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

Emily Jensen*
emily.jensen@colorado.edu
University of Colorado Boulder
Boulder, Colorado, USA

Bradley Hayes
bradley.hayes@colorado.edu
University of Colorado Boulder
Boulder, Colorado, USA

Sriram Sankaranarayanan
srirams@colorado.edu
University of Colorado Boulder
Boulder, Colorado, USA

ABSTRACT

We present a framework for the formal evaluation of human teleoperator skill level in a systematic fashion, aiming to quantify how skillful a particular operator is for a well-defined task. Our proposed framework has two parts. First, the tasks used to evaluate skill levels are decomposed into a series of domain-specific primitives, each with a formal specification using signal temporal logic. Secondly, skill levels are automatically evaluated along multiple dimensions rather than a singular number. These dimensions include robustness, efficiency, resilience and readiness for each primitive task. We provide an initial evaluation for the task of taking-off, hovering, and landing in a drone simulator. This preliminary evaluation shows the value of a multi-dimensional evaluation of human operator performance.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; • **Theory of computation → Logic**.

KEYWORDS

Automated Assessment, Temporal Logic, Skills, Teleoperation

ACM Reference Format:
Emily Jensen, Bradley Hayes, and Sriram Sankaranarayanan. 2023. More Than a Number: A Multi-dimensional Framework For Automatically Assessing Human Teleoperation Skill. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3568294.3580167

1 INTRODUCTION

Skill is a highly valued attribute for numerous human endeavors. Thus, its measurement is of great importance. The need for humans to co-operate with an autonomous system to skillfully complete safety-critical tasks is common across diverse domains such as surgery, planetary exploration, and visual inspection using drones.

The problem of measuring skill is well-known to be extremely hard. Existing approaches use examinations of an operator by qualified judges who provide numerical scores. This process is often effort intensive, subject to bias and hard to automate. As a result, the numerical scores for different operators who undergo different exams judged by a different panel of judges are hard to compare against each other. The same problem arises when it comes to judging the same operator at different points in time to measure their learning progress. Thus, we need a framework for quantifying skill that is based on simple principles that can be applied uniformly in an unbiased manner. Also, the idea of a single number representing skill level lacks *nuance*. For instance, a particular operator's performance may exhibit better safety margins for the desired specifications while sacrificing on time efficiency. This sort of nuance is often absent in a single score.

In this paper, we attempt to formulate such a framework using a combination of ideas from logic and control theory. Our proposed framework first identifies *primitive tasks* that a skillful operator needs to demonstrate. Next, we use signal temporal logic (STL) to specify these tasks in an unambiguous manner [7, 15] and define *skill* as a vector that measures the demonstrated performance of a task. Next, we reflect on the various aspects that characterize a skillful performance.

The Merriam-Webster dictionary defines skill (as in skillful performance) as "*the ability to use one's knowledge effectively and readily in execution or performance*." Analyzing the definition of skill from related sources such as the Oxford English dictionary and Wikipedia suggests the following dimensions of a skillful performance:

Robustness: Performance of the required task that is *clearly* correct (compared to *barely* correct) under nominal conditions. This aspect is especially important for safety-critical tasks.

Efficiency: Performance that minimizes time and energy.

Resiliency: Performance under varying environmental conditions for the teleoperated system.

Readiness: Performance under variations in the *human factors*: eg., at different times of day or different levels of comfort.

While we use dictionary definitions here to motivate the problem, we review relevant literature from psychology, cognitive science and human factors in Section 1.1 to provide a more rigorous background. In this paper, we design a framework that seeks to evaluate each of these aspects of skillful performance in the context of teleoperating a drone. We formalize notions of robustness and efficiency while demonstrating how these notions allow us to evaluate human teleoperation of a drone in a simulation environment.

1.1 Related Work

To define tasks for robots, previous work has focused on the use of skill primitives, which are atomic actions that may be combined and sequenced depending on the target task. For example, [12] defines manipulation primitives which are defined by parameterized twist and wrench trajectories. Other work uses these skill primitives in directed graphs [25] or a relational assembly model [19]. From a human-robot interaction perspective, robotic skill primitives can be used in interfaces to allow human users to quickly define different

* Also with U.S. Army Research Laboratory.

HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9970-8/23/03.
https://doi.org/10.1145/3568294.3580167

tasks [24] or easily transfer tasks between robotic systems [21]. These approaches are generally evaluated based on the performance of the robotic system; that is, whether the task is successfully completed and other performance metrics such as time-to-completion.

Compared to robotic systems, defining and measuring human skill proves to be less clear. Defining and assessing human skill typically involves an "I know it when I see it" approach. Our goal here is to develop a comprehensive and more formal approach. Teleoperation tasks of interest involve psychomotor skills, which require coordination between physical (e.g., limbs, muscles) movements in response to sensory stimuli (e.g., noticing wind in a teleoperation task) [20]. Since human-performed tasks are usually much more complex than those performed by robotic systems, the evaluation criteria vary considerably. For example, [6] measured skill in robotic surgical knot-tying using an expert-developed rubric of steps, where each step earned points and errors lowered the final score. Other work defines skill as the average performance over a set of attempts in a video game [5] and differentiates skill as a task-oriented value compared to ability, which is a stable trait. More generally, researchers have qualitatively [10] described motor skill improvement in terms of reduced variability and increased smoothness [23] or cognitive load reduction that allows the user to address other demanding tasks [1].

2 PROPOSED SKILL FRAMEWORK

We first describe primitive tasks specified using temporal logic. Next, we will use this in our framework to measure skill levels.

2.1 Primitive Tasks and Skills

Our proposed framework for measuring skillfulness starts from a domain-specific knowledge of what tasks are to be performed by the operator. For instance, consider the job of teleoperating a drone to perform an inspection of an oil rig. To perform the overall inspection successfully, the operator must be able to perform numerous atomic, *primitive tasks* such as taking off, maneuvering, hovering and landing. The operator will need to be skilled in performing these primitive tasks in an appropriate manner to complete the overall task at hand. Although the set of all tasks that an operator can be called to perform can be a forbiddingly large, we can enumerate a relatively smaller number of primitives that can be sequenced and combined to form complex tasks.

For drone teleoperation, we define several (but not all possible) primitive tasks and the required drone input controls in Table 1. These tasks can be sequenced and combined to form complex tasks; for example, the operator may take off, perform a circular trajectory around a landmark while also keeping the camera trained on it, and land near a charging station.

Next, we need to carefully specify each of these primitives formally. Temporal logics were originally proposed for this purpose in the field of computer-aided verification of hardware and software systems [2, 17] and subsequently adopted to robotics as a means for specifying complex robotic tasks [4, 14, 16, 18, 22]. The primitive tasks can be easily specified using a suitable temporal logic. We propose to use metric/signal temporal logics (STL) which include ability to specify real-time constraints as part of the logic [7, 15]. Let $(x(t), y(t), z(t))$ represent the position of a drone with y axis pointing vertically up, and $(v_x(t), v_y(t), v_z(t))$ represent velocities

Table 1: Selection of primitive tasks for drone teleoperation

Primitive Task	Controls Required
Angled takeoff	upward throttle + pitch or roll
Angled landing	downward throttle + pitch or roll
Straight angled line	fixed pitch + roll
Curving line	changing pitch + roll
Perspective change	yaw
Hover in place	throttle
Maintain altitude	hover in place + pitch, roll, or yaw

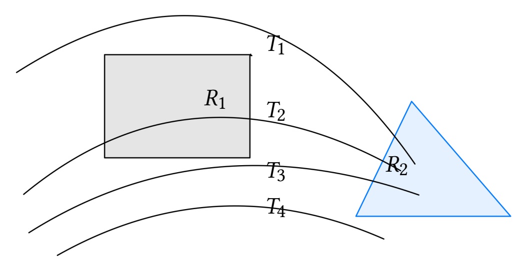


Figure 1: Illustration of robustness for various trajectories $T_1 - T_4$ for a drone. The desired task specification is to avoid the shaded region R_1 and reach region R_2.

along the respective axes. We will omit other state variables that describe attitude and control inputs $u(t)$ for simplicity.

Hover in place for some time: Let $[y_{\min}, y_{\max}]$ denote the desired range of altitude and $[-\epsilon, \epsilon]$ denote the desired limits on the velocity for some $\epsilon > 0$. Let T be the minimum amount of time we require the UAV to hover in place. In signal temporal logic, we specify the desired task as follows:

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

Note that constraints on the attitude can exist but are omitted for simplicity. This formula requires us to find a time window of at least T seconds where the y, v_y are within desired bounds.

Vertical Takeoff: We specify a vertical takeoff from the ground $y = 0$ to some altitude range $H \pm \epsilon$ and a level velocity limit $\delta > 0$ within T seconds. We specify this as:

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

The property specifies that the UAV rise up from the ground until it achieves an altitude in the range $H \pm \epsilon$ with the vertical velocity in the range $\pm\delta$. A vertical landing can be similarly specified.

Temporal Logics provide two advantages: (a) unambiguous specification language that is close to natural language and can be efficiently monitored in real-time [3]; and (b) the ability to measure compliance in terms of distance between the operator's trajectory and a desired specification. The latter property is called robustness and will be explained in the subsequent section.

2.2 Measuring Skill

Robustness: We measure robustness of a task performance as a numerical distance between the actual performance and the desired task specification. Consider four trajectories $T_1 - T_4$ in Figure 1. The overall task specification is to avoid region R_1 and reach R_2. We note that trajectories T_1 and T_3 both achieve this task. However, T_1 is seen to be more "robust" in achieving the specification than T_3 since it avoids R_1 with a larger margin. At the same time T_2 and T_4 violate the specification. However, a small perturbation of T_4 could have potentially caused it to reach the region R_2 and satisfy the specification. As a result, T_4 is a "less severe" violation than T_2.

The task in Figure 1 is specified in Signal Temporal Logic (STL):

$$\underbrace{\textsc{Always}(\neg R_1)}_{\text{avoid } R_1} \wedge \underbrace{\textsc{Eventually}(R_2)}_{\text{Reach } R_2}$$

The previous work of Fainekos et al [9] and Donze et al [8] allow us to systematically compute robustness values with respect to the STL specification. This robustness has a positive value for trajectories that satisfy this property and negative values for violating the trajectory. Robustness measures the diameter of the smallest "tube" around a trajectory such that all trajectories that stay inside this tube have the same outcome (satisfaction or violation) as the original trajectory. Thus, T_1's robustness will yield a large positive value, whereas T_3's robustness will be positive but smaller. Likewise, T_2's robustness will be a negative value with a large magnitude whereas T_4's robustness is also negative but with a small magnitude.

Efficiency: Efficiency is concerned with minimizing use of resources. We can consider resources such as: (a) *Time efficiency:* how much time is taken by the operator to complete the task? (b) *Resource efficiency:* How much energy is expended by the operator? Often, control designers express the resource efficiency as a function over the states of the trajectory and the control action of the human operator. While this is domain specific, it is easy to express and evaluate systematically. (c) *Control Variation*: is the applied control jerky or smooth? This can be measured by computing the total variation distance over the operator inputs. Depending on the domain, there may be many types of efficiency, each contributing to a different skill dimension.

Resilience: Resilience pertains to the correct execution of a task under varying environmental conditions. In a teleoperation setting these apply to the environment surrounding the remote system. Thus, for a teleoperated drone environmental conditions manifest in many ways including wind, sensor malfunctions, and damage to the drone. To measure resilience, we propose repeated execution of a task under unanticipated off-nominal conditions and measuring how robustness of the resulting trajectories vary with changing environmental conditions.

Readiness: Readiness depends on the context surrounding the operator themselves. We posit that skilled operators exhibit readiness against changing contexts that may include their physical comfort, or biological state such as time since last meal or sleepiness. This aspect of skill is the hardest to measure systematically since it is often undesirable to subject humans knowingly to adverse physical conditions.

3 EVALUATION

We conducted an initial evaluation of the proposed skill assessment approach using data collected from a convenience sample of five individuals. Using a drone piloting simulation implemented in Unity (see Figure 2) and an Xbox controller, the target task was to take off vertically to reach the floating target, hover within the floating target for five consecutive seconds, and land vertically to reach the landing pad. Note that if the drone strays from the target area during the hover segment, the timer is reset. Each person recorded two attempts of the specified task, yielding 10 different trajectories.

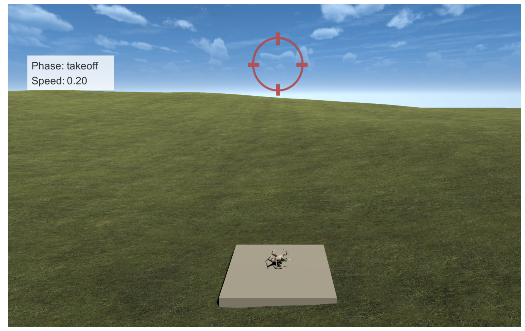


Figure 2: Unity-based Drone piloting simulator.

For each trajectory, we evaluated skill along the Robustness and Efficiency dimensions. We discuss possible elicitation of Resiliency and Readiness in the next section.

Robustness: The specifications of takeoff and hover are as specified in Section 2.1. The specification for landing is similar to takeoff but specifies a negative v_y until the drone reaches a minimum altitude with an appropriately small vertical velocity. We implemented a robustness computation engine as a simplified version of the tool TaLiRo [11]. We calculated separate robustness scores for each of the three task segments.

Efficiency: We measured time efficiency as the time required to complete each segment of the task, where longer times are considered less efficient. For the takeoff and landing segments, the minimum time (and thus most time efficient) to complete the task is essentially 0 seconds. For the hover task, time efficiency is the time required to complete the hover outside of the minimum 5 seconds (if the drone never leaves the target area).

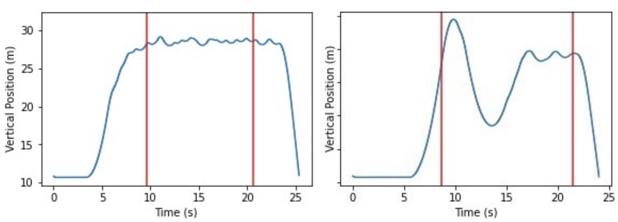
We measured control variation using mean variation distance. For each time t, we measured the distance between the control input vector $(roll, pitch, yaw, throttle)$ and the control input vector at time $t+1$, scaling by the size of the time step. Smoother control actions yield less change and smaller distances between the control input vectors. To remove redundancy with the time efficiency measure, we computed the mean control variation over the given segment. Efficiency is coded as a negative number, so larger (negative) values represent lower efficiency; this is done to align with intuition that higher numbers are better scores.

4 RESULTS AND DISCUSSION

We present the robustness and efficiency measures for each recorded segment in Table 2. One immediate finding is that all robustness estimates for the takeoff and landing segments are negative, which means that none of the recorded trajectories met the desired specifications. In particular, users flew the drone too fast at the end of the takeoff segment, overshooting the desired ending location. For the landing segment, participants often missed the task change and hovered too long before landing using a high speed. All robustness

Table 2: Robustness and efficiency measures for each recorded trajectory. The best values in each column are bolded and the worst values are italicized. RO = Robustness, TE = Time Efficiency, and CE = Control Efficiency

		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



(a) Participant 5, Trial 1 **(b) Participant 3, Trial 2**

Figure 3: Example user trajectories. The vertical position (in meters) is plotted over time (in seconds). The red vertical lines indicate the transition between task segments.

estimates for the hover segments were positive by design, since the simulator would not trigger the next segment until the drone successfully hovered at the target for the desired length of time (5 seconds).

The control efficiency values are very small. This is likely due to the fact that the simulator task only required use of one control input (throttle) and thus the overall control input needed was very small. However, we do see relative differences in control efficiency between the recorded segments, with the hover task showing overall less control efficiency than the takeoff or landing segments.

We arrive at more interesting findings by investigating the relationship between the different skill measures. First, notice in Table 2 that Participant 5 in Trial 1 achieved the best robustness and the worst control efficiency in the takeoff and hover segments (see Figure 3a). This could be caused by the user "feathering" the controls in order to correct their trajectory. On the other hand, Participant 3 in Trial 2 achieved the best robustness, time efficiency, and control efficiency in the landing segment (see Figure 3b). This shows that these dimensions of skill are not necessarily correlated, and may reflect individual user operating styles. Note that we cannot draw any generalized examples here due to our small and non-representative sample of operators; we merely attempt to highlight the nuance afforded by measuring skill along multiple dimensions.

Defining skill along multiple dimensions provides system designers with important decisions as they develop products and interventions. For a specific domain, which aspect of skill is most important? Perhaps the given task is so safety-critical that robustness is the only dimension that matters. Other tasks may depend more on resource efficiency. While we propose several possible dimensions of skill, dimension reduction techniques applied to a larger volume of data may indicate a smaller set of latent dimensions that measure skill.

A limitation of this preliminary analysis is that our data did not allow us to measure the skill dimensions of Resiliency and Readiness. Future work can specifically test these dimensions in a controlled user study by varying conditions in the simulator (e.g., wind, time of day) or the user's conditions (e.g., before/after meals, using distractor tasks). Future work can also extend the proposed framework to describe more complicated tasks that require concurrent primitive skills. Following additional user studies, we can also calculate the distribution of these skill values across a larger population as well as plotting learning curves to see how users improve in various skill dimensions over repeated trials. We plan to also incorporate user confidence ratings of their performance and structured interviews to investigate how users experience skill development. For example, we may see a change from effortful to automatic control such as in [13] or other more qualitative stages of development as users acclimate the control into their own body perception [1, 10].

5 CONCLUSIONS

This paper presents a skill measurement framework that argues for a careful identification of the various primitive tasks, their formal specification using temporal logic and a multi-dimensional approach Our preliminary evaluation on a small, non-representative set of trajectories shows some of the benefits of this approach. We have also presented a vision of how the proposed approach may lead to a comprehensive skill evaluation framework in the future.

Acknowledgments: We thank the anonymous reviewers for their detailed feedback. This work was funded in part by the US National Science Foundation (NSF) under award number 1836900 and Oak Ridge Associated Universities (ORAU) under Contract number W911NF2220077.

REFERENCES

[1] Chris Baber, Tulin Cengiz, and Manish Parekh. 2014. Tool use as distributed cognition: how tools help, hinder and define manual skill. *Frontiers in Psychology*

5 (2014). https://doi.org/10.3389/fpsyg.2014.00116
[2] Christel Baier and Joost-Pieter Katoen. 2008. *Principles of Model Checking*.
[3] Ezio Bartocci, Jyotirmoy Deshmukh, Alexandre Donzé, Georgios Fainekos, Oded Maler, Dejan Ničković, and Sriram Sankaranarayanan. 2018. Specification-Based Monitoring of Cyber-Physical Systems: A Survey on Theory, Tools and Applications. In *Lectures on Runtime Verification: Introductory and Advanced Topics*. Springer, 135–175.
[4] Amit Bhatia, Lydia E. Kavraki, and Moshe Y. Vardi. 2010. Sampling-based motion planning with temporal goals. In *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2689–2696. https://doi.org/10.1109/ROBOT.2010.5509503
[5] David Buckley, Ke Chen, and Joshua Knowles. 2017. Rapid Skill Capture in a First-Person Shooter. *IEEE Transactions on Computational Intelligence and AI in Games* 9, 1 (2017), 63–75. https://doi.org/10.1109/TCIAIG.2015.2494849
[6] 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
[7] Alexandre Donzé. 2013. On Signal Temporal Logic. In *Runtime Verification*. Springer, 382–383.
[8] Alexandre Donzé and Oded Maler. 2010. Robust Satisfaction of Temporal Logic over Real-Valued Signals. In *Formal Modeling and Analysis of Timed Systems*. Springer, Berlin, Germany, 92–106.
[9] Georgios E. Fainekos and George J. Pappas. 2009. Robustness of temporal logic specifications for continuous-time signals. *Theoret. Comput. Sci.* 410, 42 (Sept. 2009), 4262–4291. https://doi.org/10.1016/j.tcs.2009.06.021
[10] Kurt Fischer and Yan Zheng. 2002. *Conceptions of Development*. Psychology Press, Chapter The Development of Dynamic Skill Theory.
[11] Georgios Fainekos et al. . 2022. TALIRO-TOOLS. https://sites.google.com/a/asu.edu/s-taliro [Online; accessed 7. Dec. 2022].
[12] Lars Johannsmeier, Malkin Gerchow, and Sami Haddadin. 2019. A Framework for Robot Manipulation: Skill Formalism, Meta Learning and Adaptive Control. In *2019 International Conference on Robotics and Automation (ICRA)*. 5844–5850. https://doi.org/10.1109/ICRA.2019.8793542
[13] Daniel Kahneman. 2011. *Thinking, fast and slow*. Farrar, Staus and Giroux, New York, NY, USA.
[14] Joseph Kim, Christian Muise, Ankit Shah, Shubham Agarwal, and Julie Shah. 2019. Bayesian Inference of Linear Temporal Logic Specifications for Contrastive Explanations. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 5591–5598. https://doi.org/10.24963/ijcai.2019/776
[15] Ron Koymans. 1990. Specifying real-time properties with metric temporal logic. *Real-Time Syst.* 2, 4 (Nov. 1990), 255–299.
[16] Hadas Kress-Gazit, Georgios E. Fainekos, and George J. Pappas. 2009. Temporal-Logic-Based Reactive Mission and Motion Planning. *IEEE Trans. Rob.* 25, 6 (Sept. 2009), 1370–1381. https://doi.org/10.1109/TRO.2009.2030225
[17] Zohar Manna and Amir Pnueli. 1992. *The Temporal Logic of Reactive and Concurrent Systems*. Springer, New York, NY, USA. https://link.springer.com/book/10.1007/978-1-4612-0931-7
[18] Claudio Menghi, Christos Tsigkanos, Thorsten Berger, Patrizio Pelliccione, and Carlo Ghezzi. 2018. Property specification patterns for robotic missions. In *ICSE '18: Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings*. Association for Computing Machinery, New York, NY, USA, 434–435. https://doi.org/10.1145/3183440.3195044
[19] H. Mosemann and F.M. Wahl. 2001. Automatic decomposition of planned assembly sequences into skill primitives. *IEEE Transactions on Robotics and Automation* 17, 5 (2001), 709–718. https://doi.org/10.1109/70.964670
[20] Delwyn Nicholls, Linda Sweet, and Jon Hyett. 2014. Psychomotor Skills in Medical Ultrasound Imaging. *Journal of Ultrasound in Medicine* 33, 8 (2014), 1349–1352. https://doi.org/10.7863/ultra.33.8.1349 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.7863/ultra.33.8.1349
[21] Stefan Profanter, Ari Breitkreuz, Markus Rickert, and Alois Knoll. 2019. A Hardware-Agnostic OPC UA Skill Model for Robot Manipulators and Tools. In *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. 1061–1068. https://doi.org/10.1109/ETFA.2019.8869205
[22] Dorsa Sadigh and Ashish Kapoor. 2016. Safe Control under Uncertainty with Probabilistic Signal Temporal Logic. In *Proceedings of Robotics: Science and Systems*. AnnArbor, Michigan.
[23] Lior Shmuelof, John W. Krakauer, and Pietro Mazzoni. 2012. How is a motor skill learned? change and invariance at the levels of task success and trajectory control. *Journal of Neurophysiology* 108, 2 (2012), 578–594. https://doi.org/10.1152/jn.00856.2011
[24] Franz Steinmetz, Annika Wollschläger, and Roman Weitschat. 2018. RAZER—A HRI for Visual Task-Level Programming and Intuitive Skill Parameterization. *IEEE Robotics and Automation Letters* 3, 3 (2018), 1362–1369. https://doi.org/10.1109/LRA.2018.2798300
[25] U. Thomas, B. Finkemeyer, T. Kroger, and F.M. Wahl. 2003. Error-tolerant execution of complex robot tasks based on skill primitives. In *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, Vol. 3. 3069–3075 vol.3. https://doi.org/10.1109/ROBOT.2003.1242062