Robots Teaching Humans: A New Communication Paradigm via Reverse Teleoperation

Anonymous Author(s)

Abstract: Simulators offer a scalable platform to train robots, offering a path to creative and innovative solutions that are difficult for humans to envision a priori. We introduce a way to leverage this property, along with a new paradigm where robots discover creative solutions in simulation, then teach humans or other agents to physically execute the learned solutions via reverse teleoperation. We provide various examples ranging from learning new skills, to rehabilitation, to everyday activities, where such a system would be valuable.

Keywords: Teleoperation, Reinforcement Learning in Simulation

1 Motivation: Communicating New Kinesthetic Insights

The 37th move of the 2nd game between the computer program AlphaGo [1] and the 18-time world champion Lee Sedol caught the human champion by surprise. The move was puzzling to most observers as well – it did not follow any known strategies. What seemed a mistake at first turned the course of the game. ‘Move 37’ was eventually recognized as AlphaGo’s own innovation, since it could not be attributed to memorizing human strategies used during training. The successors AlphaZero [2] and MuZero [3] learned from self-play without human guidance. Go enthusiasts now recognize the potential of these programs to generate unique insights for new game strategies.

 Robots – the embodied programs – could generate valuable innovative insights for solving physical tasks. They could safely attempt a myriad tries in simulation, and the resulting strategies would be applicable to the real world if the simulation-to-reality mismatch is not large. They could also learn from real-world interactions, and instantly share progress with other robots that have a similar embodiment. However, robots lack an effective way to communicate their innovations to humans. It is unlikely that people pick up the new strategies by watching robot motions, just as the approach of learning to play the piano by watching in unlikely to succeed. What could be an effective way for the robots to teach their innovations to us? Our blue-sky idea is to develop a new paradigm that would allow robots to communicate to humans their kinesthetic insights directly via reverse teleoperation.

2 Robot-to-Human Skill Transfer: Potential Benefits

With rapid advances in learning and simulation, running large-scale distributed optimization should soon allow robots to find novel ways of solving various tasks. They are likely to improve beyond human-level performance even when constrained by similar physical and hardware characteristics. Instead of using this only for perfecting robotic performance, we propose to think about how this could help humans. For now, we will assume that we either train anthropomorphic agents in simulation, or train humanoid robots in reality and can solve the mapping to the human morphology. Below we summarize the benefits of transferring kinesthetic knowledge from robots to humans.
**Increasing motivation and reducing time to mastery**: Beginners that fail to make progress can become demotivated and disengaged. Reverse teleoperation could help execute a new skill correctly right away. A robot can then slowly attenuate the support to reduce human’s reliance on hardware, balancing the challenge vs success to maintain a high level of motivation even in unskilled beginners. Moreover, reverse teleoperation can ensure high quality for the full duration of the practice, reducing the amount of time wasted on repeating the moves incorrectly, i.e. avoid forming ‘bad habits’.

**Improving ergonomics**: Repetitive tasks are not going to disappear even if robots replace humans in factory work, because we need to do office work and household tasks that are difficult to fully automate. We can encode ergonomics guidelines into the reward signal directly when training simulated/robotic agents. Teaching humans to perform repetitive tasks in ergonomically optimal ways could ensure better well-being and prevent repetitive strain injuries. Even when people know about the correct approach they frequently lack the habit/discipline to follow it. Reverse teleoperation would be an easy way to help them acquire correct habits e.g. helping to hold the correct arm posture for typing, help keeping one’s back straight during lifting, etc.

**New ways of solving old tasks**: High jump has been a popular sport since the 19th century, but the optimality of the ‘Fosbury flop’ to go over the bar backwards has not been evident to all, until Dick Fosbury won the 1968 Olympics [4]. It could be that the skills and tool use strategies we have for many of our everyday and professional tasks are suboptimal. Ability to try daring exploration strategies during large-scale training in simulation could let us overcome these local optima.

**Designing new tools and teaching humans how to use them**: Distributed large-scale training in simulation could help create new tools and optimize their shapes. Shape optimization using physics (aerodynamics) simulation has been employed in automotive and aerospace industries [5, 6], and more recently in robotics [7, 8]. We could search for new tools, optimize tool shapes, then quickly teach humans to use these tools. We can learn basic use of a novel tool first (e.g. with [9]), then attempt risky exploration strategies when optimizing in simulation or with a simplified robust robot hardware. This would ease innovation, even for risky tasks, without injuries (e.g. test new designs of a nail holder for hammering). Overall, this could improve efficiency and safety of the tools.

**Surgeries**: Just as DaVinci robot [10] helped pioneer teleoperation for complex surgical operations, we could imagine future robots training surgeons to do delicate surgeries with known or new strategies. General surgical training takes at least five years to complete after finishing medical school. Speeding up the process of correct muscle memory formation for the surgeons would yield quicker training. Making reverse teleoperation systems accessible to people throughout the world could be crucial to ensuring enough highly skilled surgeons, in both developed and developing countries.

**Rehabilitation**: Patients with various functional limb actuation or grasp pathologies can benefit from having a glove [11, 12] or full-body exoskeleton to enable learning to move and interact in the real world with various strategies that are tailored to their situation. Importantly, the robot could also help rehabilitation patients follow the correct progression of rehabilitation exercises, and facilitate faster recovery by providing a personalized schedules based on their current recovery progress.

**Arts, Sports, and Fun**: Lastly, we envision a wide range of further scenarios in our daily lives where reverse teleoperation can help to speed up muscle memory formation, e.g. playing piano [12], using chopsticks, touch typing, mastering novel painting techniques, forming correct habits for yoga poses and other physical exercises, tricky fun skills, such as juggling, balancing for slacklining.

### 3 Reverse Teleoperation: Hardware and Algorithms

To enable the robot-to-human transfer we need a safe and effective reverse teleoperation system. Figure 1 outlines the necessary components. We start by setting up a task in simulation, specifying a general RL objective/reward function and the constraints that the learned policy must satisfy to be safe. We propose to keep objectives relatively high level, e.g. sparse rewards as opposed to dense, not too specific (no reward tweaking or demonstrations). This should allow creative solutions to emerge and avoid biasing RL to only seek solutions that humans can envision. It has been demonstrated that interesting behaviors emerge with sufficiently diversified large-scale training [13, 14]. These solutions may range from exploring novel shapes/designs (e.g. new tools) to finding new trajectories that solve the task, covering the a range of aspects to explore in simulation about the environment structure and the task. With recent progress in GPU-accelerated simulation, it is possible to train more than 10K environments in parallel on one GPU, even for advanced cases like ShadowHand[15].
Parallelizing across GPUs would enable scaling to 1 million parallel environments. We will aim to train a ‘main’ policy for an average human profile/size, then fine-tune it to accommodate humans of various ages, customizing for different levels of flexibility and other physical characteristics.

The next stage is a mechanism to actuate a person to follow trajectories proposed by the robot. The learned policies can be downloaded and run on the hardware that is worn by the user. This hardware would be composed of a sensing module to keep track of the human pose and muscle activities, and an actuation module that safely executes the next actions to move human limbs. When users resist the suggested movements, the controller would re-plan and adapt to ensure comfort and safety. Highly sensitive force-torque sensors that can detect such resistance reliably, along with impedance control, would be used to enable this [16]. If user data is shared via a central repository, reverse teleoperation systems could improve user preference models based on feedback from similar users.

Exoskeletons:

Exopulse Mollii Suit [17] is a full-body suite for neuromodulation. It includes 58 embedded electrodes positioned to stimulate 40 muscles in various parts of the body through low frequency electrostimulation. Mollii Suit is used to relax muscle spasms, activate muscles to increase blood circulation and prevent atrophy [18]. It can also alleviate tremors from Parkinson’s disease. We envision extending a suite like this to activate fine muscles for reverse teleoperation. Experimenting with various electrode types that could support more targeted muscle activation could allow to extend this suite beyond its current applications. The fact that the suite is a medical device offers a safe starting point. A safe full-body exoskeleton has been recently introduced by Sarcos (rightmost in Figure 1). It provides power-assist for users who need to lift heavy objects, apply large torques to turn industrial valves, etc. We envision that this full-body exoskeleton could be combined with work on rehabilitation exoskeletons for limbs (e.g. [19]) to create a powerful yet agile system for human actuation. Work on ‘teleoperating’ users’ own ankle prosthesis (middle in Figure 1) provides insights for how to sense user’s fine motions, and could give insights on how users react to a part connected to their body being actuated [20].

Sim and RL:

The recent success of RL has spurred massive interest in simulators to train robots. With ever improving simulators, today rendering high quality images is not a challenge anymore and physics simulators run on GPU [15] gathering 100-1000x more experience compared to CPUs. Moreover, it is becoming increasingly possible to do a high-fidelity simulation of muscles and tendons of human body [21]. Obviously, as the scene complexity increases, so does the simulation time but we believe more compute will continue to be available in future and we also anticipate that need for fast simulation will drive the progress in compute going forward. Also, we expect a similar trend in improvement in RL algorithms and their data efficiency. Since simulation is a critical component of our work, we assume that high-fidelity simulation is or will be possible in the near-term.
4 Relevant Background on Teleoperation

Reverse teleoperation is largely inspired by the standard teleoperation practice used in robotics with an aim to close the loop and recognize that teaching need not be unidirectional i.e. a robot with powerful learning algorithm can teach a human new skills by physically moving the limbs. Below we review various modern and historical teleoperation systems, and outline the concepts they share with our idea of reverse teleoperation.

Teleoperation has primarily been used as a tool in shared autonomy with a goal of augmenting the capabilities of a robotic system. A human is always in the loop to ensure that a task is completed reliably, providing ingenuity, or correcting unwanted behaviours by taking over the control. Today, commercially available robots can cope with critical tasks, such as telesurgery. Although extremely expensive, it has been estimated that da Vinci surgical systems have completed more than 7 million surgical procedures as of now [10]. DLR’s Justin robot has been used for teleoperation to rehearse scenarios where robots on a different planet are being teleoperated by humans with what is called as “supervised autonomy” — a concept somewhere between full autonomy and direct teleoperation.

An astronaut on board the International Space Station has remotely operated a humanoid robot to inspect and repair a solar farm on a simulated Mars environment set up at DLR in Munich [22]. HaptX [23] show that their teleoperation system can be used to perform extremely fine manipulations on a shadow hand by relaying human finger movements over to the robot hand. Their system also provides haptic feedback and therefore the teleoperator can get a sense of force needed to manipulate objects. Recently, [24] introduced DexPilot, a very low-cost teleoperation system that is also able to perform fine manipulations on an allegro hand. The system operates directly with RGB cameras and does not use any haptic feedback as of now but it can benefit from it just like HaptX. Telexistence [25] is another company that designs robots to do physical labour in the supermarkets. Huang et al.[26] showed a music instruction system that has finger-less gloves and vibrators on each finger which activate based on which finger is used to play a musical note. Such passive haptic gloves can teach the user to build “muscle memory” to play piano or passive haptic rehabilitation: helping people with partial spinal cord injury improve sensation and dexterity in their affected hands.

Teleoperation has also been a topic of interest in popular science fiction stories e.g. the 1942 short story “Waldo” by Robert Heinlein [27] features a man who invents and teleoperates a device using his hand and fingers. Below is a short extract from the book:

Waldo put his arms into the primary pair before him; all three pairs, including the secondary pair before the machine, came to life. Waldo flexed and extended his fingers gently; the two pairs of waldoes in the screen followed in exact, simultaneous parallelism. Such has been the popularity of this story that many real-life remote manipulators developed later also came to be called waldoes. See [28] for a history of various teleoperators, exoskeletons and industrial robots.

Conclusion:

We hope that the our paradigm of reverse teleoperation will offer a way to teach users “muscle memory” for various day to day tasks, help design new tools and teaching humans new ways to use them and have impacts in medical surgery and rehabilitation. We also imagine deployment of the system at scale would offer a chance to continually improve the system as data from the users is uploaded regularly in an ever growing central repository. Any user can therefore learn from users trialling somewhere across the world in what we think is a direct consumer-to-consumer model. Furthermore, as more users deploy this system for various innovative use cases the repository also benefits from an ever increasing category of skills leading to reduced adaptation time and faster learning for new users. Ultimately, this offers a new digital layer to humans to improve learning new skills where algorithmic creativity in simulation together with kinesthetic movements in the real world provide a fundamentally new way to reverse the teaching process.
References


