Robot Reinforcement Learning from Crowdsourced Rewards

Christian I. Penaloza[1], Sonia Chernova[2], Yasushi Mae[1] Tatsuo Arai[1]

Graduate School of Engineering Science [1]
Osaka University
Japan

Department of Computer Science [2]
Worcester Polytechnic Institute
USA

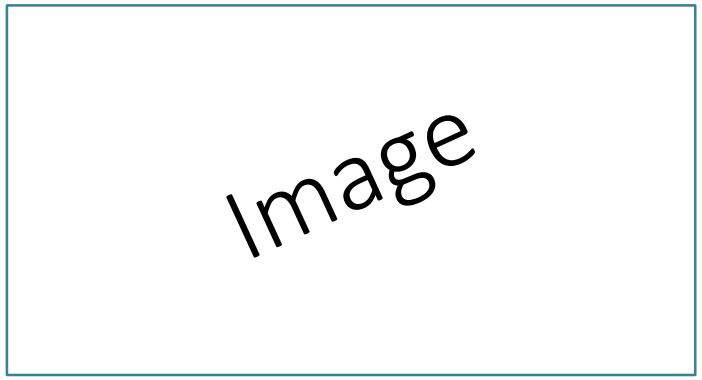
Outline

- Motivation
- Crowdsourcing
- System Architecture
- Experiment Setup
  - *Experiment 1: Rewards from Single person*
  - *Experiment 2: Rewards from a crowd (sequential ordered)*
  - *Experiment 3: Rewards from a crowd (unordered)*
- Conclusion

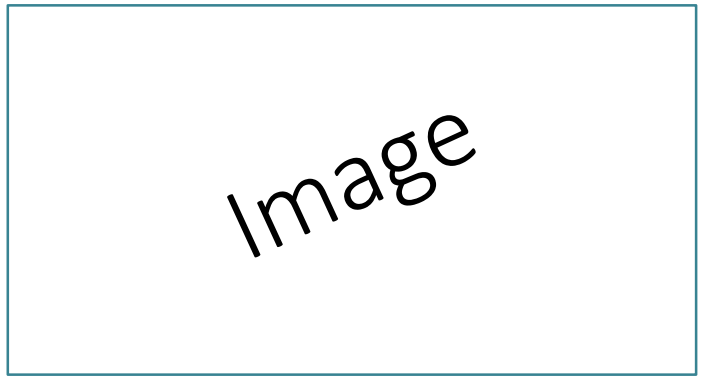
Motivation

Robot Learning from Demonstration

The main idea behind robot learning from demonstration is to derive a policy that reproduces a demonstrated behavior from a set of example demonstrations performed by a human teacher.



Problem

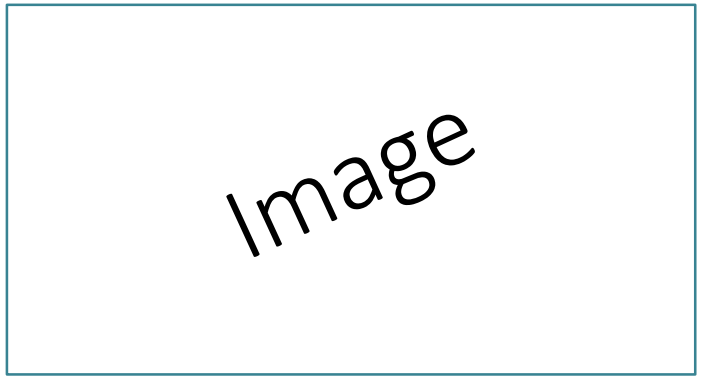


Teaching a robot to perform all sorts of different tasks is a tremendous duty!

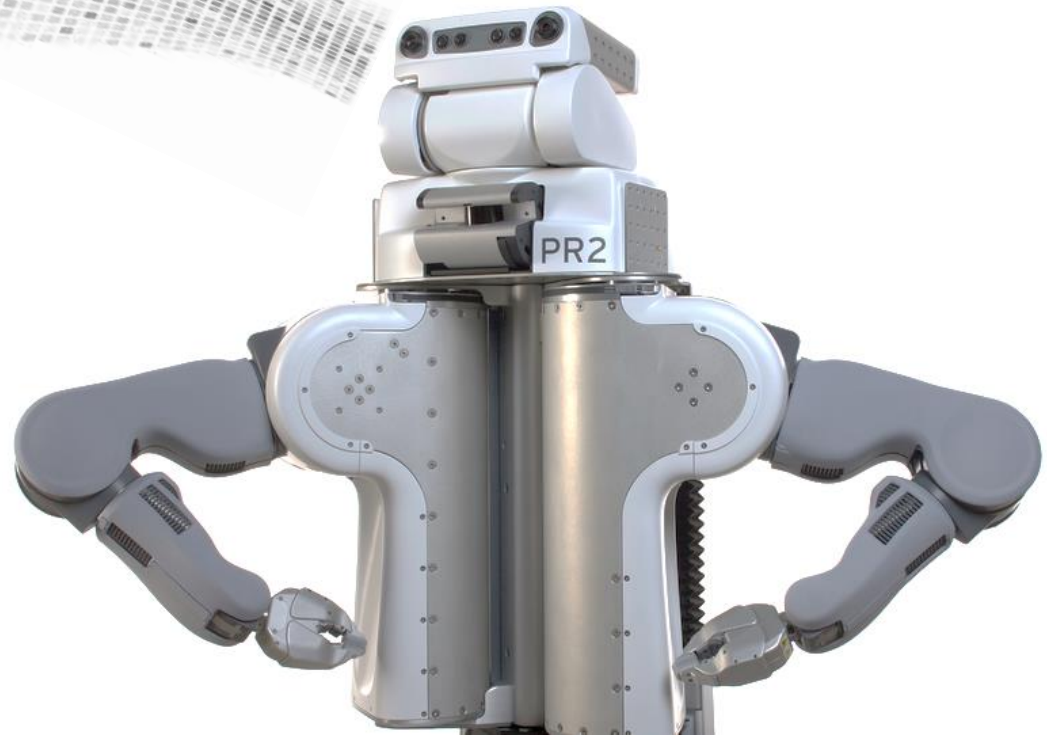
Ideally, a human could teach the robot basic skills, and the robot should **learn independently** to reuse those skills for different tasks.

Reinforcement Learning

Reinforcement Learning commonly **intractable** for real world robot applications because of large exploration times.

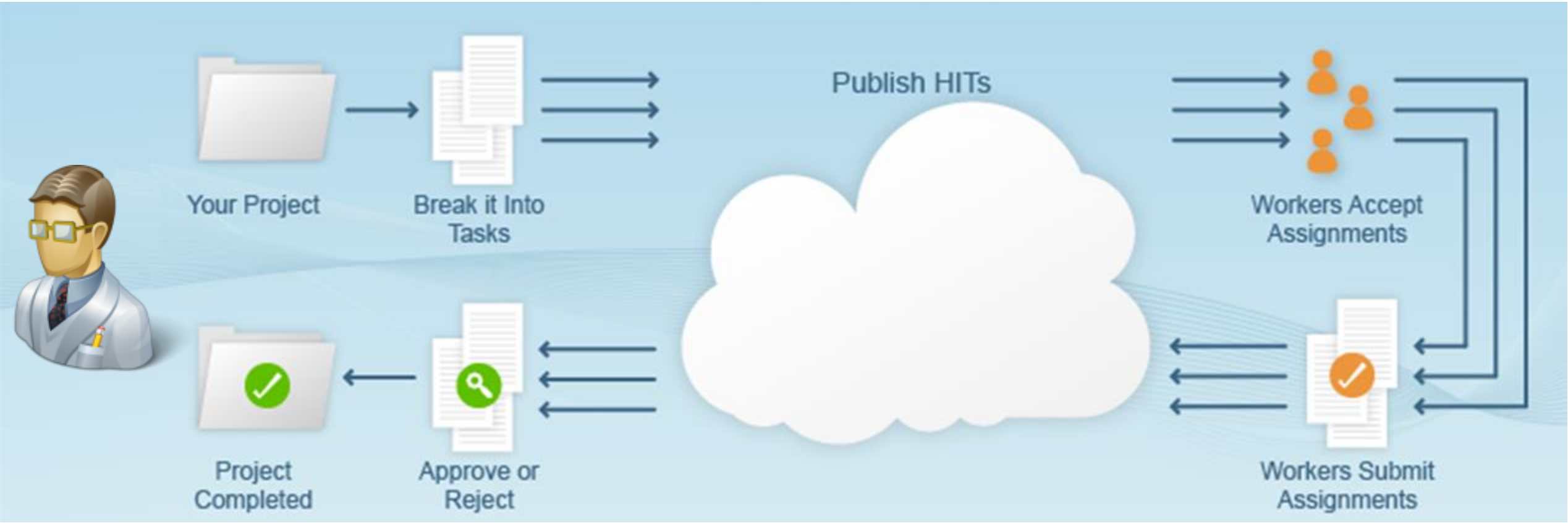


Interactive Reinforcement Learning (IRL) dedicated **human teacher** spends time guiding and **rewarding the robot** in monotonous tasks.

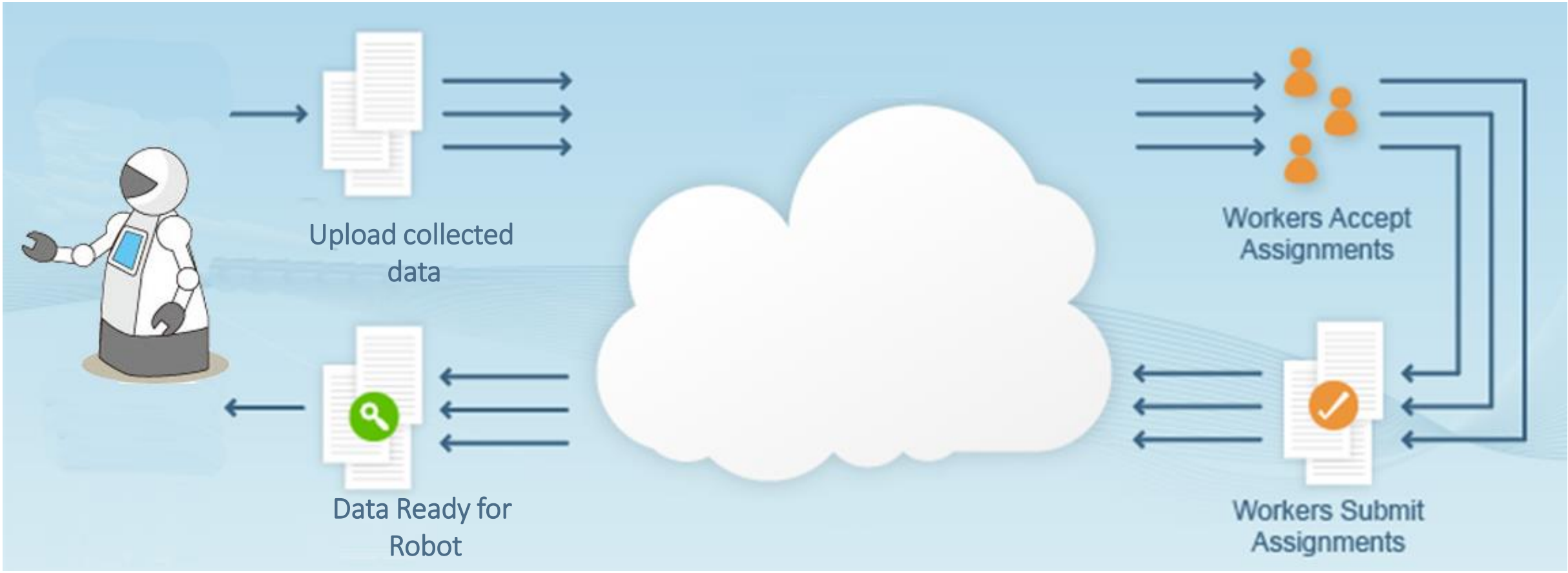


How can Robots can take advantage of the cloud?

Crowdsourcing



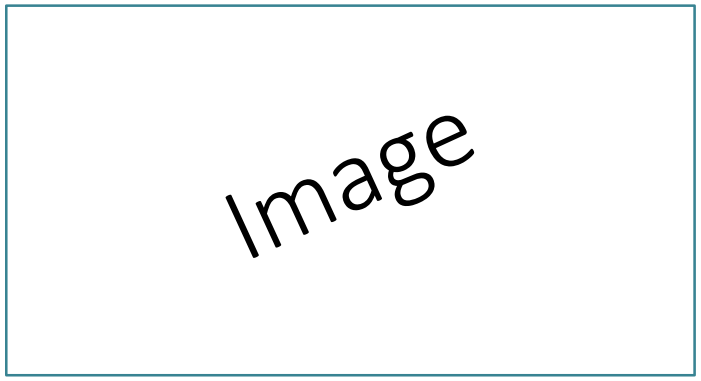
Crowdsourcing for Robots



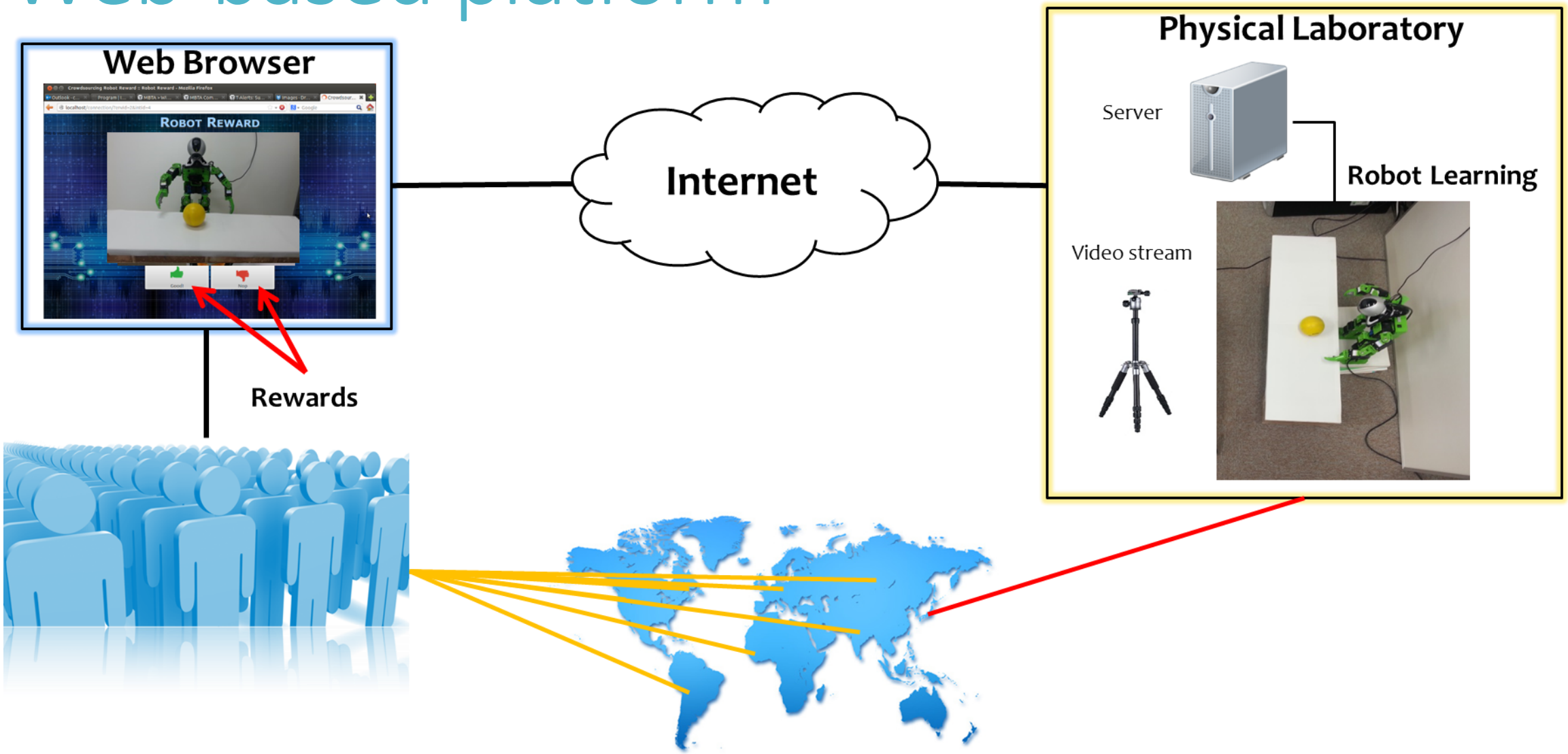
User Interface

Interfaces can be difficult to use for non-robot experts.

How can we make it more easy for users?

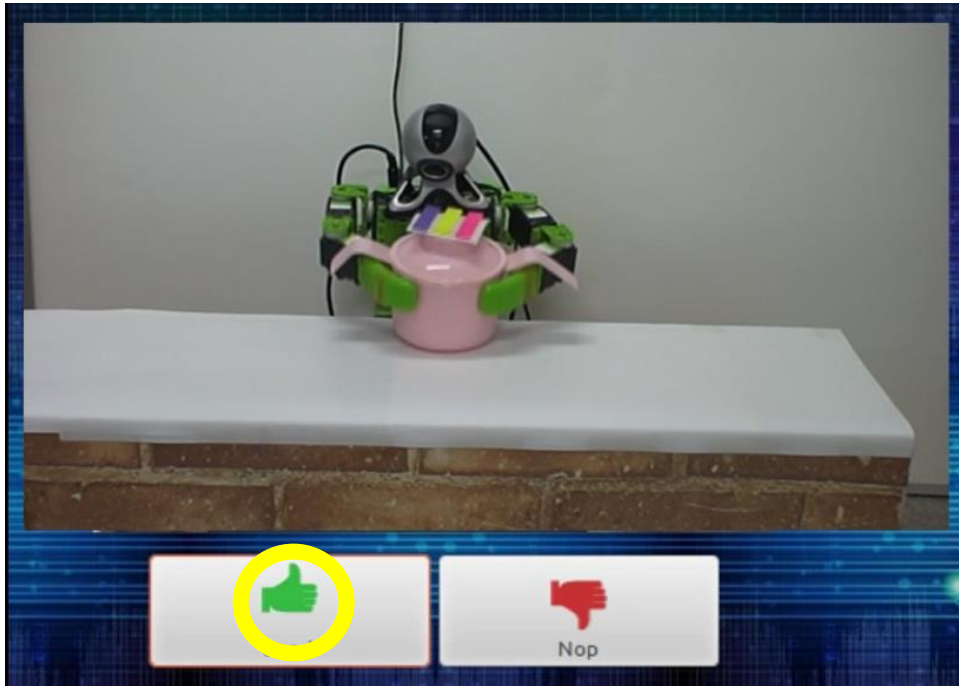


Web-based platform



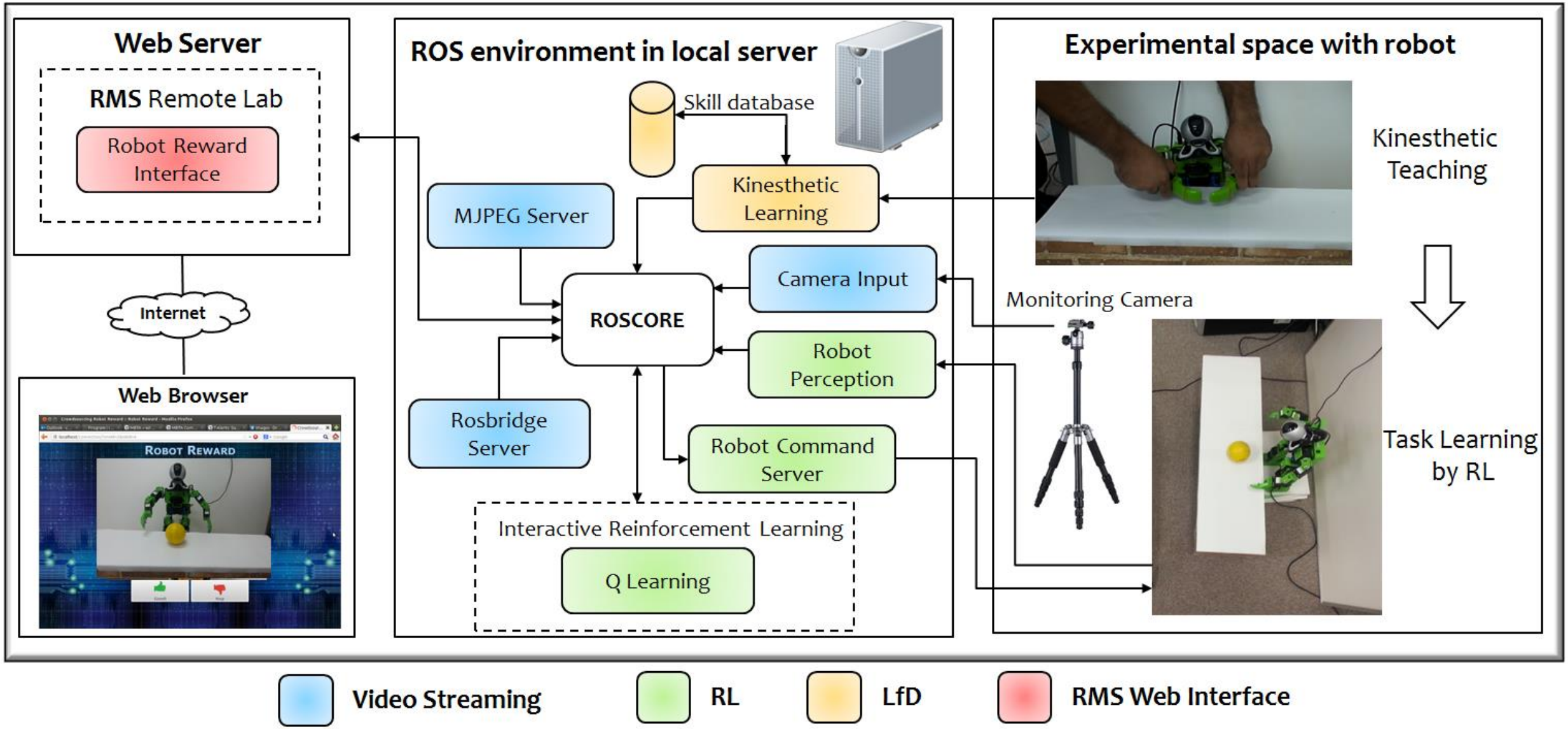
Our Approach

Learning from Demonstration
by Kinesthetic Teaching



Reinforcement Learning
using Crowdsourced Rewards

System Architecture



(RMS) Robot Management System

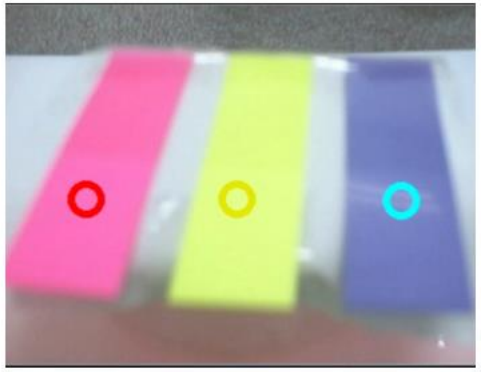
RMS (Robot Management System) is a remote lab management tool designed to control ROS enabled robots from the web.

RMS is written in a robot-platform independent manner allowing for the control of a variety of robots.

RMS allows for basic user management, interface management, and content management.

Experimental Setup

- Human teacher demonstrated basic hand motion skills to the robot.
- Objects with one or more color markers were presented in front of the robot one at a time.
- Each color marker represented a state in the RL algorithm
- Initially, the robot did not know which action corresponds to which state, and the goal of the experiment was to discover the state-action relationship.



OBJECT - STATES

Object	States
Ball	s_1(blue), s_2(green)
Toy	s_1(blue), s_2(green), s_5(purple)
Left_handle_cup	s_3(yellow)
Right_handle_cup	s_4(pink)
Two_handle_cup	s_3(yellow), s_4(pink), s_5(purple)

STATE - ACTION

States	Action
s_1: Blue	*Push_w_LeftArm*
s_2: Green	*Push_w_RightArm*
s_3: Yellow	*Lift_w_LeftArm*
s_4: Pink	*Lift_w_RightArm*
s_5: Purple	*Lift_w_BothArms*

Experiments

Experiment duration: 60 minutes.

Participants could click several times to either positive or negative reward buttons, but only last click counted.

Experiment 1

One person

Participant was instructed that he could leave the experiment whenever he felt tired or bored.

Experiment 2

Sequential order

① 12 persons participated.
② Each participant was assigned a session of 5 minutes.
③ Sessions were organized in sequential order.

Experiments

Experiment 3

Unordered

① 12 persons participated.

② Each participant was asked to provide rewards from 5 to 10 minutes.

③ Each participant was free to choose when to log in within the assigned hour.

④ Positive or Negative rewards were assigned using Majority Voting

Ex.1 Results

(One person)

Participant lasted for 22.5 minutes.

Learning performance stayed constant after participant stopped providing rewards.

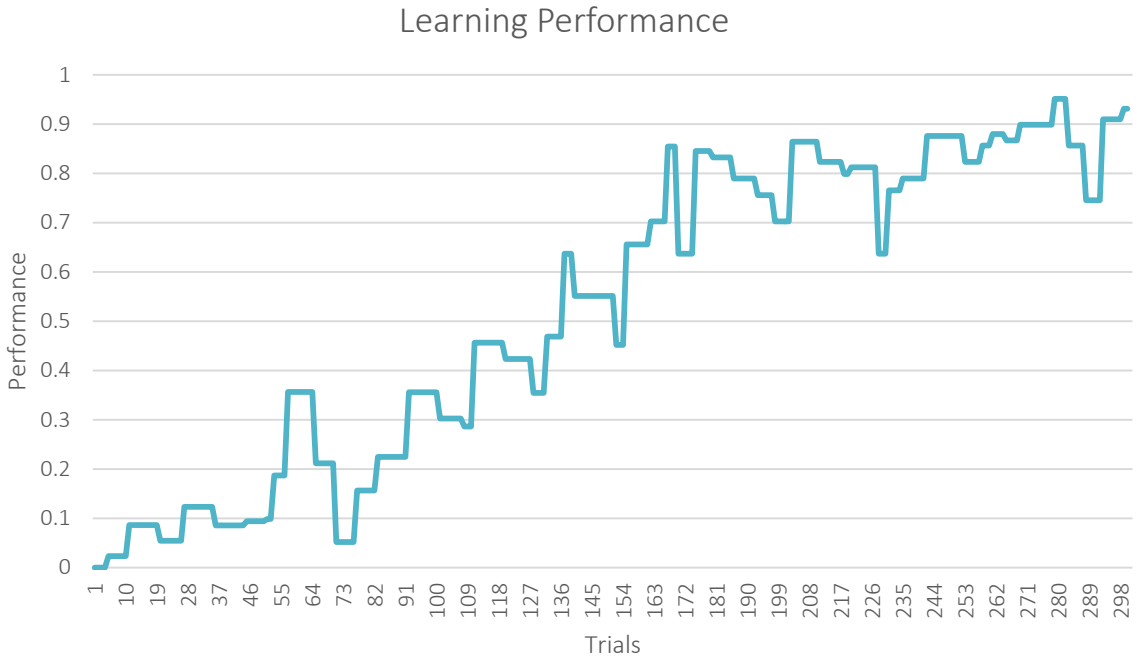
Achieved learning performance: 0.41863 (42%)

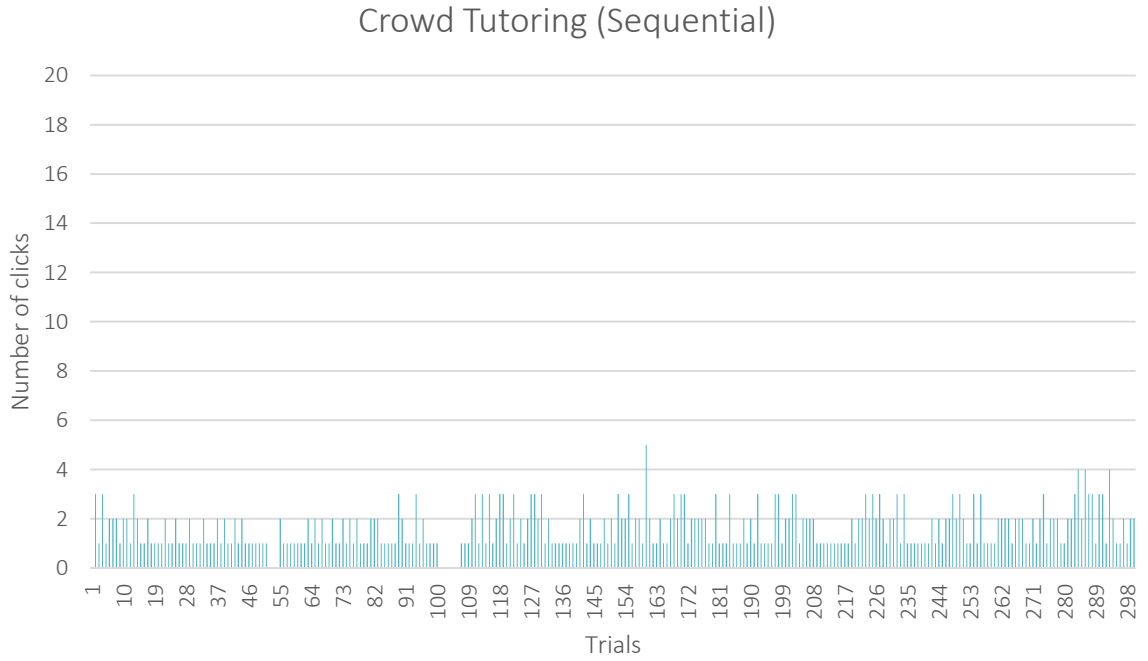
Ex.2 Results

(Sequential order)

Robot was able to achieve good learning performance after 60 minutes of continuous training.

Some participants clicked 1-3 times per reward, while other clicked 3-5 times.





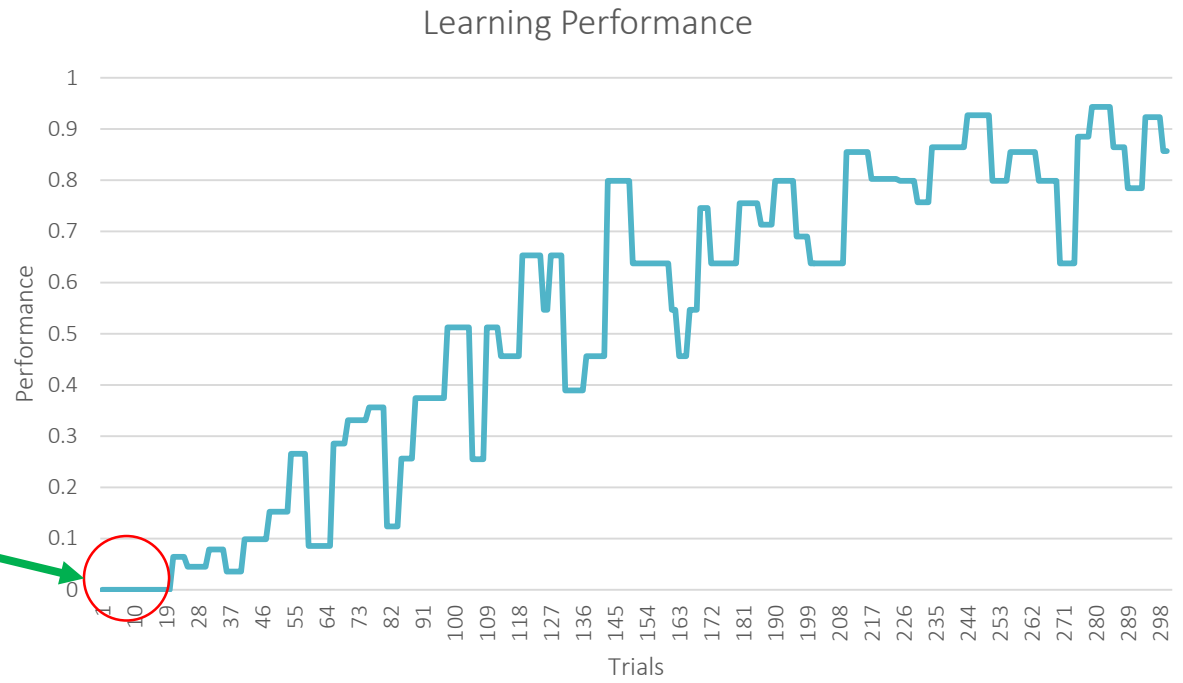
Ex.3 Results

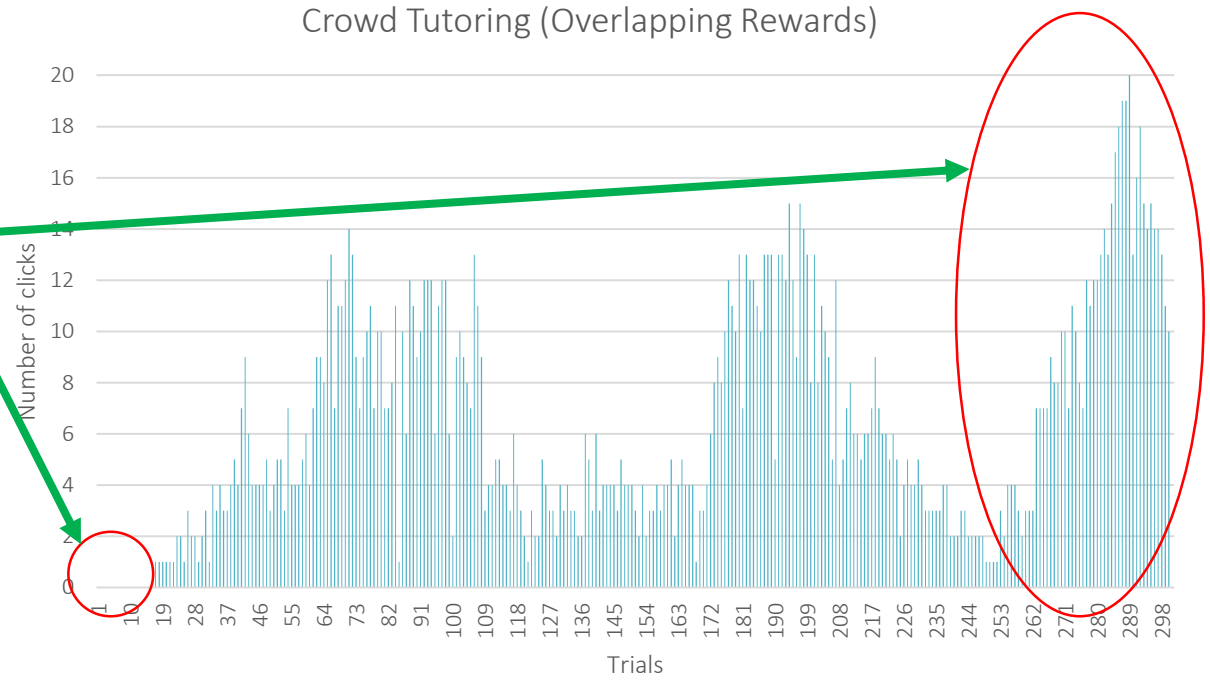
(Unordered)

Nobody logged in for the first 4 minutes.

Number of clicks increased and decreased as some participants logged in and out.

In the last 10 minutes, the number of clicks increased considerably, probably due to the fact that most participants waited until the last 10 minutes to log in.





Conclusion

- We described our approach for achieving robot learning by combining basic skill transfer through physical demonstrations, reinforcement learning and web-based crowdsourced rewards obtained via remote laboratory.

- Experimental results demonstrate that by transferring the training to the crowd, it is possible to achieve a successful robot learning performance while minimizing the effort of a dedicated teacher.

- It is necessary to consider different mechanisms to take advantage of crowdsourced rewards for more complex and time-consuming learning tasks.

Thank you

Christian I. Penaloza

penaloza@aria-laby.sys.es.osaka-u.ac.jp

Graduate School of Engineering Science
Osaka University
Japan