More Than a Number: A Multi-dimensional Framework for Automatically Assessing Human Teleoperation Skill

University of Colorado Boulder

CAIRO Collaborative AI and Robotics Lab

Emily Jensen [1,2], Bradley Hayes [1], and Sriram Sankaranarayanan [1]

[1] University of Colorado Boulder, [2] U.S. Army Research Laboratory

Contact: emily.jensen@colorado.edu

NSF

DEVCOM ARMY RESEARCH LABORATORY

Motivation

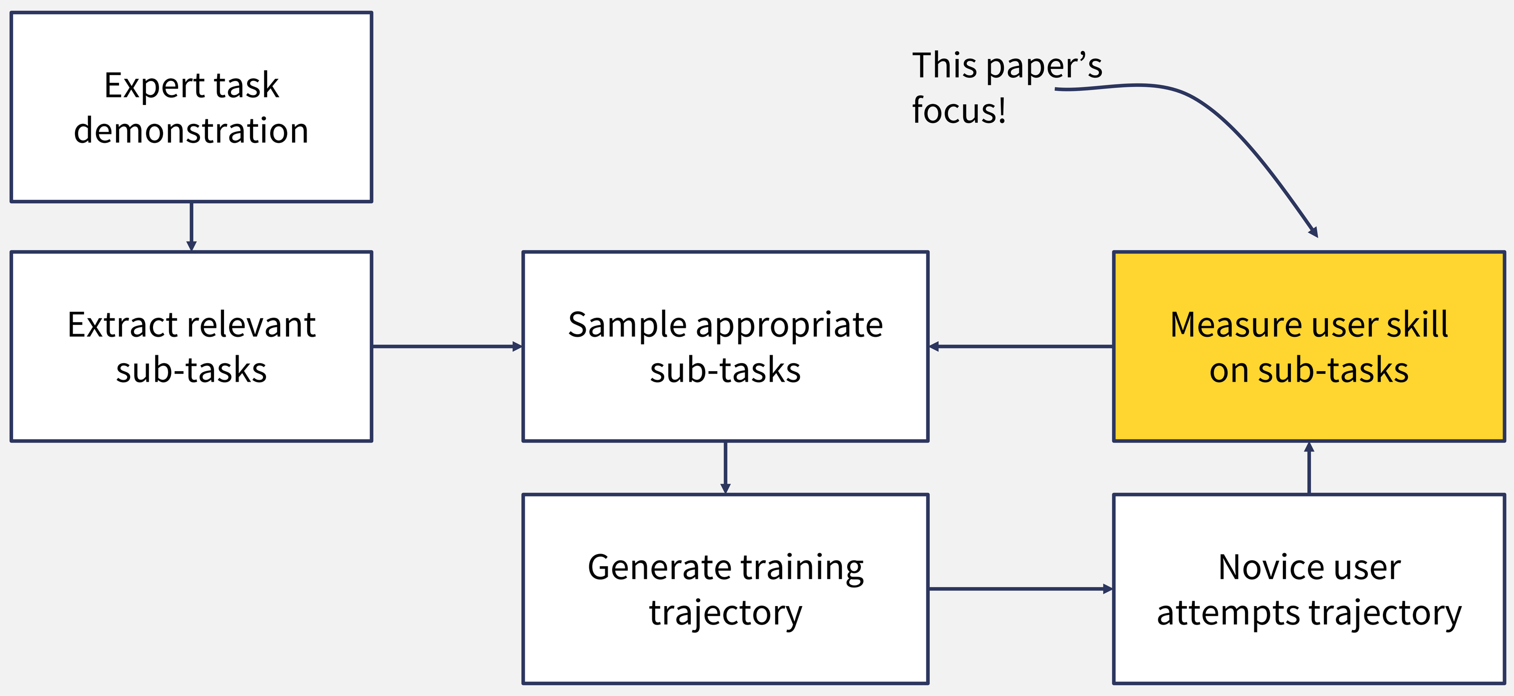
Safe teleoperation is a critical skill in domains such as construction and inspection.

We apply principles from intelligent tutoring systems to teach complex skills:

- Provide personalized support to users based on individual needs [1]
- Provide consistent and unbiased feedback during learning
- Focus practice on concepts that are difficult for the learner [2]
- Assess learner on concepts where they have mastered the prerequisites

How do we know if someone is skillful at teleoperation?

- Most work defines manual rubrics [3]. This is not ideal because it requires experts to create new rubrics for each new task and may have subjective criteria.
- Evaluation can be inconsistent over time or between judges
- Performance metrics are not well-defined for computers to assess



Big Research Question: Can we personalize the learning process by generating training examples at the user's appropriate skill level?

Measuring Drone Teleoperation Skill

Research Question: What skill representation provides a nuanced measurement of proficiency that can be operationalized by an automated curriculum generator?

Example domain: operating a drone with a remote controller

We implemented a drone piloting task in Unity

Define each trial as a **sequence of smaller tasks**

Take off vertically and stop in floating target

Hover stationary in the target for 5 consecutive seconds

Land vertically and gently on landing pad

Define **skill as a measure of performance** for each small task

Performance can be measured along **different dimensions**:

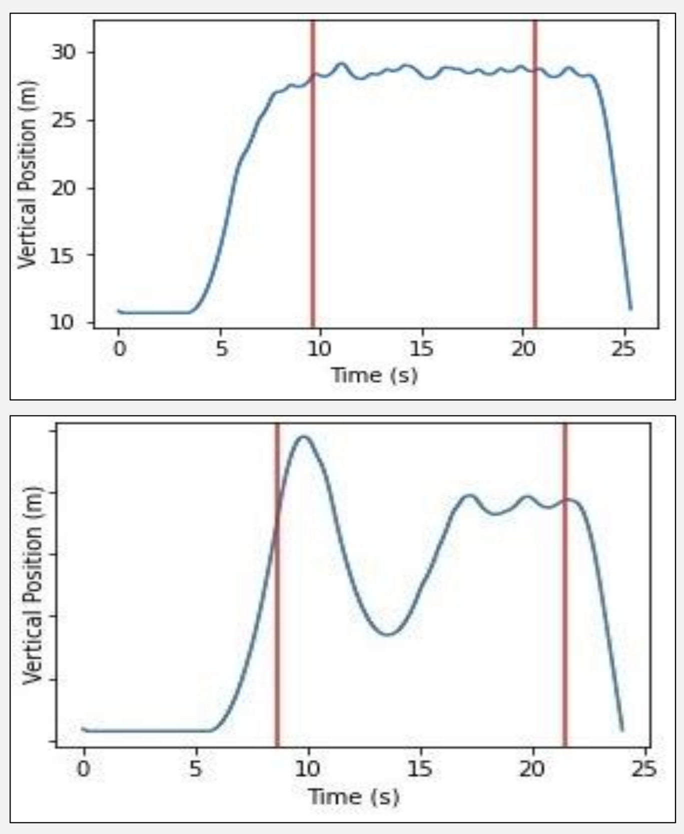
Different dimensions may be relevant for different tasks

Time efficiency: complete the task as quickly as possible

Control efficiency: use as few resources as possible

Robustness: adherence to task specifications

we specify tasks using signal temporal logic distinguishes "clearly" vs. "barely" correct performance



Example trajectories: the top shows high robustness and low control efficiency, while the bottom shows low robustness and high control efficiency

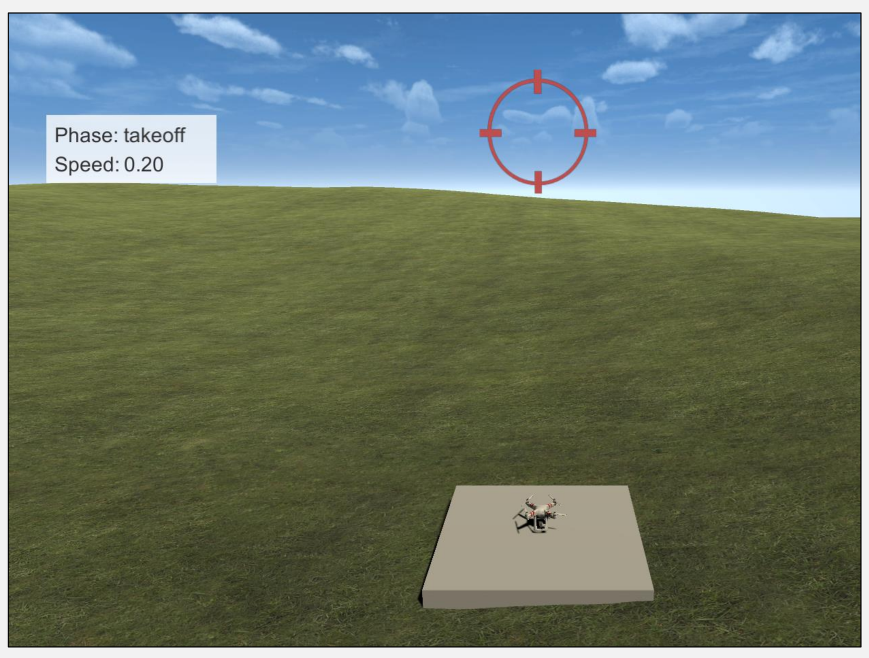
$$(v_y \geq 0)\ \text{UNTIL}_{[0,T]}\ \left(|y - H| \leq \epsilon\ \wedge\ |v_y| \leq \delta\right)$$

Takeoff: within T seconds, reach height H with velocity 0

Similar formulation for landing

$$\text{EVENTUALLY}\left(\text{Always}_{[0,T]}\left(y \in [y_{\min}, y_{\max}]\ \wedge\ |v_y| \leq \epsilon\right)\right)$$

Hover: stay in vertical range for T seconds with velocity 0



Unity task environment

Key Takeaways

Overspecification of tasks hinders skill assessment

- None of the takeoff or landing segments met the specifications because users flew too fast and overshot the hover target
- How to effectively communicate expectations to learners?

Different dimensions of skill indeed provide nuance

- Users demonstrate skillful performance along different dimensions
- There is a tradeoff between dimensions, such as speed and safety
- Different tasks may place higher priority on different skill dimensions

There are plenty of unanswered questions to consider

- We discussed a few possible dimensions of skill; how many are there?
- What level of skill is needed to declare mastery?
- How does user perception of performance relate to measured skill?

		Takeoff			Hover			Land		
Participant	Trial	RO	TE	CE	RO	TE	CE	RO	TE	CE
1	1	-4.99	-5.90	-1.79e-06	0.25	-40.32	**-1.92e-06**	-0.82	-5.38	-5.74e-06
	2	-2.54	-6.11	-1.74e-06	*0.09*	-5.94	-2.91e-06	*-2.19*	*-6.08*	-2.60e-06
2	1	*-7.14*	-6.34	-1.62e-06	0.44	*-44.72*	-4.82e-06	-1.55	-3.53	-6.00e-06
	2	-3.72	-7.60	-2.62e-06	0.12	-15.66	-4.89e-06	-1.59	-4.42	-6.58e-06
3	1	-1.66	*-12.59*	-2.51e-06	0.38	-22.40	-9.72e-06	-1.31	-3.23	-3.03e-06
	2	-7.12	-8.70	**-0.85e-06**	0.25	-7.74	-6.62e-06	**-0.65**	**-2.58**	**-2.07e-06**
4	1	-1.63	-7.95	-1.97e-06	0.71	**-5.22**	-6.84e-06	-1.46	-3.85	*-6.96e-06*
	2	-0.51	-10.33	-1.97e-06	0.93	-7.10	-6.81e-06	-0.13	-4.44	-4.46e-06
5	1	**-0.00**	-9.60	*-5.67e-06*	**1.08**	-6.06	*-11.55e-06*	-1.85	-4.69	-5.65e-06
	2	-4.84	**-5.42**	-2.56e-06	0.95	-13.92	-8.15e-06	-1.32	-2.73	-6.01e-06

Note: RO = Robustness, TE = Time Efficiency, CE = Control Efficiency

best in column = **bold**, worst in column = *italics*

Download the paper!

Next Steps

Collect a dictionary of sub-task trajectory templates

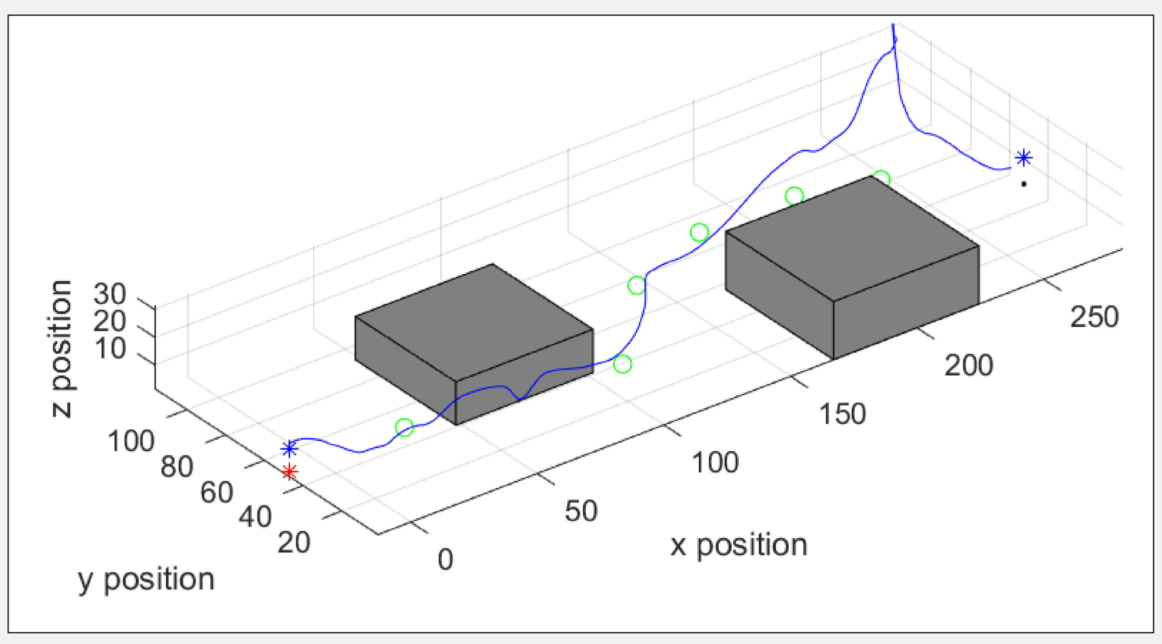
- Use expert demonstrations for task exemplars
- Organize into a hierarchy of prerequisite tasks
- Define skill estimation metric(s) for each task

Create training examples that adapt to the learner's skill level

- Select sub-tasks that are in the Zone of Proximal Development [4]
- Combine sub-task templates to create a longer trajectory
- More sub-tasks become available as prerequisites are met
- Update skill estimates based on adherence to training trajectory

Compare adaptive curriculum to standard fixed approaches

- How many training examples are needed to reach mastery?
- How confident do learners feel about their abilities?
- How much time is wasted practicing mastered concepts?



3D drone teleoperation scenario: this task requires complex maneuvers such as turning while avoiding collisions. Users will need to master smaller tasks such as moving in a straight line before attempting this scenario.

References

[1] Mikić, V., Ilić, M., Kopanja, L., and Vesin, B., Personalisation methods in e-learning—A literature review, Comput. Appl. Eng. Educ. 2022; 30: 1931–1958. https://doi.org/10.1002/cae.22566

[2] Ericsson, K. Anders, et al. The Role of Deliberate Practice in the Acquisition of Expert Performance. Psychological Review, vol. 100, no. 3, 1993, pp. 363–406., https://doi.org/10.1037/0033-295x.100.3.363

[3] L. Chang, R. M. Satava, C. A. Pellegrini, and M. N. Sinanan. 2003. Robotic surgery: Identifying the learning curve through objective measurement of skill. Surgical Endoscopy 17, 11 (2003), 1744–1748. https://doi.org/10.1007/s00464-003-8813-6

[4] Vygotsky, L. S. (1978). Mind in Society: Development of Higher Psychological Processes. United Kingdom: Harvard University Press.