In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)* (IROS 2021), Virtual, September 2021

Capturing Skill State in Curriculum Learning for Human Skill Acquisition*

Keya Ghonasgi[1], Reuth Mirsky[2], Sanmit Narvekar[2], Bharath Masetty[1],
Adrian M. Haith[3], Peter Stone[2,4], Ashish D. Deshpande[1]

***Abstract*— Humans learn complex motor skills with practice and training. Though the learning process is not fully understood, several theories from motor learning, neuroscience, education, and game design suggest that curriculum-based training may be the key to efficient skill acquisition. However, designing such a curriculum and understanding its effects on learning are challenging problems. In this paper, we define the Human-skill Curriculum Markov Decision Process (H-CMDP) to systematize the design of training protocols. We also identify a vocabulary of performance features to enable the approximation for a human's skill level across a variety of cognitive and motor tasks. A novel task domain is introduced as a testbed to evaluate the effectiveness of our approach. Human subject experiments show that (1) participants can learn to improve their performance in tasks within this domain, (2) the learning is quantifiable via our performance features, and (3) the domain is flexible enough to create distinct levels of difficulty. The long-term goal of this work is to systematize the process of curriculum-based training toward the design of protocols for robot-mediated rehabilitation.**

I. INTRODUCTION

Human skill acquisition is a complicated phenomenon that is difficult to capture or characterize. Understanding human learning is of great interest to several fields including neuroscience, physical therapy, and sports training [2]–[5]. The theory that motor re-learning post a neurological injury is akin to novel motor learning [3] has further increased the interest in skill acquisition from a rehabilitation perspective. Recent advances in rehabilitation robots have greatly expanded the realm of possibilities for providing repeatable and consistent training. However, design of effective protocols for rehabilitation remains an open challenge. A robot controller must first understand the human's current abilities – what is easy, difficult, or impossible to perform. This estimation needs be comprehensive enough to inform the training process, and responsive to the learner's training progress. Crucial to this model of learning is the design

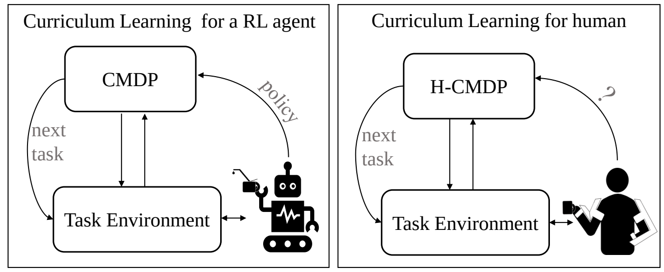


Fig. 1. Curriculum learning settings: The existing CMDP for RL agents [6] (left) and our parallel framework for human training, say with a rehabilitation robot (right).

of performance measures that allow the inference of motor learning [4], [5]. We propose a vocabulary based on existing literature that will be used to discuss and evaluate human skill learning in the context of training protocol design.

While the characterization of human ability is formidable in itself, leveraging this information to improve training paradigms is even more challenging. Curriculum-based training is a well-known method to improve the efficiency of skill acquisition in the fields of motor learning and neuroscience [7]–[10] and robotic surgery training [11]. Previous studies have implemented manual selection of successive task difficulty for efficient learning [11]–[17]. However, this manual process, being subjective rather than systematic, may result in sub-optimal ordering of tasks. Recent work in Reinforcement Learning (RL) has introduced means to construct automatic curricula for RL agents [18]–[20].

The first contribution of this paper is the formulation of a Curriculum Markov Decision Process for Human skill acquisition which extends a Curriculum MDP for RL agents (CMDP, Fig. 1, left) to be used for human learners (H-CMDP, Fig. 1, right). The main challenge in this formulation is that the instantaneous skill state is fully observable in RL agents, but is hidden in human learners. We address this challenge by proposing a set of performance features to evaluate the learner's skill state. *The second contribution of this paper is a novel platform, Reach Ninja*, designed to serve as a testbed for our formulation and implementation of the H-CMDP in this and future works. We use this testbed to demonstrate how a skill state is captured using specific metrics that quantify the features described in Section III. We further demonstrate the validity of this task domain in enabling and characterizing learning in the human agent.

In the future we plan to determine and tackle the challenges of implementing a H-CMDP for human skill acquisition across a variety of task domains. This research

* This paper contains material from a preliminary workshop version [1]. Compared to the workshop version, this manuscript has been extended with additional formalization and experiments. None of the material has appeared in a previous conference paper. This work is supported by NSF (M3X-2019704).

[1]Rehabilitation and Neuromuscular Robotics lab, Mechanical Engineering department, The University of Texas at Austin, TX 78712, USA (keya.ghonasgi,bmasetty)@utexas.edu, ashish@austin.utexas.edu

[2]Learning Agents Research Group, Computer Science department, The University of Texas at Austin, TX 78712, USA (reuth, sanmit, pstone)@cs.utexas.edu

[3]BLAM lab, Neuroscience department, Johns Hopkins University, MD 21218 USA adrian.haith@jhu.edu

[4] Sony AI

is particularly motivated by the potential application to the design and evaluation of training protocols for rehabilitation.

II. BACKGROUND

Motor learning is central to the rehabilitation goal of promoting recovery by facilitating learning of impaired motor behavior [13], [21]–[23]. Previous research has also shown that humans can learn or adapt to specific motor behaviors [24]–[26]. Training on sub-tasks that build in difficulty and intensity – or a *curriculum* – has also been shown to be effective in neurorehabilitaiton, e.g., in stroke treatment. Several theories in motor learning and sport science (challenge point [7], flow channel [8], dynamic difficulty adjustment [9], zone of proximal development [10]) have suggested a possible method to use this information towards protocol design through *curriculum-based training*.

In this paper, we propose a framework for curriculum design for motor skill acquisition using techniques from Reinforcement learning. Reinforcement learning is a paradigm for learning sequential decision making tasks for an artificial agent acting in an environment. It models a *task* as a Markov Decision Process (MDP) [27]. A MDP, M, is a 4-tuple $(\mathbb{S}, \mathbb{A}, p, r)$, where $\mathbb{S}$ is the set of states in the environment, $\mathbb{A}$ is the set of actions the agent can take, $p(s'|s,a)$ is a transition function that gives the probability of transitioning from state s to state s' after taking action a, and $r(s,a,s')$ is a reward function that gives the immediate reward for taking action a in state s and transitioning to state s'. At each time step t, the agent observes its state and chooses an action according to its *policy* $\pi(a|s)$. Its goal is to learn an *optimal policy* π^*, which maximizes the expected *return* (cumulative sum of rewards) until the episode ends.

Learning this optimal policy requires the RL agent to explore the environment and accumulate rewards, which may be difficult due to sparse rewards or the presence of adversarial agents or elements. One way to accelerate learning in such complex settings is to first train the agent on an easier *source task* that requires fewer actions to reach the goal, or has fewer elements in the environment that the agent needs to learn about. The knowledge acquired in this simpler environment can then be *transferred* to improve learning on the more challenging *target task* [28]. Moreover, an agent can train on a *sequence* of source tasks, called a *curriculum*, where each subsequent task becomes progressively harder and builds upon skills learned in previous tasks. Curriculum learning [20] is a methodology to optimize the order in which tasks are presented to the agent, so as to improve learning speed or performance on a final target task.

In this paper, we draw inspiration from a hierarchical model for curriculum design that poses curriculum generation as an interaction between two MDPs [29]. The first is a lower level MDP acting as the *student* agent, which is the recipient of the curriculum. This agent interacts in the standard way with a given task. The second is a higher level Curriculum MDP (CMDP, Fig. 2, top) for the *teacher*, whose goal is to select tasks for the student to train on.

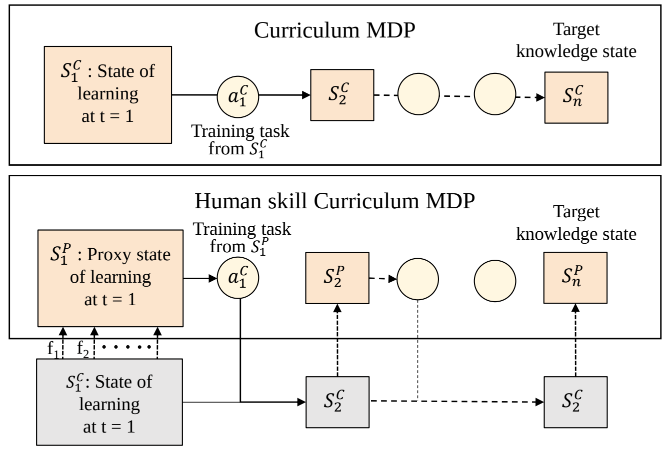


Fig. 2. Schematic representing a CMDP (top) and a H-CMDP (bottom). $\mathbb{S}^C$: states of the CMDP, or the learner's skill, which is hidden in the case of the human learner; a^c: action space of the CMDP, or the source task selection; $\mathbb{S}^P$: proxy state of the human learner; and f: performance features.

The reward function in a CMDP is defined to maximize asymptotic performance by rewarding transitions into terminal states that improve the final performance on the target task. This reward may also be tailored towards achieving other goals such as reaching a threshold level of performance. In this paper, we take the first step to extend the CMDP formulation for a human learner. We begin by demonstrating the challenges offered by the human learner and describe the state space for the human learner CMDP.

III. EVALUATING HUMAN PERFORMANCE

Adapting a CMDP to human motor learning is impeded by two main challenges introduced by the human learner to the representation of the CMDP states. First, unlike RL agents, the human learner's state of knowledge (or policy) is not observable. This setting is equivalent to the problem of designing a CMDP for a black box agent. Second, an RL agent may be retrained from a specific initial skill level any number of times. On the other hand, not only is it impossible for the human learner to unlearn a task, there is no way to stop the learning process to evaluate performance, or to control when learning will occur. To aid our discussion in handling these differences, we define the following terms:

Definition 1: The **Skill State** of the learner is its internal state, which holds all information about the learner's skill. For an RL agent, this state is represented by the agent's policy that is visible to the teacher. However, in human learners, this internal state is hidden from the teacher.

Definition 2: The **Proxy State** of the learner is a belief about the learner's skill state. For a set of proxy states, we assume that there is a surjective function $f : \mathbb{S}^C \to \mathbb{S}^P$ that maps each skill state to a proxy state.

Definition 3: The **Performance Features** of the learner on the task is a list of measurable quantities that can be observed, and that can be mapped to the learner's proxy state.

Definition 4: The **Probe Task** may be the same as the target task being learned, or one that is designed to evaluate the learner's performance. This task is used to query a learner's instantaneous proxy state, and must be incorporated into the training paradigm as learning cannot be stopped.

As a first step to bridge the gap in performance evaluation, we propose to formalize a CMDP for human motor learning such that *the person's performance on different metrics on the target task will be used as a proxy for their skill state*. Rather than direct access to a student MDP, we have access to a list of features that correlate with the proxy state of the learner, as depicted in Fig. 2. This setting can be modeled as a Partially Observable MDP (POMDP) [30]. In a POMDP, the true state of the agent (in our case, the learner's skill state) is unknown, so observations (performance features) are used to create a belief (proxy state) about the learner's hidden skill state. Further, as it is impossible to pause or reverse learning in humans, we rely on extensive and carefully designed human subject experiments to validate our method.

A. Human-skill Curriculum Markov Decision Process

Definition 5: The Human-skill Curriculum Markov Decision Process (H-CMDP), M^C, is a CMDP with each state in the state space $s^C \in \mathbb{S}^C$ is a *proxy to the skill state*, and is mapped to a list of features $s^C = \langle f_1, f_2, \ldots, \rangle$ that estimate the learner's skill using a variety of metrics in the target task.

For a given target task, a feature f, and a threshold value f^*, we define a set of *proxy states* that pass the feature threshold $S^C_{f^*}$ as all the states in which the student's performance on the target task on the feature f is greater than or equal to f^*. The learner is said to have **acquired a skill** when they reach a proxy state s_n in the set $\bigcap_f S^C_{f^*}$ for all of the features that define the proxy state space.

The main modifications in a H-CMDP compared to a CMDP are: (1) the state space does not consist of the true knowledge state, but of a set of *performance features* that estimate the human skill overcoming the challenge of observing the true skill state of the learner; and (2) the evaluation of the learner's skill state must take place through *probe tasks* that are incorporated into the training.

B. State Evaluation through Performance Features

The efficacy of the proxy state in approximating the true skill state depends primarily on the selection of the performance features used to define the proxy state during a probe task. These features must be selected such that they form a clear picture of human skill learning. We propose a set of general features based on existing literature on measuring human performance. Note that all of the features chosen represent a current state rather than an aggregated history of states, so they preserve the Markovian property of an MDP.

Confidence: level of belief the human has in their own ability. For example, making the task more challenging might affect the confidence of a learner who has already reached some level of confidence on an easier version of the task [31].

Accuracy: may be a binary, if the task ends with either "success" or "failure", or have a numeric value quantifying the most important factor the learner needs to improve upon, such as a task specific score.

Precision: reduction in variability or inconsistency in performance, that may be caused due to noisy environment. For example, this metric can be a success rate in a multi-trial task, or the variability in a repetitive human movement [32].

Strategy: characterization of a possible strategy employed by the learner. For example, this metric may capture similarity to a specific motion pattern [33]. The difference between strategy and precision is a subtle one, equivalent to that between making a conscious decision to pursue a specific movement pattern (strategy), and performing the movement exactly as intended (precision).

The learner's performance in these features is expected to provide a comprehensive picture of the learner's proxy state. In the next section, we present a specific game designed to evaluate curriculum learning for a novel learning task. Specifically, we address the challenges of designing such a task, defining the performance features and evaluating the features in human subject experiments.

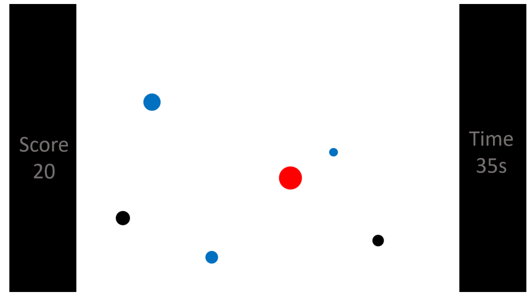
IV. DYNAMIC TASK DOMAIN: REACH NINJA

Prior research in skill acquisition has often focused on controlled lab-based tasks, like target reaching, that are easy to analyze, but too simplistic to benefit from curriculum-based training. On the other hand, skills in applied settings, such as surgical skill, tend to be more complex and harder to learn and could benefit from curriculum-based training. However, such skills can be challenging to analyze and do not provide a convenient platform for broader research [17]. In this paper, we aim to design a target task that lies closer to the center of the spectrum of simpler tasks (that are easy to analyze) and highly complex tasks (that are not broadly relevant). We propose a challenging and novel task such that repeated practice is expected to result in quantifiable learning. Moreover, source task difficulty modulation should be clearly defined, and the effect of changing this difficulty should result in quantifiable changes in performance. The ultimate goal of constructing a curriculum for rehabilitation motivates the generalizability of this discussion to a variety of source and target tasks.

A. Game Design

Inspired by the popular phone game Fruit Ninja [34], we present a game called "Reach Ninja". The game is developed in Python 3.4 using OpenCV 4.4.0 [35] and Pygame 1.9.3 [36]. The game is packaged into locally executable files using Pyinstaller 4.0 [37]. The application tracks the movement of an object held in the player's hand through their webcam (Fig. 3a), and the position is presented on the screen as a circular red marker called the cursor (Fig. 3b). During the game, other circles (targets), either blue or black in color, appear on the screen, entering from the bottom edge with randomly chosen initial velocities and radii. The player's goal is to capture these targets using the red marker on screen, through their hand motion, in order to maximize their score in a fixed period of time (heuristically set to 40s). The targets are also acted upon by a gravity-like force, resulting in a predictable projectile motion. The blue (positive) targets are

(a) Webcam view



(b) Gamescreen view

Fig. 3. Reach Ninja task domain. The left panel shows the webcam view (not visible to the player) and the player's hand movement tracked online during gameplay. The right panel shows the game screen (visible to the player) with the red marker tracking the player's hand, and blue and black targets. Total score and time remaining are always displayed on screen.

assigned a positive score depending on their size (smaller size gives larger score) and velocity (faster gives larger score), with the maximum possible score from a positive target being 30. The black (negative) targets are assigned a fixed negative score of -10 independent of size and speed. The supplementary video includes clips of the game and demonstrates the intuitive environment.

In a pilot study [1], it was unclear whether the baseline version of the task satisfied the first guideline of being challenging and novel. Towards making the task more challenging, we introduce two further dynamic interactions. The first is *partial feedback*, where the red cursor is displayed on screen only intermittently (1s on, 1s off), while the hand tracking and game continue normally. This interaction is designed to encourage players to learn the relationship between their true hand position and the cursor displayed on the screen. Second, we introduce a *magnetic field* around the red marker such that the positive targets are repelled while the negative targets are attracted. This interaction is expected to encourage higher speeds that are required to overcome the effect of the magnetic field. As these interactions can be turned on or off easily, we inherently satisfy the second requirement of easy source task difficulty modulation.

B. H-CMDP in the context of Reach Ninja

As discussed in section III-B, the use of the H-CMDP depends on the features used to define the proxy skill state of the learner. In this section, we instantiate these features in the context of the Reach Ninja task, based on the extensive 'expert knowledge' of the game designers.

Confidence – Mean Speed (MS): speed of the red marker (pixels/s) averaged over a game session of 40s. During experiments, it is observed that a higher MS correspond with more active play, implying that the learner has higher confidence in their ability to control their movement.

Accuracy – Final Score Percentage (FSP): score at the end of a 40s game as a percentage of the highest possible score. A higher score directly corresponds with better accuracy hitting the positive targets while avoiding the negative ones.

Precision – Captured Target Percentage (CTP): the percentage of targets that were successfully captured by the participant with a positive increase in the score. A larger CTP corresponds to the participants' precision in avoiding the negative targets while actively reaching for positive targets.

Strategy – Score Per Capture (SPC): ratio of the total score from positive targets with the number of positive targets the player captured as a percentage of the range of possible positive scores (between 0 and 30). A larger SPC implies on a choice to approach smaller and faster positive targets, suggesting the use of a strategy to maximize the final score.

V. EXPERIMENTAL VALIDATION

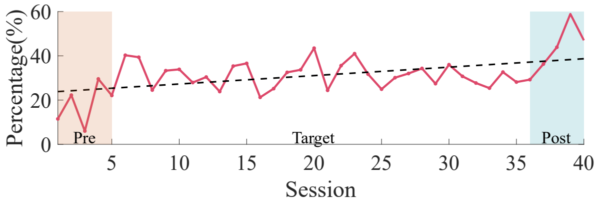
Given the novelty of the Reach Ninja task domain and the H-CMDP model, we focus our experimental efforts on validating the suitability of the novel task towards studying human learning. We test the hypotheses that the Reach Ninja task domain satisfies the predetermined guidelines through a human subject experiment validating the suitability of this environment for learning.

A. EXPERIMENT PROTOCOL

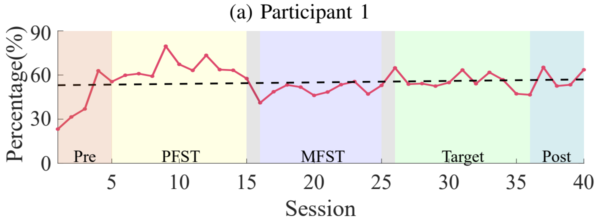
10 subjects (6 male, 4 female, aged 27.1 ± 3) participated in the study. Each subject played 40 sessions of the target game. The target game, which included both the partial feedback and magnetic field interventions was used to probe the players' performance at various times during the experiment. By turning off one of the two interventions, we define two possible source tasks, Partial Feedback Source Task (PFST), and Magnetic Field Source Task (MFST). To study the effect of these source tasks on performance, 5 of the 10 subjects (referred to as the curriculum group) are trained on a rudimentary curriculum. The curriculum is ordered as follows: Sessions 1 – 4 are pre-training probe tasks (same as the target task); sessions 5 – 14 are the PFST (magnetic field off); session 15 is a probe task; sessions 16 – 25 are MFST (partial feedback off); sessions 26 – 37 are the target task; and sessions 38 – 40 are the post-training probe tasks. The remaining 5 subjects (referred to as the control group) only practice the target task for all 40 sessions. The comparison of the performance in three stages is expected to *demonstrate learning through practice*. This experiment was approved by the Institutional Review Board at the University of Texas at Austin under the protocol number 2020-07-0156.

B. EXPERIMENT RESULTS

Fig. 4 shows the FSP across 40 sessions for two representative subjects (Participants 1 and 6). The slope of the fitted



(a) Participant 1



(b) Participant 6

Fig. 4. Final Score Percentage trend for participants 1 (control group) and 6 (curriculum group) over 40 session. The dashed line represents the trend across all sessions.

trend line (0.4) demonstrates the participant's tendency to improve their FSP through practice over time. Fig. 5 shows that in comparing the pre-training (sessions 2 – 4) versus post-training (sessions 38 – 40) in terms of Final Score Percentage, the overall increasing trend seen in Fig. 4a holds for most participants regardless of their training group. This trend is analyzed statistically and found to be significant across all participants, as seen in Fig. 6 (top-right). The remaining 3 metrics MS, CTP and SPC, are also compared between pre- and post-training (see section IV-B). Repeated measures ANOVA verifies that all feature values change over the training with statistical significance indicated by $p < 0.05$. In each case the change in the metric corresponds to an increase in performance. Using the Holm-Bonferroni correction with an ad-hoc contrast analysis, the FSP shows a statistically significant increase only when comparing either the pre-training task or MFST, to the PFST (p = 0.002 and 0.011 respectively, refer to Figure. 4b).

VI. Discussion

In section III of this paper, we define the H-CMDP and introduce the challenge of evaluating the hidden skill state of a human learner. We then demonstrate our method to address this challenge through the use of performance features and a task domain designed for this project. These features form the state space of the H-CMDP and are crucial to the future steps of designing training curricula for human learners.

Our experimental results demonstrate that, through training, all 4 performance metrics (confidence, accuracy, precision and strategy) show statistically significant improvement across the 10 subjects, corresponding to an improvement in their overall skill state (Figure 6). As all of the performance features show significant improvement, we deduce that *learning occurs during the experiment's training sessions*. Further, *this learning process can be captured* as exemplified in Fig. 4a by a single participant's FSP. We also observe an interesting result in the source tasks used in the training for the curriculum group (participants 6 - 10 in Fig. 5) where performance is statistically different only when the magnetic field intervention is removed. This observation is reflected in

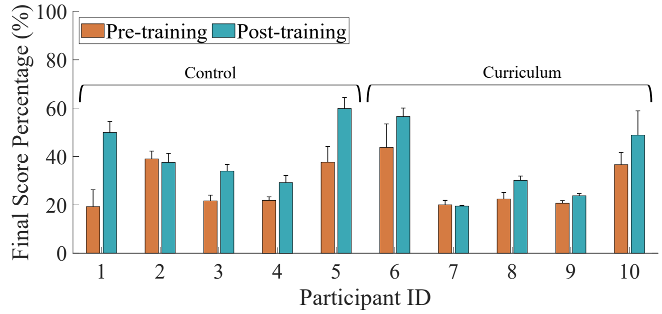


Fig. 5. Final Score Percentage: this plot compares the percentage of the overall score achieved by players in pre-training versus post-training sessions. The control group results are displayed as participants 1 - 5 and the curriculum group as participants 6 - 10.

the change in FSP for Participant 6 (Figure. 4b), and suggests that a majority of the challenge in the target task can be attributed to the magnetic field intervention.

Though we observe statistically significant learning between pre-and post-training performance (Fig. 6), there is large inter-trial variability in a single participant's FSP over the 40 sessions (Fig. 4). Our manually designed curriculum is unable to elicit distinct improvements in learner performance compared to repeated practice on the target task. These observations demonstrate the challenges to the generalizability of a curriculum for different participants due to human variability. Further, the improvement in performance between pre- and post-training trials does not necessarily demonstrate motor learning, and additional retention and transfer tests are needed for completeness [5]. Finally, the metrics discussed in this paper are primarily outcome based. Future work will focus on the measurement and comparison of kinematic and coordination metrics [4], [5]. Together with an increased number of experimental trials, these metrics will provide a more complete picture of motor learning, transfer, and the effect of curriculum-based training.

VII. Conclusion

This work introduces a formalization of the motor skill acquisition process through curriculum-based training. We begin by defining the Human-skill Curriculum Markov Decision Process and discussing the challenge of evaluating the state space of this MDP, the human learner's skill state. We address this challenge by selecting a vocabulary of performance features to characterize the human learner's skill referred to as their proxy skill state. We then define and validate a new testbed task, Reach Ninja, for studying human

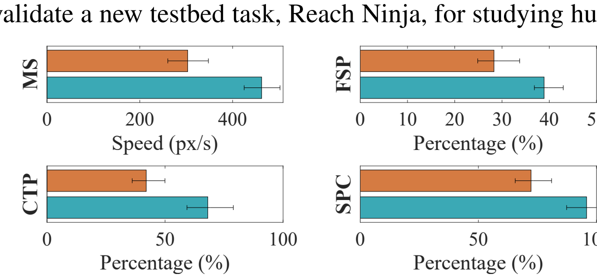


Fig. 6. Pre-training versus post-training results for all metrics, averaged across participants. All changes are statistically significant (p-values: MS = 0.0043*, FSP = 0.0082*, CTP = 0.011*, SPC = 0.031*).

learning, and use the testbed to demonstrate our solution implementation through human experiments. The results in this paper constitute an important first step towards the design and deployment of a H-CMDP for automatic selection. This work will be extended to a larger set of target tasks and across a diverse set of skills and performance measures, with the ultimate goal of supporting robotic rehabilitation in impaired populations.

ACKNOWLEDGMENTS

This work has taken place with joint efforts from the ReNeu Robotics Lab and Learning Agents Research Group (LARG) at UT Austin. Effort in the ReNeu Lab is supported, in part, by Facebook Reality Lab and NSF (M3X-2019704). LARG research is supported in part by NSF (CPS-1739964, IIS-1724157, NRI-1925082, CMMI-2019704), ONR (N00014-18-2243), FLI (RFP2-000), ARO (W911NF-19-2-0333), DARPA, Lockheed Martin, GM, and Bosch. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

REFERENCES

[1] K. Ghonasgi, R. Mirsky, B. Masetty, S. Narvekar, A. Haith, P. Stone, and A. D. Deshpande, "Leveraging reinforcement learning for human motor skill acquisition," in *Social AI for Human-Robot Interactions of Human-Care Service Robots Workshop in the International Conference on Intelligent Robots and Systems (IROS)*, vol. 165, 2020, p. 166.

[2] D. L. Adams, "Develop better motor skill progressions with Gentile's taxonomy of tasks," *Journal of Physical Education, Recreation & Dance*, vol. 70, no. 8, pp. 35–38, 1999.

[3] J. H. Carr, *Movement science: Foundations for physical therapy in rehabilitation*. Aspen Publishers, 1987.

[4] R. Magill and D. Anderson, *Motor learning and control*. McGraw-Hill Publishing New York, 2010.

[5] R. A. Schmidt, T. D. Lee, C. Winstein, G. Wulf, and H. N. Zelaznik, *Motor control and learning: A behavioral emphasis*. Human kinetics, 2018.

[6] S. Narvekar, J. Sinapov, M. Leonetti, and P. Stone, "Source task creation for curriculum learning," in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2016, pp. 566–574.

[7] M. A. Guadagnoli and T. D. Lee, "Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning," *Journal of Motor Behavior*, vol. 36, no. 2, pp. 212–224, 2004.

[8] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an understanding of flow in video games," *Computers in Entertainment (CIE)*, vol. 6, no. 2, pp. 1–27, 2008.

[9] M. Zohaib, "Dynamic difficulty adjustment (DDA) in computer games: A review," *Advances in Human-Computer Interaction*, 2018.

[10] L. Vygotsky, "Zone of proximal development," *Mind in society: The development of higher psychological processes*, vol. 5291, p. 157, 1987.

[11] A. P. Stegemann, K. Ahmed, J. R. Syed, S. Rehman, K. Ghani, R. Autorino, M. Sharif, A. Rao, Y. Shi, G. E. Wilding *et al.*, "Fundamental skills of robotic surgery: a multi-institutional randomized controlled trial for validation of a simulation-based curriculum," *Urology*, vol. 81, no. 4, pp. 767–774, 2013.

[12] P. Agarwal and A. D. Deshpande, "Subject-specific assist-as-needed controllers for a hand exoskeleton for rehabilitation," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 508–515, 2017.

[13] L. Marchal-Crespo and D. J. Reinkensmeyer, "Review of control strategies for robotic movement training after neurologic injury," *Journal of Neuroengineering and Rehabilitation*, vol. 6, no. 1, p. 20, 2009.

[14] E. T. Wolbrecht, V. Chan, D. J. Reinkensmeyer, and J. E. Bobrow, "Optimizing compliant, model-based robotic assistance to promote neurorehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 3, pp. 286–297, 2008.

[15] E. Basalp, L. Marchal-Crespo, G. Rauter, R. Riener, and P. Wolf, "Rowing simulator modulates water density to foster motor learning," *Frontiers in Robotics and AI*, vol. 6, p. 74, 2019.

[16] R. Feingold Polak and S. L. Tzedek, "Social robot for rehabilitation: Expert clinicians and post-stroke patients' evaluation following a long-term intervention," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 151–160.

[17] R. S. Wang and S. N. Ambani, "Robotic surgery training: Current trends and future directions," *Urologic Clinics*, 2020.

[18] S. Narvekar, J. Sinapov, and P. Stone, "Autonomous task sequencing for customized curriculum design in reinforcement learning." in *IJCAI*, 2017, pp. 2536–2542.

[19] T. Matiisen, A. Oliver, T. Cohen, and J. Schulman, "Teacher-student curriculum learning," *IEEE transactions on neural networks and learning systems*, 2019.

[20] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum learning for reinforcement learning domains: A framework and survey," *Journal of Machine Learning Research*, vol. 21, no. 181, pp. 1–50, 2020.

[21] J. A. Kleim and T. A. Jones, "Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage," *Journal of speech, language, and hearing research*, 2008.

[22] D. J. Reinkensmeyer and V. Dietz, *Neurorehabilitation Technology*. Springer, 2016.

[23] R. L. Sainburg and P. K. Mutha, "Movement neuroscience foundations of neurorehabilitation," in *Neurorehabilitation Technology*. Springer, 2016, pp. 19–38.

[24] E. Burdet, R. Osu, D. W. Franklin, T. E. Milner, and M. Kawato, "The central nervous system stabilizes unstable dynamics by learning optimal impedance," *Nature*, vol. 414, no. 6862, p. 446, 2001.

[25] E. Burdet, D. Franklin, R. Osu, K. Tee, M. Kawato, and T. Milner, "How are internal models of unstable tasks formed?" in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2. IEEE, 2004, pp. 4491–4494.

[26] A. Kadiallah, G. Liaw, M. Kawato, D. W. Franklin, and E. Burdet, "Impedance control is selectively tuned to multiple directions of movement," *Journal of Neurophysiology*, vol. 106, no. 5, pp. 2737–2748, 2011.

[27] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[28] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey." *Journal of Machine Learning Research*, vol. 10, no. 7, 2009.

[29] S. Narvekar and P. Stone, "Learning curriculum policies for reinforcement learning," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 25–33.

[30] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[31] A. Gentile, "Skill acquisition: Action, movement, and the neuromotor processes," *Movement Science: Foundations for Physical Therapy in Rehabilitation*, 1987.

[32] A. K. Dhawale, M. A. Smith, and B. P. Ölveczky, "The role of variability in motor learning," *Annual review of neuroscience*, vol. 40, pp. 479–498, 2017.

[33] J. Rieskamp and P. E. Otto, "Ssl: a theory of how people learn to select strategies." *Journal of Experimental Psychology: General*, vol. 135, no. 2, p. 207, 2006.

[34] Halfbrick Studios, "Fruit ninja," Apr 2010.

[35] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[36] PyGame Core Development Team, "PyGame 1.9.6," 2020.

[37] PyInstaller Core Development Team, "PyInstaller 4.0," 2020.