Adapting Task Difficulty in a Cup-Stacking Rehabilitative Task

Melina Daniilidis
melinad@seas.upenn.edu
University of Pennsylvania

Nathaniel Dennler
dennler@usc.edu
University of Southern California

Maja Matarić
mataric@usc.edu
University of Southern California

Stefanos Nikolaidis
nikolaid@usc.edu
University of Southern California

ABSTRACT

As the need for accessible upper-body stroke rehabilitation grows, it becomes increasingly important to investigate how the difficulty level of rehabilitation tasks can be personalized to a patient and automatically adapted based on the patient's progress in therapy. We introduce a framework that uses Fitts' Law to define task difficulty and iteratively apply it to dynamically adjust difficulty levels and to assign therapy tasks within the context of a cup-stacking occupational therapy activity. Our preliminary simulation results support the hypothesis that the model can adapt its difficulty levels based on a user's time taken to stack a cup at various points on a table. Future work includes exploring the impact of different variables on the model's adaptability and integrating personalized verbal feedback from a socially assistive robot.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design**; • **Computing methodologies** → *Modeling and simulation.*

KEYWORDS

Assistive robotics, Upper-limb rehabilitation, Task adaptation

ACM Reference Format:
Melina Daniilidis, Nathaniel Dennler, Maja Matarić, and Stefanos Nikolaidis. 2024. Adapting Task Difficulty in a Cup-Stacking Rehabilitative Task. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640558

1 INTRODUCTION

Eighty percent of stroke patients experience impairment in their upper bodies post-stroke [1]. Without proper and consistent rehabilitation therapy, patients may experience functional decline in their affected limb [6]. With 795,000 new stroke cases per year and a need for frequent training in the US alone, it is becoming increasingly important to create accessible rehabilitation devices that can administer therapy tasks to the patient at home or in a rehabilitation center [13].

HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3640558

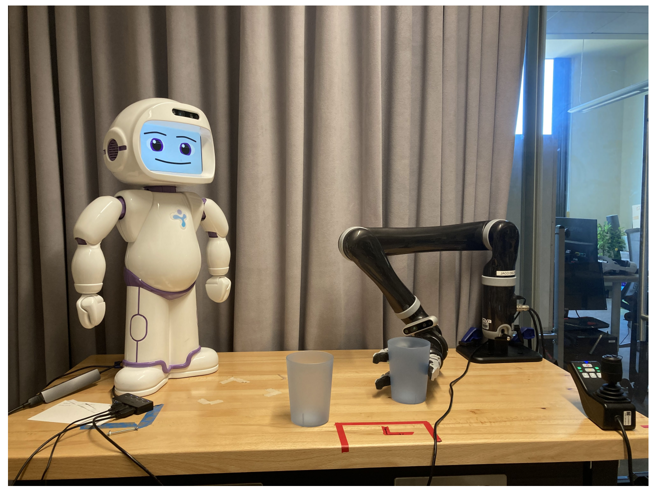

Figure 1: A depiction of the environment. In a rehabilitative therapy session, the Kinova JACO2 assistive arm (on the right) moves to various points across the table while holding a cup. The patient, sitting in front of the table and centered at the origin (the red box at the end of the table), stacks a cup on top of the robot's cup using their paretic arm. The Lux AI QTRobot (on the left) provides instructions for the user.

We focus on investigating the task difficulty adaptation in the context of a cup-stacking task, as shown in Figure 1. Literature in the areas of task difficulty adaptation and defining task difficulty in rehabilitative tasks are addressed in various ways, but these approaches require significant human input [14] or focus on robots that maintain physical contact with their users [9].

In this paper, our contributions are twofold. We first define a difficulty metric for cup-stacking using Fitts' Law, which describes the relationship between the time required to move to a target in a graphical user interface and the distance to the target [4]. We then apply the same equation iteratively to produce tasks that adapt to the difficulty level of the user. We perform an initial evaluation of this procedure in simulation. Our preliminary results indicate that the model can exhibit adaptive behavior throughout a therapy session.

2 BACKGROUND

A standard approach of administering the Fugl-Meyer Assessment on stroke patients is useful in defining task difficulty and matching tasks to a patient's ability, but the procedure is manual and requires an occupational therapist to reassess and re-evaluate the patient's difficulty level and tasks every few sessions [14]. Other approaches require external equipment, such as a virtual component. Error amplification [10], for example, is shown to improve user correction and performance by using a screen to display a target to be a little more off-target than it truly is. Previous approaches use a robot that comes into contact with the subject to measure and assess motor recovery [9]. One study, however, leverages an equation, Fitts' Law, to delineate between levels of task difficulty [15]. It established three levels of difficulty based on Fitts' Law, but did not use an adaptation technique. This inspired us to use the same equation not only as a metric of task difficulty, but also as a method of generating tasks that match the level of difficulty.

3 TECHNICAL APPROACH

3.1 Estimating Difficulty

To define our task difficulty, we use Fitts' Law, which is a linear predictive model of human movement in terms of time and distance [4, 15]:

$$T = a + b \cdot log_2(1 + \frac{D}{W})$$

Here, T is the time it takes for the user to reach the target point, D is the distance of the target from the origin, and W is the width of the target. The Index of Difficulty (ID) is defined as the logarithmic term in the equation and represents the measurement of task difficulty.

$$ID = log_2(1 + \frac{D}{W})$$

In our case, D will be the distance of the point the robotic arm moves to, which is where the patient must stack the cup. W is the width of the cup and T is the amount of time (in seconds) it takes for a stroke patient to stack a cup at a D distance away from the origin. When these values are given, a and b can be calculated with linear regression [15].

3.2 Adapting Difficulty

By adapting the difficulty, we aim to personalize the therapy session of the user in real-time and task them with points that will challenge them at the right level based on their ongoing performance. Our proposed framework takes in a target time to reach a set of points, defined by a physical therapist, and the actual amount of time it takes the user to reach these points. It then outputs a new difficulty level, that is defined by the distance of the next set of points from the origin. We describe our framework as follows:

(1) Perform an initial assessment of the patient, where the robotic arm moves to points that are uniformly scattered throughout the table area. Measure the distance of these points and the time it took the patient to stack a cup at each point.
(2) Calculate a and b with linear regression using the T, D, and W of these points.
(3) Select randomly N points throughout the task space and for each point, use the Fitts' Law equation to find an estimated time given the point's distance from the origin (D), a, and b. This results in a mapping of N points to their estimated times.
(4) Set a lower and upper time bound that represent the points of "just right" difficulty of appropriate knowledge, using therapist domain knowledge. From the mapping, select the points that are within this time bound. These will be the tasks for the patient in this round.
(5) Continue storing the distances of these points and the actual time it took for the user to reach them. Then, use this information and the data from all previous points reached to recalculate and update a and b. This step is meant to improve the predictive model as more data is gathered about the patient's movements throughout the therapy session.

Fig. 2 displays a flowchart of the framework.

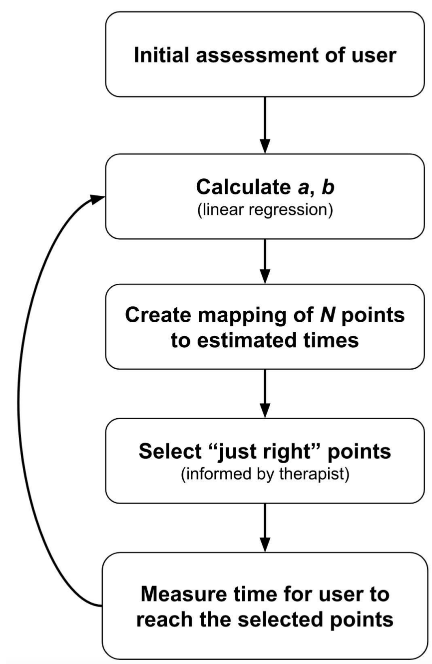

Figure 2: Flowchart of task adaptation framework

4 SIMULATED EXPERIMENT

We preliminarily tested our approach in virtual simulation. We established 3 experimental groups: faster, slower, and control. The faster group would represent a user that is improving and stacking the cup with greater ease, meaning that they are reaching the points faster at each session. The slower group would represent a user that is having difficulty or is being challenged by the selected points, meaning that they are reaching the points slower at each session. In the control group, the user's performance is not changing significantly, meaning that they are reaching the points at similar speeds. To simulate these groups, we first empirically selected a time bound of 4.08-4.73 seconds by conducting an initial assessment with a non-stroke patient user on the physical robot and finding the range of the slowest 25 points in the generated mapping, out of $N = 100$ points. We then simulate the control group with a Normal distribution with $\mu_c = 4.41$, the midpoint of the time bound range, and $\sigma = 0.12$, to represent the entire time bound, as Normal distributions have been shown to fit the variance in reaching times accurately [3]. We specify the μ_f of the faster group to be 6 standard deviations below μ_c and the μ_s of the slower group to be 6 standard deviations above μ_c, to minimize overlap between the three distributions. All three distributions have the same standard deviation. Thus, we have that the faster group is $\mathcal{N}(3.69, 0.12)$, the control group is $\mathcal{N}(4.41, 0.12)$, and the slower group is $\mathcal{N}(5.13, 0.12)$. We then performed the procedure as follows:

(1) For a given group, 10 points along the Normal distribution of that group were selected from the generated mapping. These points represent the points that the user would reach for that round and the time for each point represents the amount of time it would take the user to reach that point (in simulation).
(2) These 10 points were added to the data of the initial assessment. a and b were recalculated from this updated dataset and a new mapping was generated. The average D of the 10 points was also calculated and used for analysis.
(3) Steps 1 and 2 were repeated for five rounds.
(4) Steps 1-3 were performed on each experimental group.

We hypothesized that the average index of difficulty (ID) of the faster group would increase at each round, simulating that the algorithm would detect that the user needs to be challenged. Along similar lines, we hypothesized that the average ID of the slower group would decrease at each round, simulating that the algorithm would detect that the user is having difficulty in the session. We expected the average ID of the control group to not change significantly between rounds.

5 PRELIMINARY RESULTS

To evaluate our approach, we examine how well Fitts' Law describes task difficulty, and how well our framework can adapt to different users.

5.1 Evaluating the Efficacy of Fitts' Law

First, we evaluate that Fitts' Law extends to the cup reaching task. To do this, we collected the time for a non-stroke user to reach each point in the initial assessment and had the user perform the initial assessment three times to observe the fitness.

We visualize the fitness of Fitts' Law, in Figure 3. We found that the Index of Difficulty (ID) significantly predicted the time taken to reach a point ($R^2 = 0.467$, $\beta = 1.56$, $p = 0.000032$). This indicates that Fitts' Law is a reasonable approximation for time to complete the cup stacking task.

5.2 Evaluating Adaptivity with Simulated Users

Figure 4 shows the behavior of the average distance of each group as the rounds progressed. We show the average distance D as the metric of difficulty because the width W of the cup remains constant since the same cup is used in all rounds. The trends in the graph support our hypotheses for each group. In the faster group, the average distance is constantly increasing from round to round, which implies that the level of difficulty is increasing. Analogously, the average distance of the slower group is constantly decreasing, implying that the level of difficulty is decreasing, and the average distance/difficulty of the control group remains about the same.

Figure 5 shows the behavior of the average Index of Difficulty (ID) of each group throughout the rounds. Here, we observe a slight

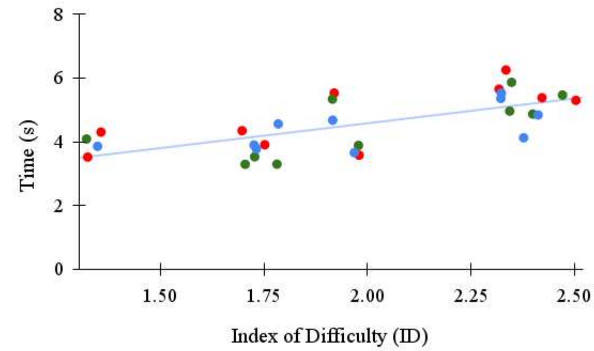

Figure 3: Linear fitness of the Index of Difficulty (ID) over three runs of the initial assessment.

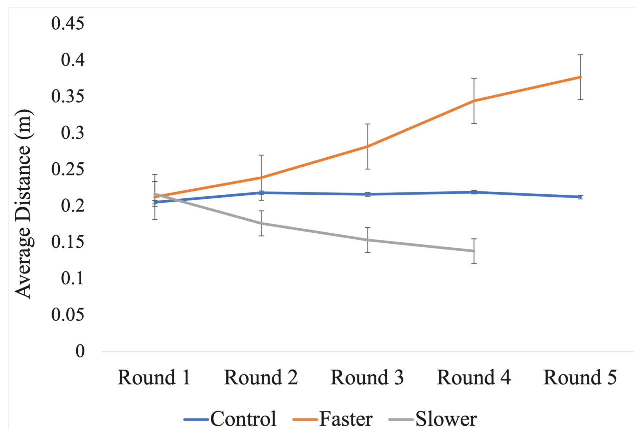

Figure 4: Average Distance (m) per round

decrease in ID for the faster group from round 3 to round 4, but this is expected because the simulated times are stochastic.

It is important to address some other observations and limitations of our design. In both the "faster" and "slower" groups, the number of selectable points at each round will eventually be zero. This means that after a certain number of rounds, there are no points from the latest mapping that exist within the 4.08-4.73 second time bound. Specifically, we determined outside of the experiment that the "faster" group runs out of selectable points after approximately 80 points (8 rounds) and the "slower" group runs out of selectable points after approximately 40 points (4 rounds). This is why there is no data for Round 5 of the "smaller" group in Figure 4.

This is expected behavior, since as the simulated user keeps performing faster or slower, the times in the mapping will eventually be completely below the time bound or completely above the time bound, respectively. Therefore, there is a possibility that the algorithm would not be able to adapt after a certain point if the user is consistently reaching farther points even faster or if the user is consistently reaching closer points even slower. The likelihood of this scenario is something that we would like to observe in physical experimentation. We hypothesize that it is unlikely that a user from

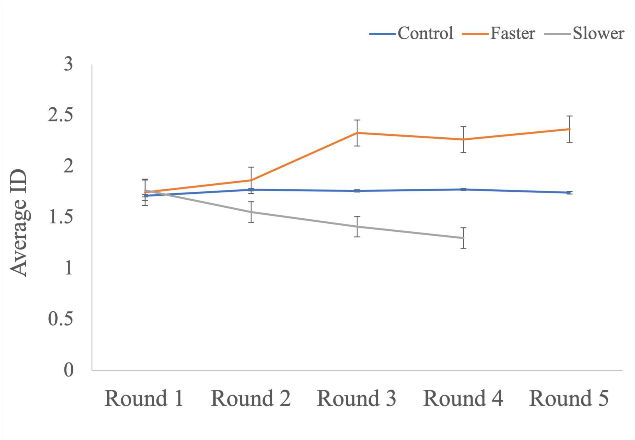

Figure 5: Average Index of Difficulty (*ID*) per round

the faster/slower group will keep reaching increasingly challenging points faster/slower than in the previous round. One method that could potentially prevent the exhaustion of selectable points would be to generate a mapping for more points at each round. Sampling more than 100 points creates a denser set of candidate points that are within the time bound.

The initial assessment was also not performed on a stroke patient, and is hence not representative of a stroke patient's performance. On the non-stroke patient user, the time to reach a point in relation to the point's distance from the user does not vary as much. They could reach closer and farther points at similar times (between 3-5 seconds), which is why the empirical time bound of 4.08-4.73 seconds is narrow. It is also important to note that the simulation was performed on a single participant's data and is therefore susceptible to bias resulting from a small sample size. We plan to run a user study with more participants for future work in order to draw stronger conclusions.

6 FUTURE WORK

We are interested in addressing our observations from the simulation by testing in the physical world with stroke patients. In particular, we will pay attention to the evaluation of both the proposed framework and the user experience from actual users of the system.

We are interested in assessing if changing certain factors (e.g., the points used in the initial assessment, the linear regression score from the initial assessment, the number of points selected at each round, etc.) significantly influence when the groups will no longer be able to select points from within the time bound and any other results. We would want to observe when or if a plateau occurs in the number of selectable points based on these changes. For one of these variables, the time bounds, we are curious to see if setting a threshold would be difficult for a physical therapist to determine in reality or if there are other empirical methods for dynamically determining a time bound of the "just right" points for a specific user. We reason that a rehabilitation therapist is able to establish a suitable time bound from observing the patient making reaches to the cup. Rehabilitation therapists can monitor patient exertion levels and ask them qualitative questions (e.g., "Was it easier or harder for you to reach this point compared to the previous point?") during the initial assessment to establish an appropriate time bound for the specific patient. Based on these inspections, the therapist would determine a number of seconds to input as the target time bound.

We are also interested in looking at other metrics for evaluating the proposed framework beyond the time taken to reach an object. Some potential metrics include effort assessed through facial affect [2], attention assessed through gaze [7], enjoyment assessed through physiological response [8], and arm function assessed through the score from the Fugl-Meyer Assessment for upper extremity [5].

In order to assess the user experience of the system, we will conduct semi-structured interviews with the patients and therapists from the physical experiment by asking questions to evaluate the usability and effectiveness of our approach.

To expand these realms, we would also like to incorporate personalized verbal feedback from the socially assistive robot (QTRobot pictured in Figure 1) in the therapy session along with task difficulty adaptation. Research in Socially Assistive Robotics (SAR) investigates approaches to automating verbal feedback to the patient that will keep them motivated and engaged in the task. There exists substantial literature that examines the effect of certain types of verbal feedback on user performance, such as normative feedback [11] and personality-matching feedback [12]. We would like to design a therapy session that provides verbal feedback (with the QTRobot) to the patient based on how well they perform the task given by the robotic arm and the task difficulty adaptation algorithm. This is with the intention of comparing a session that combines both components with ones that implement each component in isolation. It would be interesting to determine if the combination is more effective in increasing user engagement, motivation, and performance.

7 CONCLUSION

This paper addresses the need for accessible and automated upper-body rehabilitation for stroke patients, emphasizing the significance of task difficulty adaptation in improving patient performance. Leveraging Fitts' Law, the proposed framework defines a difficulty metric for a cup-stacking task and iteratively applies it to dynamically adjust the task difficulty levels and the therapy tasks. The simulation results show the adaptability of the model to users' varying performance levels. However, limitations, such as the eventual exhaustion of selectable points in certain scenarios, hint at the need to transition to physical experimentation in order to determine if these behaviors would occur in real environments. Future work aims to explore the impact of different variables on the model's adaptability and incorporate personalized verbal feedback from a socially assistive robot to enhance user engagement and motivation.

8 ACKNOWLEDGEMENTS

This research is supported by the National Science Foundation under Grant No. CNS-2051117.

REFERENCES

[1] Justin A Beebe and Catherine E Lang. 2009. Active range of motion predicts upper extremity function 3 months after stroke. *Stroke* 40, 5 (2009), 1772–1779.

[2] Helma M De Morree and Samuele M Marcora. 2010. The face of effort: Frowning muscle activity reflects effort during a physical task. *Biological psychology* 85, 3 (2010), 377–382.

[3] Nathaniel Dennler, Amelia Cain, Erica De Guzman, Claudia Chiu, Carolee J Winstein, Stefanos Nikolaidis, and Maja J Matarić. 2023. A metric for characterizing the arm nonuse workspace in poststroke individuals using a robot arm. *Science Robotics* 8, 84 (2023), eadf7723.

[4] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[5] AR Fugel-Meyer, L Jaasko, I Leyman, S Ollson, and S Steglind. 1975. The post-stroke hemiplegic patient1, a method for evaluation of physical perpormance. *Scand. J. Rahabil. Med* 7 (1975), 13–31.

[6] Cheol E Han, Sujin Kim, Shuya Chen, Yi-Hsuan Lai, Jeong-Yoon Lee, Rieko Osu, Carolee J Winstein, and Nicolas Schweighofer. 2013. Quantifying arm nonuse in individuals poststroke. *Neurorehabilitation and neural repair* 27, 5 (2013), 439–447.

[7] Shamsi T Iqbal and Brian P Bailey. 2004. Using eye gaze patterns to identify user tasks. In *The Grace Hopper Celebration of Women in Computing*, Vol. 4. 2004.

[8] Regan L Mandryk, Kori M Inkpen, and Thomas W Calvert. 2006. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & information technology* 25, 2 (2006), 141–158.

[9] Jean-Claude Metzger, Olivier Lambercy, Antonella Califfi, Daria Dinacci, Claudio Petrillo, Paolo Rossi, Fabio M Conti, and Roger Gassert. 2014. Assessment-driven selection and adaptation of exercise difficulty in robot-assisted therapy: a pilot study with a hand rehabilitation robot. *Journal of neuroengineering and rehabilitation* 11, 1 (2014), 1–14.

[10] Navid Shirzad and HF Machiel Van der Loos. 2012. Error amplification to promote motor learning and motivation in therapy robotics. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3907–3910.

[11] Katelyn Swift-Spong, Elaine Short, Eric Wade, and Maja J Matarić. 2015. Effects of comparative feedback from a socially assistive robot on self-efficacy in post-stroke rehabilitation. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 764–769.

[12] Adriana Tapus and Maja J Mataric. 2008. Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance.. In *AAAI spring symposium: emotion, personality, and social behavior*. 133–140.

[13] Connie W Tsao, Aaron W Aday, Zaid I Almarzooq, Cheryl AM Anderson, Pankaj Arora, Christy L Avery, Carissa M Baker-Smith, Andrea Z Beaton, Amelia K Boehme, Alfred E Buxton, et al. 2023. Heart disease and stroke statistics—2023 update: a report from the American Heart Association. *Circulation* 147, 8 (2023), e93–e621.

[14] Michelle L Woodbury, Kelly Anderson, Christian Finetto, Andrew Fortune, Blair Dellenbach, Emily Grattan, and Scott Hutchison. 2016. Matching task difficulty to patient ability during task practice improves upper extremity motor skill after stroke: a proof-of-concept study. *Archives of physical medicine and rehabilitation* 97, 11 (2016), 1863–1871.

[15] Lukas Zimmerli, Carmen Krewer, Roger Gassert, Friedemann Müller, Robert Riener, and Lars Lünenburger. 2012. Validation of a mechanism to balance exercise difficulty in robot-assisted upper-extremity rehabilitation after stroke. *Journal of neuroengineering and rehabilitation* 9, 1 (2012), 1–13.