Quantifying Haptic Quality: External Measurements Match Expert Assessments of Stiffness Rendering Across Devices

Farimah Fazlollahi[1], Hasti Seifi[2], Giulia Ballardini[1], Zahra Taghizadeh[1], Andrew Schulz[1], Karon E. MacLean[3], and Katherine J. Kuchenbecker[1]

***Abstract*— Grounded force-feedback devices employ diverse mechanical, electrical, and computational designs, making it difficult to assess their quality fairly. Device specifications often have little relation to the opinions experts report during hands-on testing. To close this gap, we invited 16 haptics experts to evaluate four representative devices rendering stiff spheres while we recorded data with Haptify, our measurement-based benchmarking system. Our results indicate that expert opinions of each device's maximum renderable stiffness correlate with a performance metric based on measured vibrations. This approach of pairing expert assessments with external measurements promises a systematic way to characterize the capabilities of haptic devices.**

I. Introduction

A force-feedback device is a mechatronic system that renders haptic sensations by measuring the user's motion and/or force, and outputting forces and/or motions in response. Device documentation (e.g., research articles, datasheets) includes a subset of specifications the inventors provided for evaluation [1], [2]. Although Haptipedia visualizes these specifications for more than one hundred haptic devices [3], Seifi et al. showed that expert opinions about a device's capabilities go beyond low-level specifications and instead center on their experience in physically testing it [4]. To map connections between expert opinions and external sensor measurements recorded during the interaction, we conducted a user study in which 16 expert hapticians tested four commercial haptic devices (Novint Falcon, Force Dimension Omega.3, 3D Systems Touch X, and 3D Systems Touch) in two rounds: first unpowered, then actively rendering different virtual environments. We recorded 3D force, position, velocity, and acceleration with our benchmarking system, Haptify [5], along with the haptic parameters, sensed position, and commanded force from CHAI3D [6].

II. User Study and Analysis

After consenting and completing a background questionnaire, experts examined each device unpowered and then while actively rendering five virtual benchmark environments: stiffness, damping, force field, textures, and magnetic marbles. This paper reports on the stiffness benchmark. This

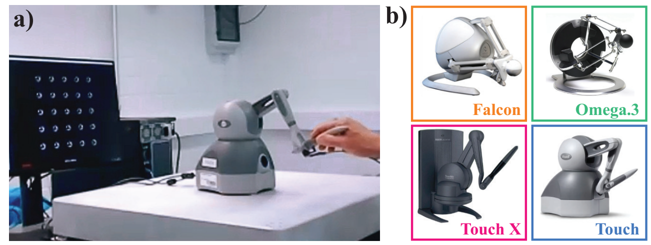

Fig. 1. a) The Touch device being tested. b) The study's four haptic devices.

environment presents 125 stationary frictionless spheres, each rendered with the same stiffness and distributed across the entire workspace. The expert could adjust sphere stiffness using a keyboard, and was asked to identify the maximum stiffness the device could render with acceptable quality. The experts also rated other aspects of the device's performance and answered open-ended questions about the factors they considered in determining the stiffness rendering quality.

1) Thematic Analysis: Expert responses to open-ended questions were imported into MAXQDA (VERBI GmbH) and subjected to thematic analysis [7]. Two authors open-coded the survey responses and then combined similar codes, resulting in a list of factors the experts used to assess stiffness rendering quality. We calculated the percentage of experts who mentioned each factor to identify perceived differences between devices for this benchmark. The most frequently mentioned factor was the *perceived vibration level at different stiffness values*: 81% (for Touch X), 68% (Touch), and 56% (Falcon and Omega.3). Next most-mentioned were workspace limitations (44% Falcon), background forces (37% Falcon), and auditory noise (37% Omega.3).

2) Statistical Analysis: The experts also rated five performance features for each device, including its ability to render stiffness and its rendering uniformity across the workspace (Fig. 2a). As the assumption of normality was not met for these data, we applied the Friedman test to investigate the differences in expert ratings between the devices. Significant effects were analyzed with post-hoc Wilcoxon tests.

3) Quantitative Analysis: We identified a quantitative performance metric from the recorded sensor measurements based on the expert responses to the open-ended questions about the parameters that are important for the stiffness benchmark. Since vibrations were mentioned most often, we calculated the root-mean-square (RMS) of the measured 3D vibration signal recorded on the end-effector, as in [5], over 20-ms time windows. We then calculated the average of the vibrations experienced by all experts at each commanded stiffness value to compare this vibration metric with the

[1]F. Fazlollahi, G. Ballardini, Z. Taghizadeh, A. Schulz, and K. J. Kuchenbecker are with the Haptic Intelligence Dept., Max Planck Institute for Intelligent Systems, Stuttgart, Germany (fazlollahi, ballardini, taghizadeh, aschulz, kjk)@is.mpg.de

[2]H. Seifi is with the School of Computing and Augmented Intelligence, Arizona State University, Tempe, USA hasti.seifi@asu.edu

[3]K. E. MacLean is with the Department of Computer Science, University of British Columbia, Vancouver, Canada, maclean@cs.ubc.ca

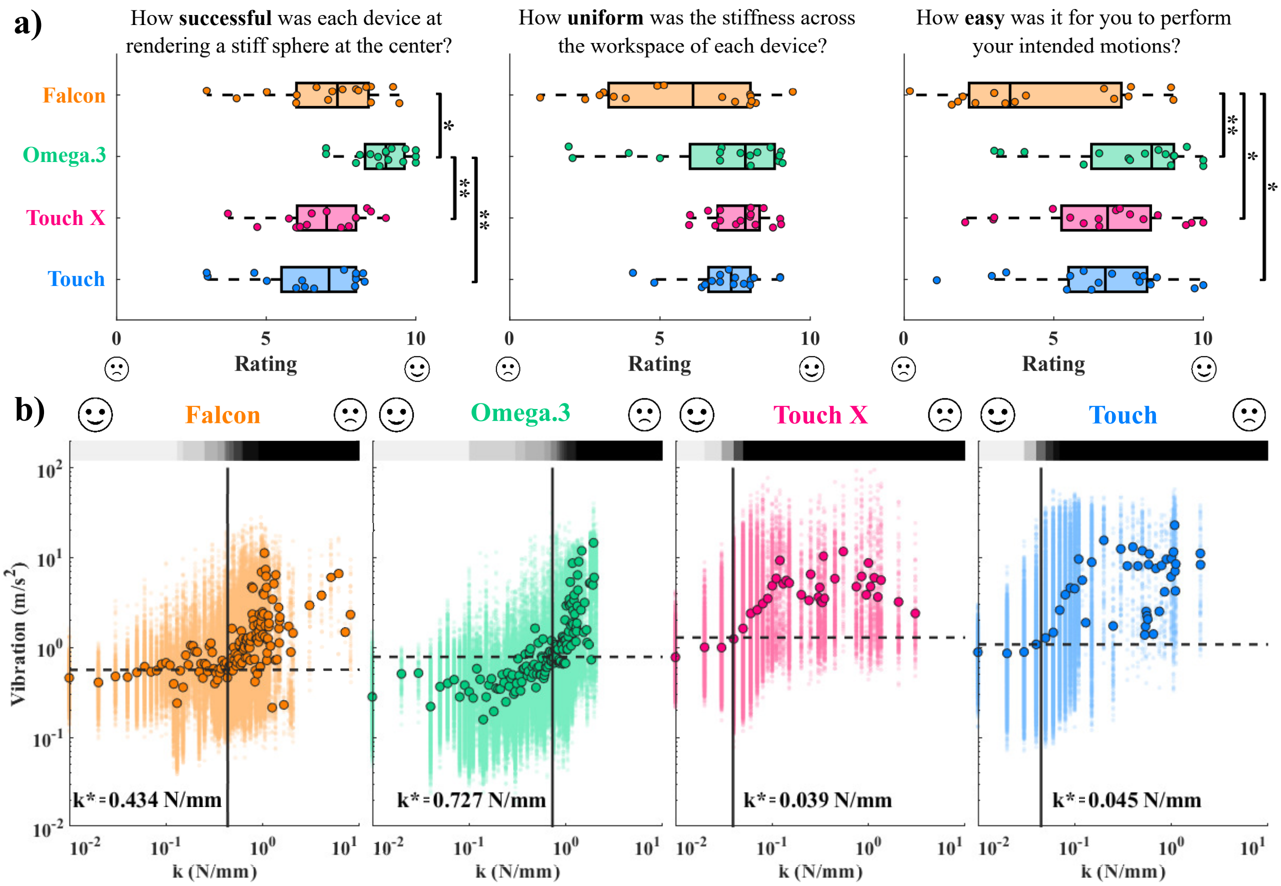

Fig. 2. a) Expert ratings of the four devices on three criteria (from 0 indicating negative to 10 indicating positive assessment); significant differences are marked with * ($p < 0.05$) and ** ($p < 0.01$). b) RMS vibrations measured at each rendered stiffness value. k^* (vertical line) is the mean of the maximum acceptable stiffness values chosen by the experts. The horizontal dashed line indicates the average of the vibration magnitudes near k^*. The white ($n = 16$ experts) to black ($n = 0$) spectrum above each device's vibration response represents the number of experts who found that stiffness to be acceptable.

experts' opinions about stiffness rendering quality (Fig. 2b).

III. Preliminary Results and Future Plans

The proposed vibration metric begins to rise at the average of the expert-chosen maximum stiffness values (Fig. 2b), as hypothesized. The number of experts mentioning unacceptable levels of vibration as a limitation is also consistent with the average RMS vibration at this stiffness value. The two parallel-structured devices, the Omega.3 and the Falcon, can render the highest stiffness values at an acceptable vibration level. The inexpensive Falcon had the lowest vibration level, but its stiffness rendering uniformity and ease of performing intended motions were rated worst (Fig. 2a). The two hybrid serial-structured devices, the Touch X and the Touch, exhibit higher vibration levels at lower stiffness; however, the spheres felt adequately rigid at the lower stiffnesses, so their overall ratings were not different from the Falcon. Next, we plan to calculate other quality metrics based on each device's actual rendered stiffness and maximum force [5]. Then we will use these same methods to analyze the unpowered device interactions and four other active benchmarks.

Acknowledgments

The authors thank Lijuan Wang, Alborz Aghamaleki Sarvestani, Ravali Gourishetti, Bernard Javot, Joey Burns, the MPI-IS Robotics ZWE, and IMPRS-IS.

References

[1] A. Abbasimoshaei, J. Reisinger, C. Neupert, C. Hatzfeld, and W. Zhou, *Evaluation of Haptic Systems*. Cham: Springer International Publishing, 2023, pp. 587–623.

[2] E. Samur, *Performance Metrics for Haptic Interfaces*. Springer Science & Business Media, 2012.

[3] H. Seifi, F. Fazlollahi, M. Oppermann, A. Sastrillo, John, J. Ip, A. Agrawal, G. Park, K. J. Kuchenbecker, and K. E. MacLean, "Haptipedia: Accelerating haptic device discovery to support interaction and engineering design," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[4] H. Seifi, M. Oppermann, J. Bullard, K. E. MacLean, and K. J. Kuchenbecker, "Capturing experts' mental models to organize a collection of haptic devices: Affordances outweigh attributes," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[5] F. Fazlollahi and K. J. Kuchenbecker, "Haptify: A measurement-based benchmarking system for grounded force-feedback devices," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 1622–1636, Apr. 2023.

[6] F. Conti, F. Barbagli, R. Balaniuk, M. Halg, C. Lu, D. Morris, L. Sentis, J. Warren, O. Khatib, and K. Salisbury, "The CHAI libraries," in *Proceedings of Eurohaptics 2003*, Dublin, Ireland, 2003, pp. 496–500.

[7] G. Terry, N. Hayfield, V. Clarke, and V. Braun, "Thematic analysis," *The SAGE Handbook of Qualitative Research in Psychology*, vol. 2, pp. 17–37, 2017.