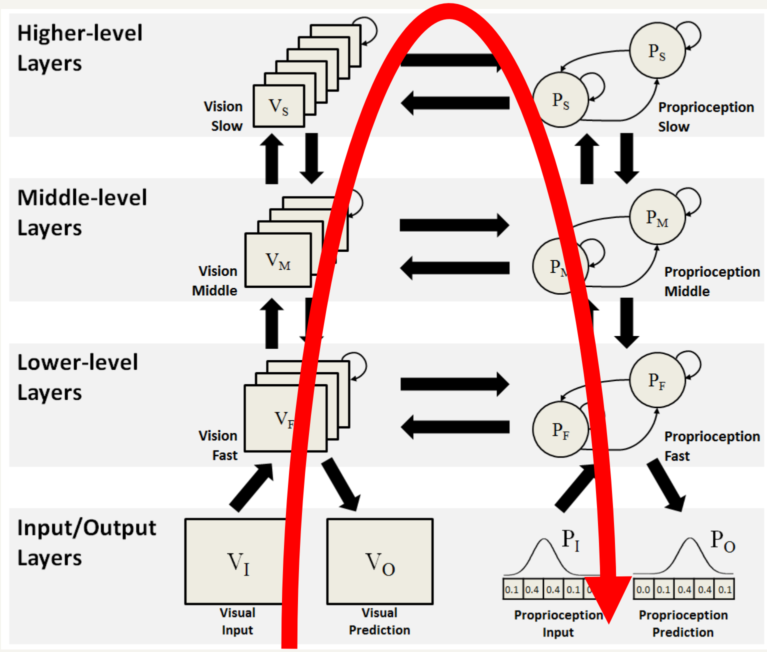

**Prediction Error**

## Minimizing Visual PE

- Minimizing the difference b/w
  - Visual Prediction (predicted gesture)
  - Observation (observed gesture)
- No External Proprioceptive Signal
  - Robot's action was generated simultaneously while minimizing Visual PE



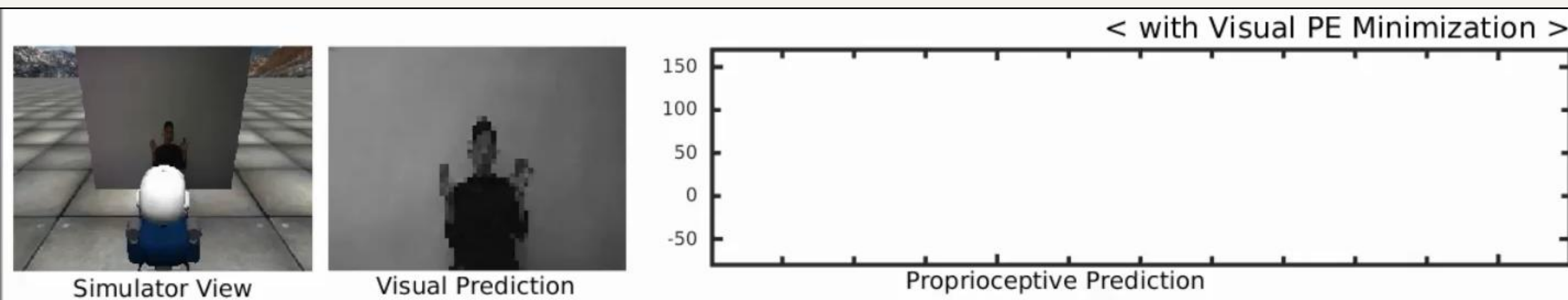
# Prediction Error Minimization




**Prediction Error**

## Minimizing Visual PE

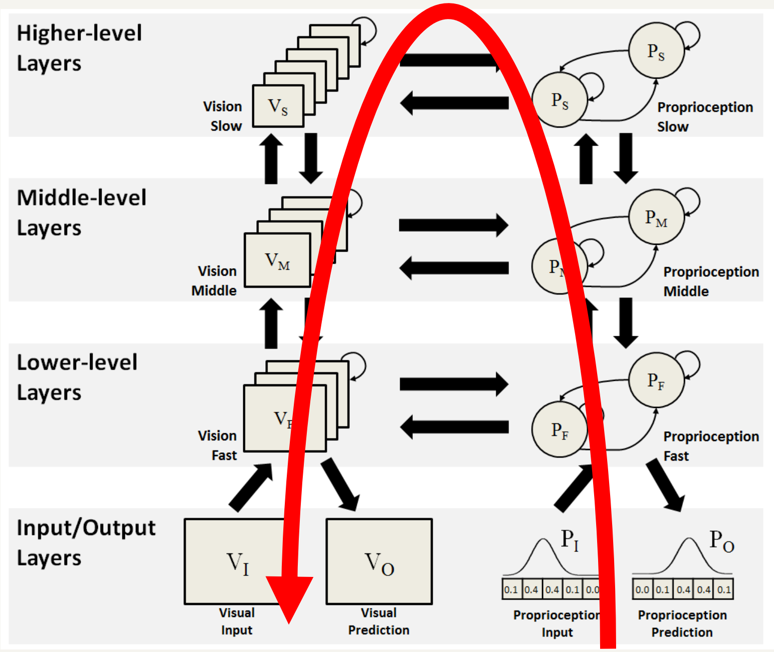
- **With Visual PE Minimization**
  - Predicted coherent visual & Proprioceptive patterns ➔ Successful imitation
- Without Visual PE Minimization
  - Did NOT predict Visual & Proprioceptive patterns ➔ Unsuccessful imitation



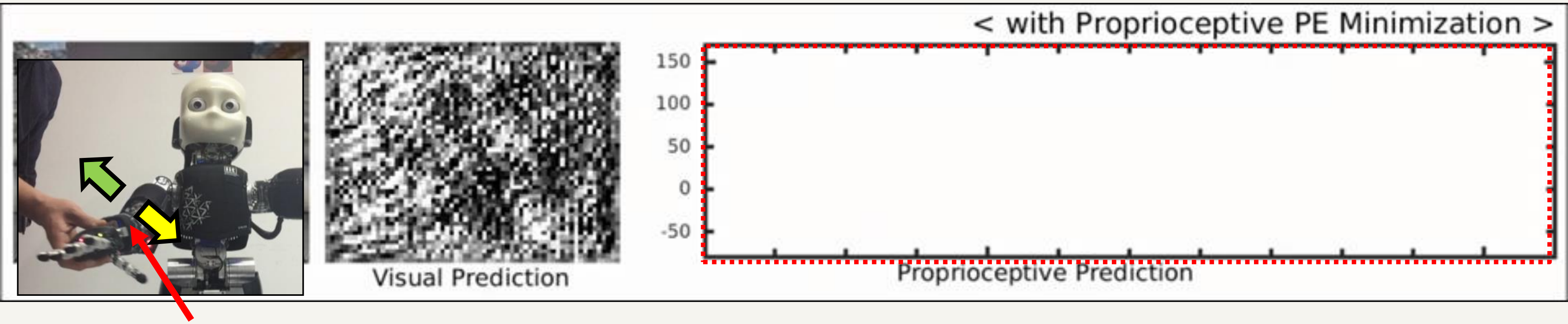
# Prediction Error Minimization

## Minimizing Prop. PE

- Minimizing the difference b/w
  - Prop. Prediction (Predicted joint position)
  - Observation (Perceived joint position)



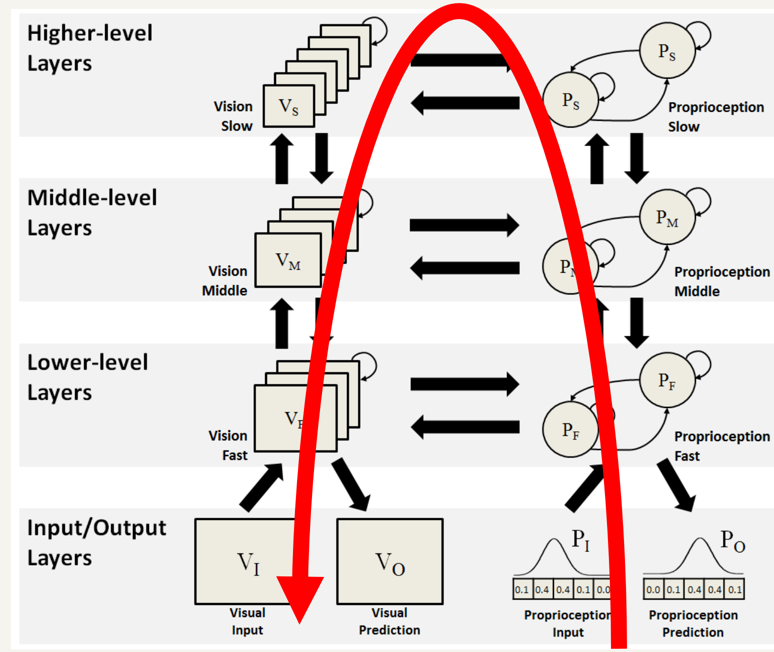

Difference ( Observed Jnt Position, Predicted Jnt Postion ) = **Prediction Error**



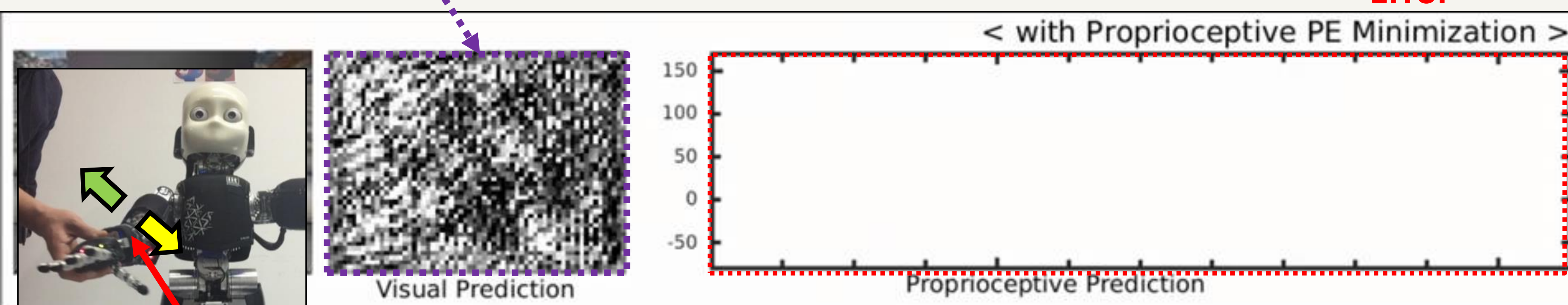
# Prediction Error Minimization

## Minimizing Prop. PE

- Minimizing the difference b/w
  - Prop. Prediction (Predicted joint position)
  - Observation (Perceived joint position)
- No External Visual Target Signal
  - Visual Prediction was generated simultaneously while minimizing Prop. PE




Prediction Error

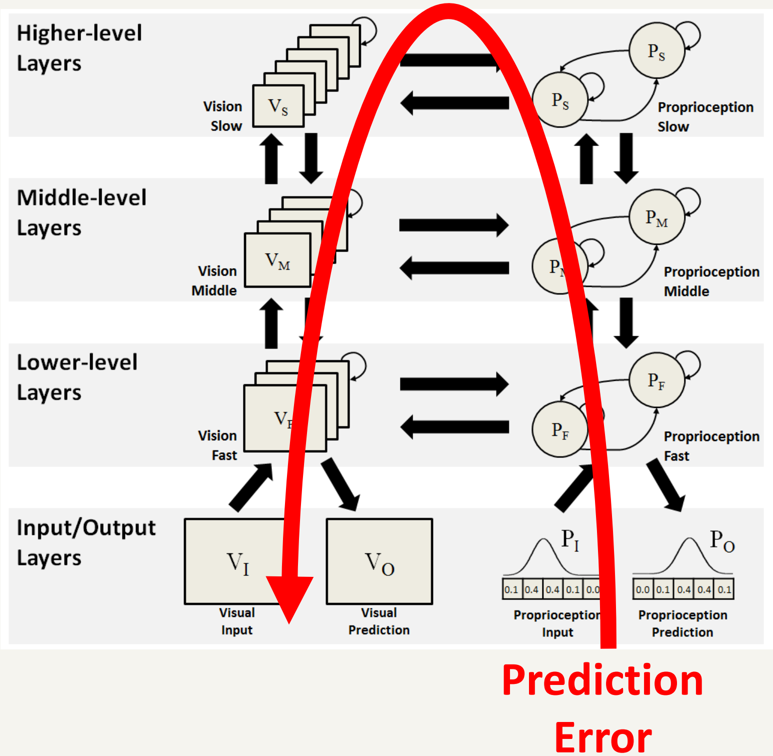


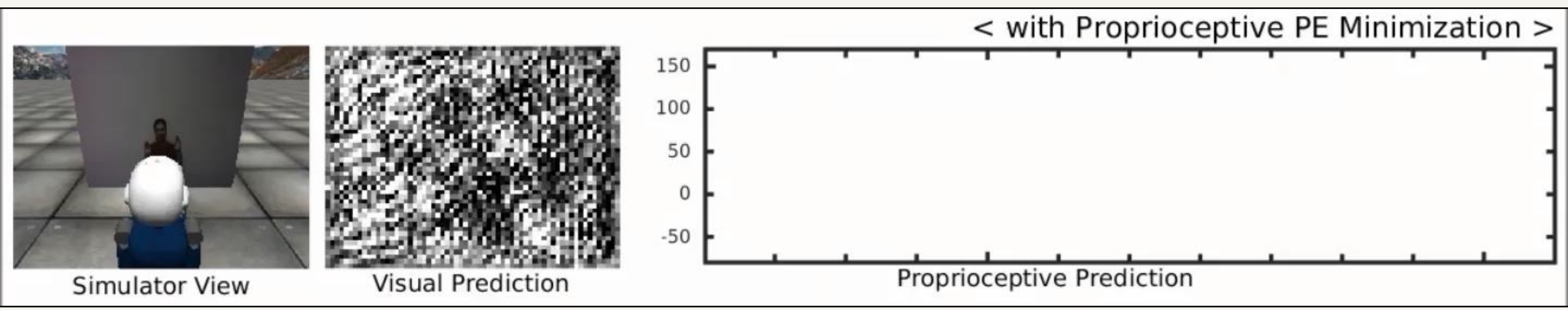

PROP. PREDICTION ERROR

# Prediction Error Minimization

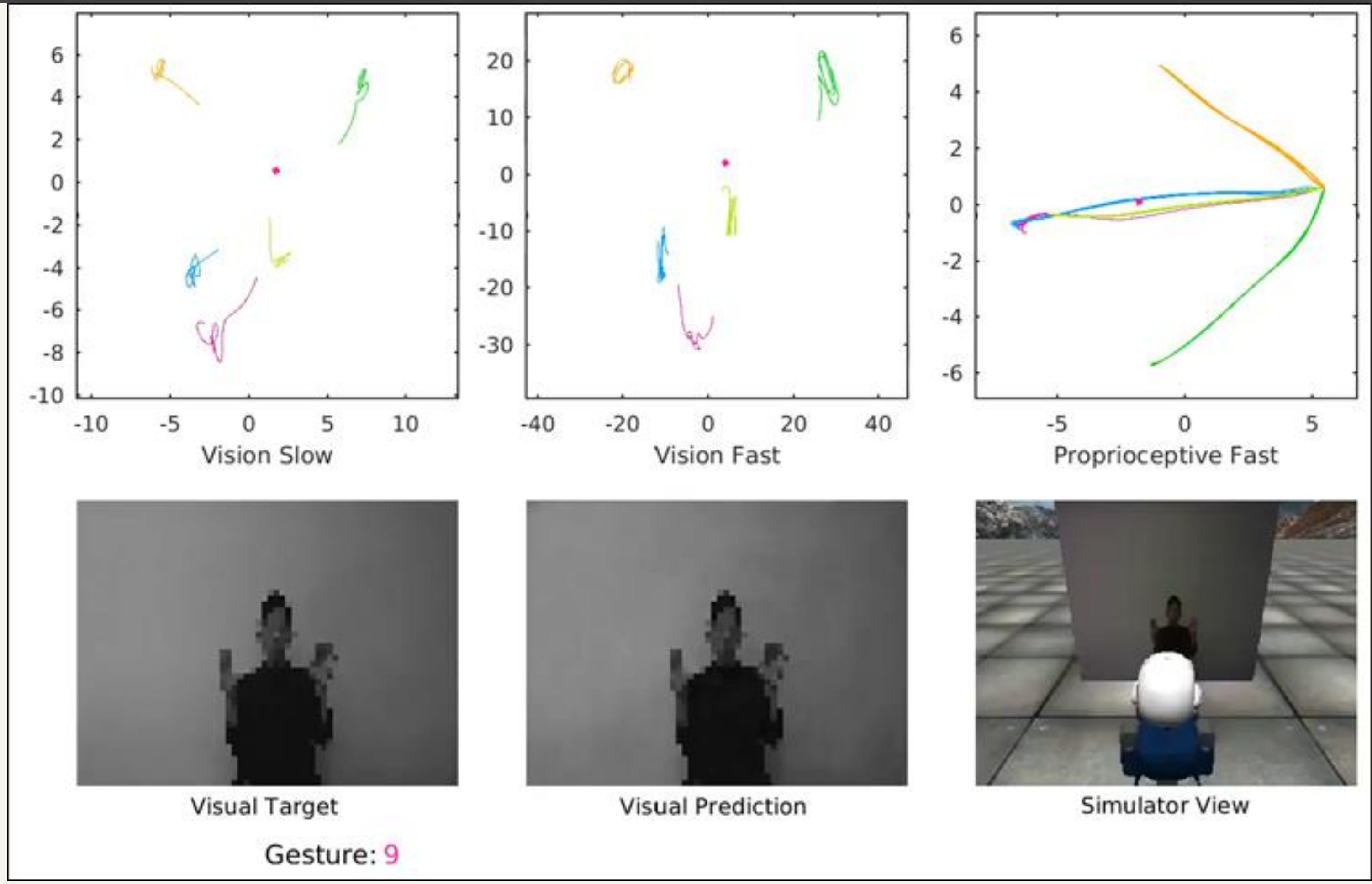
## Minimizing Prop. PE

- **With Proprioceptive PE Minimization**
  - Successfully minimized Proprioceptive PE
  - **Generated corresponding Visual Prediction (imaginary)**
- Without Proprioceptive PE Minimization
  - Not able to adapt to incoming Proprioceptive signal



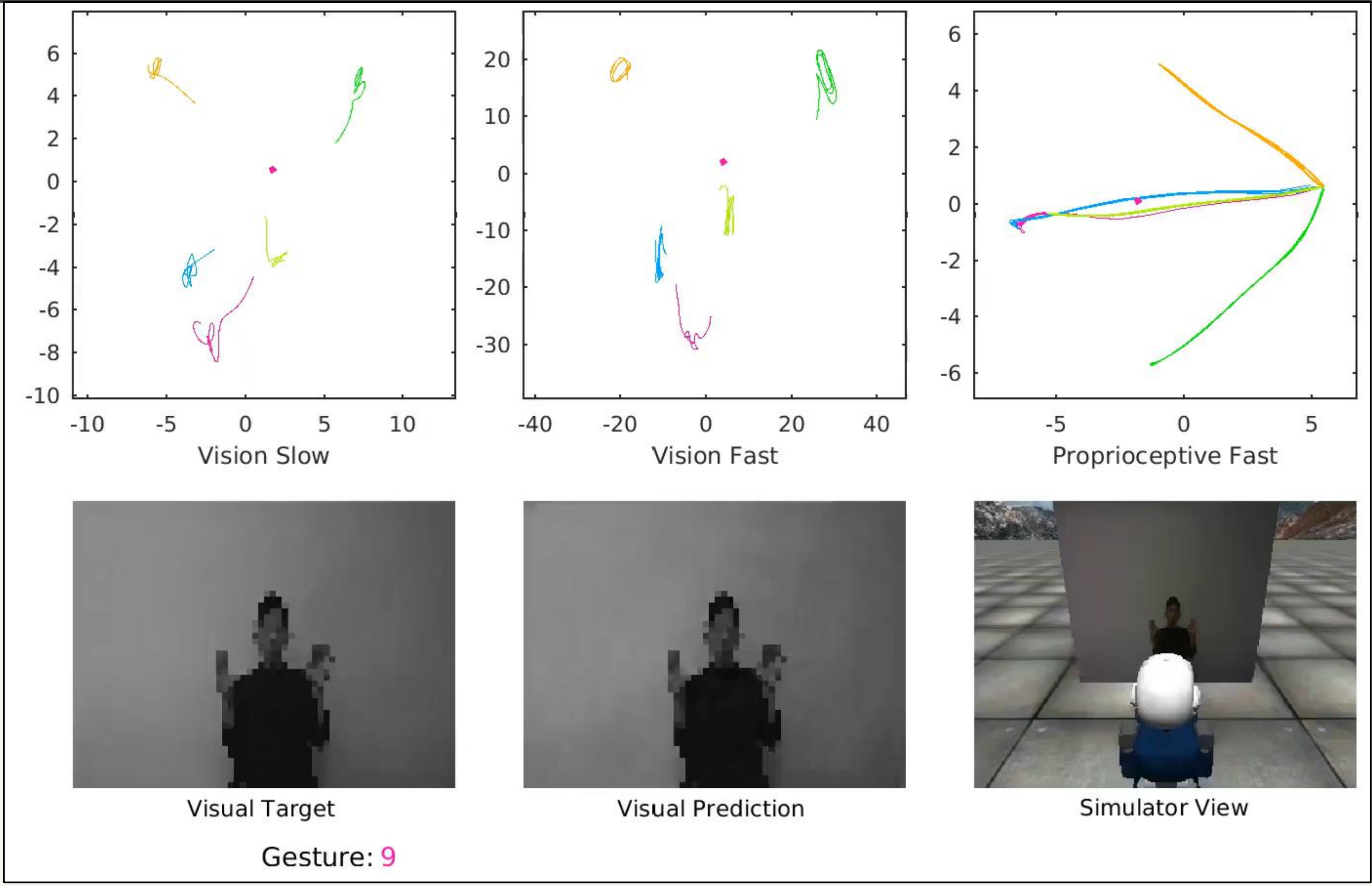
# Neural Activation while Minimizing Visual Prediction Error



# Neural Activation while Minimizing Visual Prediction Error

**Overlapping Trajectories**

b/w Training & Testing




- Inferring intention latent in observed patterns @ Higher-level
- Recalling the corresponding representations @ Lower-level

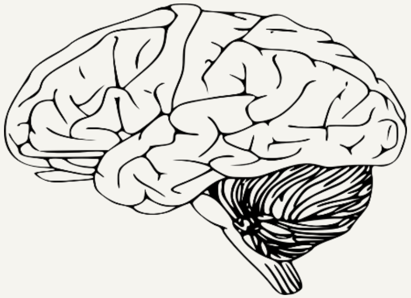
➔ Retrieval of missing sensorimotor signals

# Neural Activation while Minimizing Visual Prediction Error

- **MNS-like Behavior** emerged from
  1. Neural connectivity (between two pathways)
  2. Learning sensorimotor experience
  3. **Prediction Error Minimization**
     - "Predictive Coding Account of MNS" (Kilner, Friston and Frith, 2007)
     - "Within predictive coding, <u>recognition of causes is simply the process of jointly minimizing prediction error</u> at all levels of a cortical hierarchy."

# Conclusion

Build a **Cognitive Agent** based on



**Embodiment**

"Learning from sensorimotor experience"

acquired from dynamic interaction with the world

**Prediction**

"Brain = a Prediction Machine"

- Complex cognitive behaviors emerged
  - Mental simulation, Intention recognition, MNS-like behavior, etc.
  - From "Visuo-Motor associative learning under the predictive coding framework"

# Thank you

Please see the following paper for more information.

J. Hwang, J. Kim, A. Ahmadi, M. Choi and J. Tani, "Dealing With Large-Scale Spatio-Temporal Patterns in Imitative Interaction Between a Robot and a Human by Using the Predictive Coding Framework," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. doi: 10.1109/TSMC.2018.2791984, 2018

## CNRU @ OIST

Photo by Takazumi