3rd Workshop Toward Robot Avatars
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Abu Dhabi, UAE, 2024.

Dexterous Manipulation with a Bi-Manual Anthropomorphic Teleoperation Robot

Toby Buckley, J. Edward Colgate

***Abstract*— We present a new upper-limb anthropomorphic dexterous telemanipulation system, the Dexterity Testbed Nexus (DexNex). DexNex is teleoperated by a human user in the Operator Station who controls the Avatar Station to complete manipulation tasks. The Avatar replicates the upper limbs of a human and is statically mounted to the workspace. Three benchmarking tasks were used: box & blocks, the Minnesota Turning Test revised form (MTTrf), and a table setting task. Subjects completed the tasks with their natural bodies to provide normative data. Subjects then attempted the same tasks with haptic feedback enabled or disabled. The utility of haptics was computed for four metrics. Haptic feedback improved performance for three of the four metrics (26% increase in Box & Blocks score, 12% increased Table Setting success rate, and 1.3x faster time per success in Table Setting).**

Figure 1. Dexterity Testbed Nexus. Left: Operator station, right: Avatar station. Only the upper limbs are tracked and mirrored. Feedback is provided to the user visually and haptically.

I. INTRODUCTION

This paper presents a new dexterous manipulation testbed, the Dexterity Nexus (DexNex), with anthropomorphic arms, hands, and vision (Fig. 1). DexNex's objective is to test advanced hardware and software to improve the performance of upper-limb manipulation systems. DexNex is composed of an Operator station and an Avatar station. This paper uses the DexNex base system which mirrors the Operator's actions 1-to-1 onto the Avatar. Inspiration for DexNex's design was taken from the NimbRo Avatar [1] and the Tactile Telerobot [2].

Similar teleoperation systems have benchmarked their performance compared to normative data. I. A. Kuling et al. [3] compared the utility of haptics and found that haptic feedback was preferred but didn't have an impact on performance in a Box & Blocks task. Their system reported a **13.3x** lower teleoperation score compared to their natural bodies. Fishel et al. [2] also performed the Box & Blocks task with shadow robot hands and UR10 arms. They reported a **4.6x** lower teleoperation score.

II. SYSTEM OVERVIEW

The Operator station uses two HaptX DK2 gloves (Seattle, WA), three SteamVR base stations (Bellevue, WA), and a Varjo Aero VR headset (Arlington, VA). The headset provides stereoscopic visual feedback while the gloves provide fingertip haptic feedback and passive finger force feedback.

The Avatar station uses two ABB Gofas for arms, two Shadow Dexterous Hands, a UFACTORY xArm6 for its neck, two FLIR Blackfly cameras with Fujifilm fisheye lenses for eyes, and 6 Biotac SP- fingertip pressure sensors on the thumb, index, and middle fingers. The Biotac sensors each provide a single analog output of pressure.

The Operator and Avatar stations are computationally separate; the only connection is a single ethernet cord which puts each station onto the same local area network (LAN). ROS2 is the middleware used to allow different applications and hardware to communicate. GPU's on Avatar and Operator PC's use FFMPEG with the H.265 codec to encode & decode 4k camera feeds at 60 FPS with about **16ms** latency.

Double Spherical Televisualization (DST) is used to render the wide field-of-view (FOV) Avatar cameras to the Operator's head-mounted-display (HMD) [4]. The Unity 3D graphics engine is used to facilitate DST and provide 3D vision to the user.

The ROS2 package Moveit2 is used to compute inverse kinematics (IK) for each arm. The neck arm uses the Kinematics and Dynamics Library (KDL) IK solver to mirror the Operator HMD pose 1-to-1. The Avatar arm+wrist (8 DoF) tracks the user's palm using a custom cost function in a gradient descent-based solver (BIO-IK) with a term for L2 normalization (limits total movement in joint space) [5].

Each finger uses a uniquely tuned custom cost function in BIO-IK with terms for L2 normalization, target orientation, target position, and planar position. Each finger IK was calibrated to facilitate gestures when the user adopts an open hand state, and fine manipulation when the user adopts a closed hand state. A simple linear scaling function is used to transition between "gesture" mode and "manipulation" mode.

Biotac fingertip pressures are normalized and then passed onto the HaptX fingertip tactors & finger brakes. All tactors of a fingertip are inflated according to a simple linear scaling function. Finger brakes activate once a static threshold is surpassed.

III. TEST PROCEDURE

Seven subjects performed three tasks: Box & Blocks, Minnesota Turning Test revised form (MTTrf), and Table Setting [6, 7].

Each subject performed each task three times: once with their natural bodies, once teleoperated with haptics on, and once with haptics off. The order of teleoperation was switched for each participant to minimize the effects of training time on the calculation of haptics utility. Sessions were stopped if hardware or software failures occurred. After each session, a post-test survey was conducted for qualitative data.

Subjects were first trained in DexNex for 5 minutes doing various non-task warmup exercises. Subjects then had 5 minutes to practice one trial of the MTTrf task (the same procedure as the normative data).

IV. Results

Four metrics were used to compare results: Box & Blocks score, MTTrf success rate, Table Setting (T.S.) success rate, and Table Setting time per success (Fig. 2).

The data for no haptic feedback vs. haptic feedback was compared to inform the utility of haptics. For Box & Blocks haptic feedback increased the score by **26%** (3.9 to 4.9, n=7). For MTTrf haptic feedback decreased the success rate by **20%** (4.3 to 3.4, n=7).

For the Table Setting task, haptics improved the success rate **12%** (5.0 to 5.6, n=5), The Table Setting time per success was about **1.3x** faster (75.4s to 56.3s, n=5) with haptic feedback.

Data were also compared to those for a highly experienced teleoperator (lead author) with around 30 hours of training. Compared to the average teleoperator, the experienced user scored **40%** better on Box & Blocks, had a **2x** higher MTTrf success rate, had a **1.1x** higher Table Setting success rate, and was **2.9x** faster per success on Table Setting.

Compared to the natural body performance, the average participant scored **15x** worse on Box & Blocks, had **50%** worse success rate on MTTrf, had **10%** worse success rate on Table Setting, and took **40x** longer per success on Table Setting. Compared to the natural body performance, the highly experienced teleoperator scored **10x** worse on Box & Blocks, had the same MTTrf success rate (**100%**), had the same Table Setting success rate (**100%**), and took **13.5x** longer per success on Table Setting.

V. Discussion

The small sample size limits the strength of any conclusions. For example, in the MTTrf task, one would expect success rate to increase with haptic feedback, but in fact it decreased. This trend may well reverse with additional data.

The MTTrf task required users to delicately grasp lightweight pucks and rotate the Avatar forearm 180°. Neither of these actions were heavily dependent on haptic feedback. Regardless, users did report a preference for haptics enabled as it reduced mental burden since they relied less on tiresome visual processing to determine if objects were grasped or not.

Based on the user surveys and observations, the biggest bottlenecks to performance were difficulty in achieving desired finger positions, difficulty completing grasps and obtaining high quality grasps, difficulty avoiding collisions due to bulky Avatar components, and physical delays from moving or rotating bulky equipment. In addition, users desired more training time which would let them become better acquainted with how the system functioned. Users also reported feeling physically fatigued after the trials. Some experienced mental fatigue too, although this was reduced when haptic feedback was provided.

Based on this study, future work will provide assistive features such as automatically aligning grasps and stabilizing object interactions. Such features will enable higher performance teleoperation while reducing mental burden, frustration, and fatigue of the user. Hardware could also be improved; lighter robot hands & forearms would lead to much quicker movements, especially wrist rotations. More advanced features that incorporate computer vision and AI/ML may help to speed up teleoperation further. For instance, an AI copilot could transform Operator actions into a more useful command signal, increasing the likelihood of task success, and a dynamic world model would allow movement planning to achieve grasps and avoid collisions. Lastly, the system must also be evaluated on more tasks to explore its generalizability.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 2221571. The authors thank Jeremy Fishel, Rohan Kota, Kevin Nella, Matt Elwin, and Carl Moore for fruitful discussions.

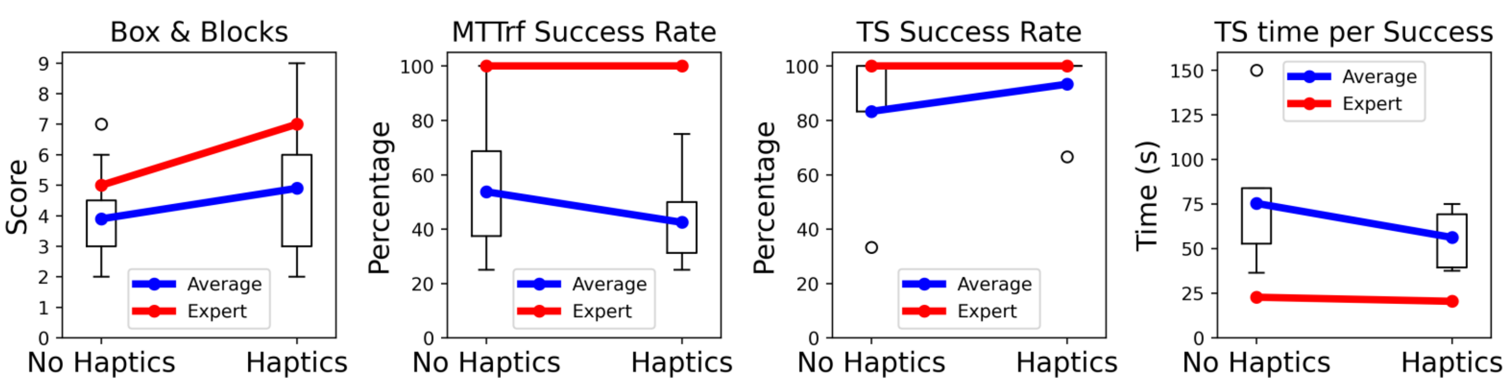

Figure 2. Results of all participants' teleoperation for all four metrics. Each chart compares the performance without and with haptic feedback enabled. From left to right: Box & Blocks score (normative score: **63.8**), MTTrf Success Rate (normative score: **100%**), Table Setting Success Rate (normative score: **100%**), Table Setting time per success (normative score: **1.6s**)

REFERENCES

[1] Schwarz, Max, Christian Lenz, Andre Rochow, Michael Schreiber, and Sven Behnke. "Nimbro avatar: Interactive immersive telepresence with force-feedback telemanipulation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5312-5319. IEEE, 2021.

[2] Jeremy A. Fishel, Toni Oliver, Michael Eichermueller, Giuseppe Barbieri, Ethan Fowler, Toivo Hartikainen, Luke Moss, and Rich Walker. “Tactile telerobots for dull, dirty, dangerous, and inaccessible tasks,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11305–11310, 2020.

[3] Irene A. Kuling, Kaj Gijsbertse, Bouke N. Krom, Kees J. van Teeffelen, and Jan B. F. van Erp. “Haptic feedback in a teleoperated box & blocks task,” in Ilana Nisky, Jess Hartcher-O’Brien, Micha ̈el Wiertlewski, and Jeroen Smeets, editors, *Haptics: Science, Technology, Applications*, pages 96–104, Cham, 2020. Springer International Publishing

[4] Max Schwarz and Sven Behnke. “Low-latency immersive 6d televisualization with spherical rendering,” in 2020 *IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, pages 320–325. IEEE, 2021.

[5] Starke, Sebastian. "Bio IK: a memetic evolutionary algorithm for generic multi-objective inverse kinematics," *PhD diss.*, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky, 2020.

[6] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel and A. M. Dollar, "The YCB object and Model set: Towards common benchmarks for manipulation research," *2015 International Conference on Advanced Robotics (ICAR)*, Istanbul, Turkey, 2015, pp. 510-517. IEEE, 2015.

[7] Tesio, L., Simone, A., Zebellin, G., Rota, V., Malfitano, C., & Perucca, L. “Bimanual dexterity assessment: validation of a revised form of the turning subtest from the Minnesota Dexterity Test,” in *International journal of rehabilitation research. Internationale Zeitschrift fur Rehabilitationsforschung. Revue internationale de recherches de readaptation*, *39*(1), 57–62.