# Adaptive Robot Assisted Therapy using Interactive Reinforcement Learning

Konstantinos Tsiakas[1,2], Maria Dagioglou[2], Vangelis Karkaletsis[2], Fillia Makedon[1]

[1]HERACLEIA - Human Centered Computing Lab
Department of Computer Science
University of Texas at Arlington, USA

[2]Software and Knowledge Engineering Lab
Institute of Informatics and Telecommunications
National Center for Scientific Research, Demokritos, Greece

ICSR 2016

# Outline

Robot Assisted Therapy has been extensively applied to assist users with cognitive and/or physical impairments

- **Robotic Assistive Systems** can establish a productive interaction with the user, improving the effects of a therapy session
- Two significant attributes of such systems (or agents) are **personalization** and **safety**
- **Reinforcement Learning** is an appropriate framework for modeling and optimizing the behavior of an agent that interacts with a user

# Related Work

- A robotic seal that interacted with patients to explore the role of the robot in the improvement of conversation and interaction skils (Kidd et al. 2006)
- A Socially Assistive Robot (SAR) motivator to increase user enjoyment and performance on physical/cognitive task (Fasola & Mataric 2010)
- A social robot for personalized Robot-Child Tutoring (Ramachandran & Scassellati 2014)

These works validate that a **personalized** and **adaptive** robotic system can establish a tailored and productive interaction with the user

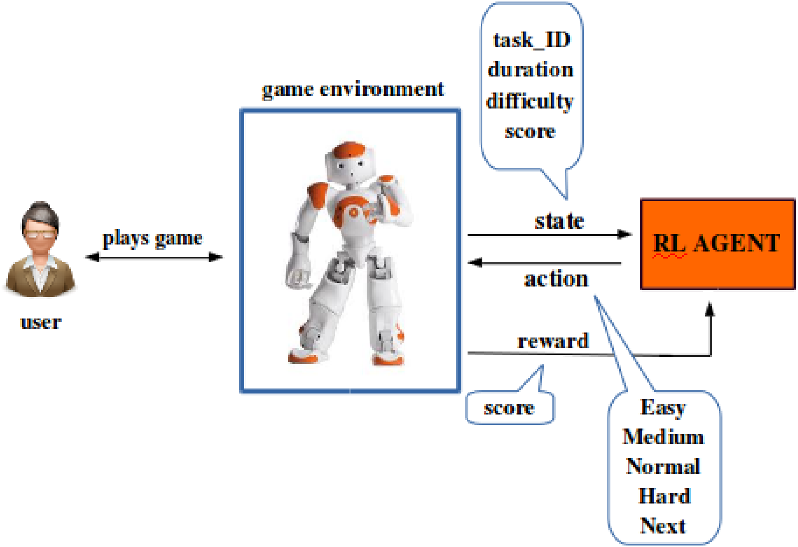
## Adaptive Robot Assistive Therapy

### Adaptive Training Task

- *Adaptive Training Task example:* A training task designed to improve participant's specific cognitive or physical aspects
- The user needs to complete a set of three tasks
- The robot models the difficulty level based on the user's performance during the interaction, to assist the user to finish the session, while keeping them engaged

# Adaptive Cognitive Task

## Robot Behavior Modelling as an MDP



# Learning Experiments

- The agent needs to learn the **optimal policy**; the mapping from states to actions that maximizes the accumulated reward (or total return) during each interaction
- In order to evaluate this modeling, we apply $\epsilon$-**greedy Q-learning** for four different user models
- These user models depict different user skills under different game parameters (task difficulty and duration)

# Learning Experiments

## User Models

| User 1 (expert) | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 |
| | 2 | 1 | 1 | 1 | 1 |
| | 3 | 1 | 1 | 1 | 1 |
| | 4 | 1 | 1 | 1 | 1 |
| | 5 | 1 | 1 | 1 | 1 |
| | 6 | 0 | 0 | 0 | 0 |

| User 2 (novice) | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 |
| | 2 | 1 | 1 | 0 | 0 |
| | 3 | 1 | 0 | 0 | 0 |
| | 4 | 1 | 0 | 0 | 0 |
| | 5 | 1 | 0 | 0 | 0 |
| | 6 | 0 | 0 | 0 | 0 |

| User 3 | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 0 | 0 |
| | 1 | 1 | 1 | 0 | 0 |
| | 2 | 1 | 1 | 0 | 0 |
| | 3 | 1 | 0 | 0 | 0 |
| | 4 | 1 | 0 | 0 | 0 |
| | 5 | 1 | 0 | 0 | 0 |
| | 6 | 0 | 0 | 0 | 0 |

| User 4 | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 0 | 0 |
| | 1 | 1 | 1 | 0 | 0 |
| | 2 | 1 | 1 | 1 | 0 |
| | 3 | 1 | 1 | 1 | 1 |
| | 4 | 1 | 1 | 1 | 1 |
| | 5 | 0 | 0 | 0 | 0 |
| | 6 | 0 | 0 | 0 | 0 |

# Learning Experiments

## User Models

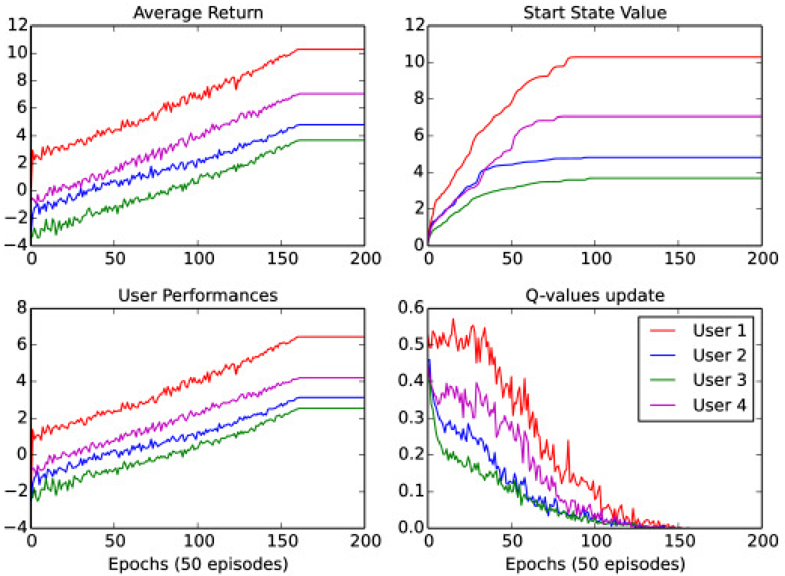
| User 1 (expert) | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 |
| | 2 | 1 | 1 | 1 | 1 |
| | 3 | 1 | 1 | 1 | 1 |
| | 4 | 1 | 1 | 1 | 1 |
| | 5 | 1 | 1 | 1 | 1 |
| | 6 | 0 | 0 | 0 | 0 |

| User 2 (novice) | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 1 | 1 |
| | 1 | 1 | 1 | 1 | 1 |
| | 2 | 1 | 1 | 0 | 0 |
| | 3 | 1 | 0 | 0 | 0 |
| | 4 | 1 | 0 | 0 | 0 |
| | 5 | 1 | 0 | 0 | 0 |
| | 6 | 0 | 0 | 0 | 0 |

| User 3 | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 0 | 0 |
| | 1 | 1 | 1 | 0 | 0 |
| | 2 | 1 | 1 | 0 | 0 |
| | 3 | 1 | 0 | 0 | 0 |
| | 4 | 1 | 0 | 0 | 0 |
| | 5 | 1 | 0 | 0 | 0 |
| | 6 | 0 | 0 | 0 | 0 |

| User 4 | | Difficulty level | | | |
|---|---|---|---|---|---|
| | | easy | medium | normal | hard |
| duration | 0 | 1 | 1 | 0 | 0 |
| | 1 | 1 | 1 | 0 | 0 |
| | 2 | 1 | 1 | 1 | 0 |
| | 3 | 1 | 1 | 1 | 1 |
| | 4 | 1 | 1 | 1 | 1 |
| | 5 | 0 | 0 | 0 | 0 |
| | 6 | 0 | 0 | 0 | 0 |

# Learning Experiments



# Policy Transfer Experiments

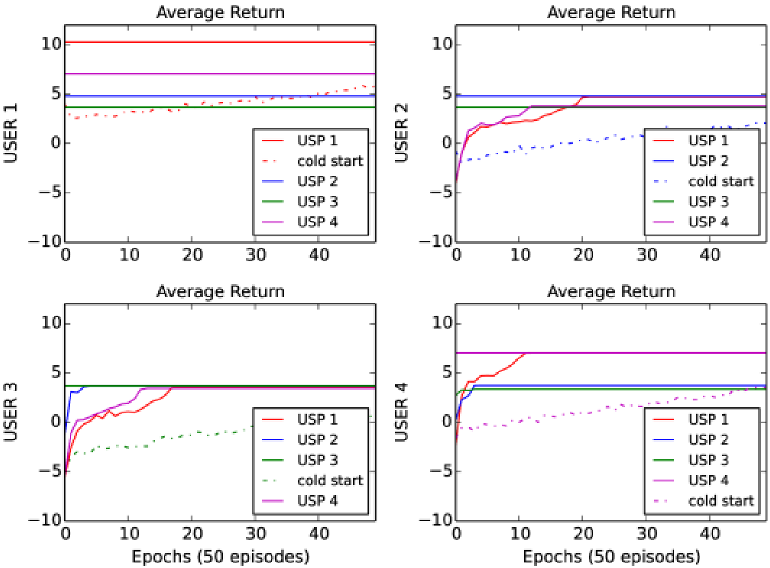
We make the following assumptions:

- A user specific policy (USP) is the optimal policy (for the corresponding model); the one that maximizes total return, thus user performance
- Applying a learned policy to a different user model may not be efficient but better than learning from scratch

We apply the four different USPs to the four different user models, following an exploitation-only approach (minimize risks and harmful actions)

# Policy Transfer Experiments

## Results

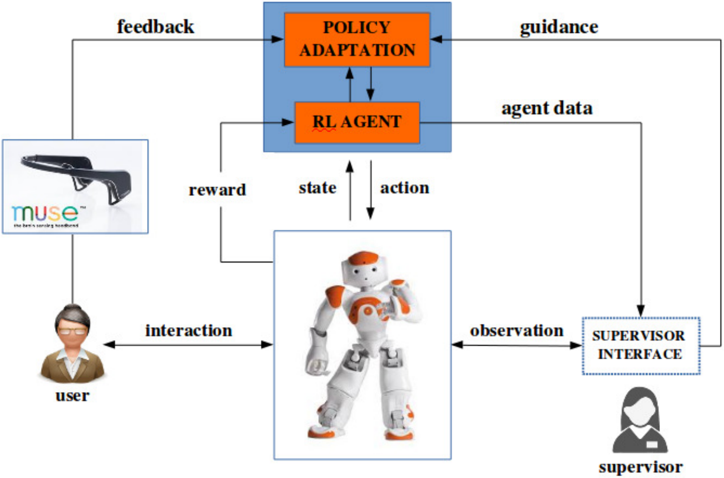


# Interactive Learning and Adaptation Framework

## Description

- A framework that utilizes **Interactive Reinforcement Learning** for dynamical adaptation
  - **Learning from Feedback** treats the human input as a reward after the executed action
  - **Learning from Guidance** allows human intervention to the selected action before execution, proposing (corrective) actions
- Integration of user feedback and expert guidance towards a *lifelong* learning setup

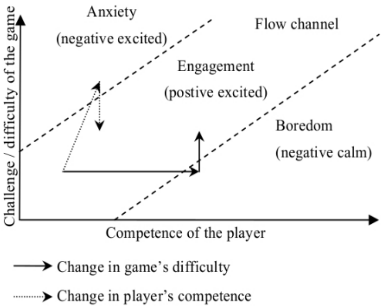
# Interactive Learning and Adaptation Framework



# Interactive Learning and Adaptation Framework

## Learning from Feedback

*Strong involvement in a task occurs when the skills of an individual meets the challenge of a task* (Chanel et al. 2008)




$$feedback = |diff - diff'|/N_{diff} \in [-1, 0]$$
$$Q(s,a) = Q(s,a) + \beta \cdot feedback$$

# Interactive Learning and Adaptation Framework

## Learning from Guidance

Combination of two techniques:

- Human-guided exploration - SPARC (Senft et al. 2016)
- Teaching on a budget– mistake correcting (Torrey at al. 2013)

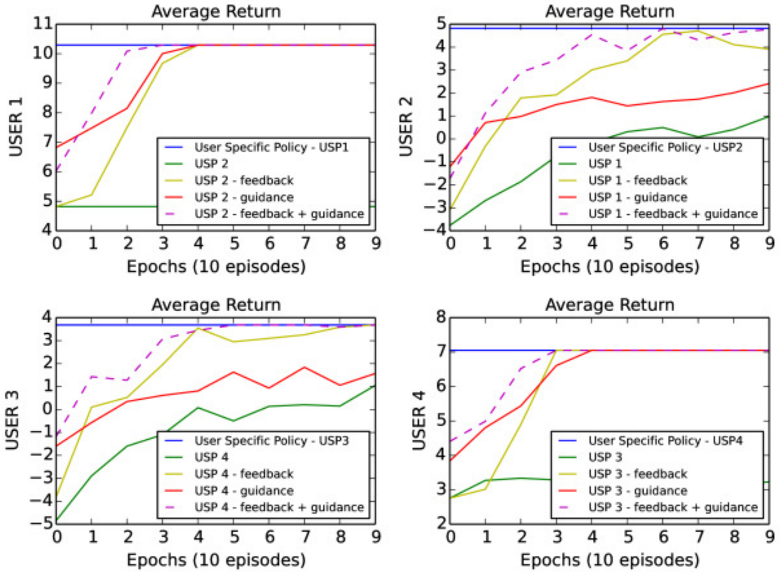
**Assumption:** Almost perfect supervisor – follows the corresponding USP with probability p

If $n < \mathrm{N}$ && $action_{\mathrm{MDP}} \neq action_{\mathrm{supervisor}}$:

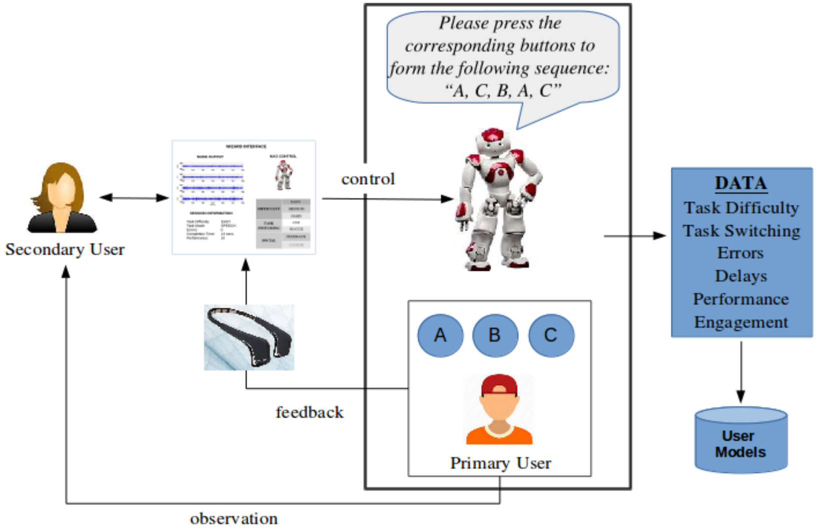
perform $\mathrm{action}_{\mathrm{supervisor}}$ with probability $p = 0.7$ (otherwise $\mathrm{action}_{\mathrm{MDP}}$)

$n$ += 1

# Preliminary Adaptation Experiments



# Ongoing Work



# Thank you...

Contact
konstantinos.tsiakas@mavs.uta.edu