2025 IEEE World Haptics Conference (WHC)
Suwon Convention Center, Suwon, Korea
July 8 ~ 11, 2025

Quantifying Texture-Rendering Quality Across Haptic Devices

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

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, including research articles and datasheets, presents a subset of technical specifications provided by the inventors for evaluation purposes [1]. The online tool Haptipedia visualizes these specifications for more than one hundred kinesthetic haptic devices [2]. However, researchers have shown that expert opinions about a device's capabilities go beyond low-level specifications, instead emphasizing hands-on experience with the devices [3]. To better capture the haptic interaction from the user's perspective, Fazlollahi and Kuchenbecker created Haptify [4], a benchmarking system that uses external sensors to evaluate grounded force-feedback (GFF) devices and compare them from the user's perspective. Seeking to connect expert insights with sensor measurements recorded during haptic interactions, we engaged 16 expert hapticians in a user study to guide the development of a reliable benchmarking approach, collecting rich data on the quality of four representative GFF haptic devices: Novint Falcon, Force Dimension Omega.3, 3D Systems Touch X, and 3D Systems Touch (Fig. 1A). The experts assessed the devices in two rounds: first unpowered, then actively rendering five different virtual environments. We recorded 3D force, position, velocity, and acceleration with Haptify [4], along with the sensed position and commanded force from CHAI3D, and qualitative feedback from surveys. By combining the expert opinions with Haptify's quantitative data, we aim to propose a metric for assessing texture-rendering quality in this WIP paper.

II. User Study and Analysis

After consenting and completing a background questionnaire, the expert participant examined each device unpowered and rated five key characteristics: workspace size, gravity balance, effective inertia, friction, and mechanical smoothness. Based on this interaction, they also estimated the haptic rendering quality of each device. Subsequently, the expert assessed the performance of each device in five active

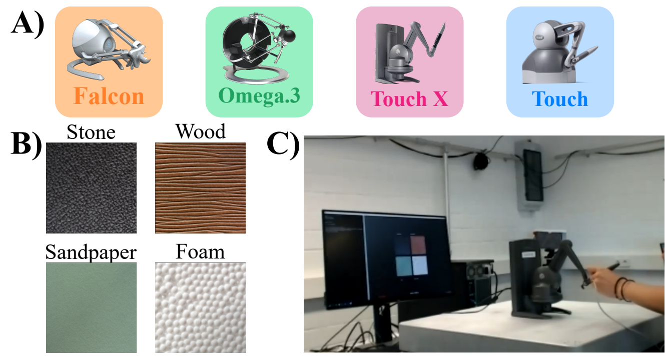

Fig. 1. A) The four force-feedback devices tested. B) The four textures rendered in the virtual environment. C) The Touch X device being tested.

virtual benchmark environments: stiffness, damping, force field, textures, and magnets. Each benchmark was designed to assess a specific aspect of GFF device performance. This work reports on only the textures benchmark, which includes four different virtual materials: hard stone, wooden board, foam, and sandpaper (Fig. 1B). These four textured squares were located at the center of the device's workspace. The expert used each device to freely interact with the textures for as long as they wanted and rated the realism and texture-rendering capabilities of the four devices (Fig. 1C). Each expert also answered open-ended questions about the criteria they used to determine texture-rendering quality.

1) Thematic Analysis: Responses to the open-ended questions were imported into MAXQDA (VERBI GmbH) and analyzed using thematic analysis [5]. Two authors independently open-coded the responses and merged similar codes to generate a list of issues used by experts to evaluate texture-rendering quality. To identify perceived differences among the devices, we calculated the percentage of experts who referenced each issue. The most frequently mentioned by the experts was *textures felt non-realistic*.

2) Statistical Analysis: Experts also rated how successful each device was at rendering realistic textures on a visual analog scale from 0 (very unsuccessful) to 10 (very successful). Since the data violated the assumption of normality (Shapiro-Wilk test), we performed the Friedman test followed by pairwise comparisons using Wilcoxon signed-rank tests to assess differences in expert ratings between devices.

3) Quantitative Analysis: Based on expert responses to the open-ended questions regarding key parameters for texture rendering, we identified a quantitative performance metric from the recorded sensor measurements. Since the most frequently mentioned issue was that textures did not feel

[1]F. Fazlollahi, G. Ballardini, Z. Taghizadeh, A. K. Schulz, and K. J. Kuchenbecker are with the Haptic Intelligence Department, 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

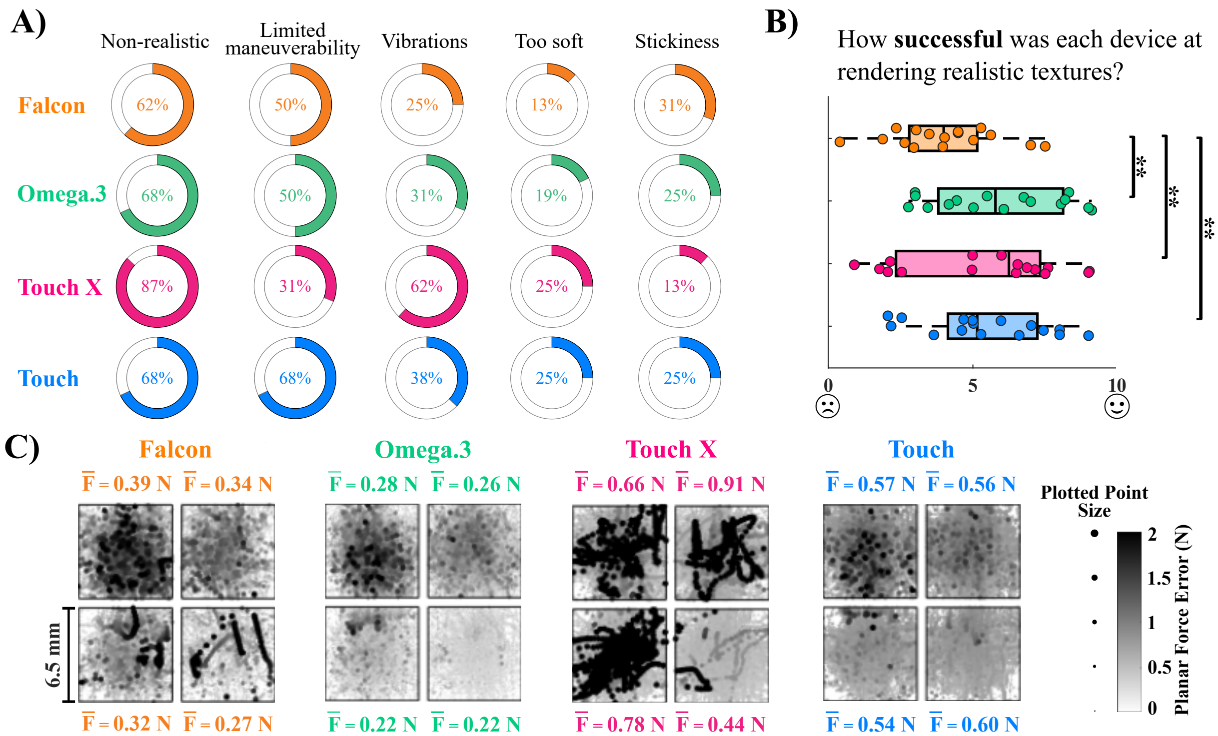

Fig. 2. A) The top five issues mentioned by experts, with the percentage of experts who mentioned each issue. B) Expert ratings of the four devices on texture-rendering success (from 0 indicating negative assessment to 10 indicating positive assessment); ** indicates $p < 0.01$. C) Magnitude of planar force error by device and texture, with texture locations and sizes. The four textures are arranged in the same way they were seen by participants (Fig. 1B). Larger errors are represented by larger points and darker colors, ensuring they remain visible and are not obscured by those with smaller force errors. The average of the magnitude of the planar force error for each texture and device is presented near the respective texture.

realistic, we computed the planar force error by subtracting the force measured by Haptify's force plate from the corresponding commanded force at each timestamp for all experts. This metric quantifies the deviation between the device's output and the intended texture rendering, with larger values indicating greater discrepancies and thus lower realism.

III. Preliminary Results and Future Plans

The issue of *textures felt non-realistic* was reported by 87% of experts for Touch X, 68% for Touch and Omega.3, and 62% for Falcon. The percentages of the other commonly mentioned issues—*limited maneuverability*, *vibrations*, *too soft*, and *stickiness*—are shown in Fig. 2A. The Falcon was rated significantly less successful at rendering realistic textures than the other three devices (all $p < 0.01$; Fig. 2B). This poor overall assessment, despite relatively few issue mentions, likely reflects experts' perception of its textures as uniformly mediocre rather than one or two textures being distinctly problematic. The heat map of planar force error for each device and texture is shown in Fig. 2C. Notably, there is an agreement between the qualitative and quantitative data for this benchmark, as the most frequently mentioned non-realistic device was the Touch X, which exhibited the highest planar force errors for three of the textures. However, the Touch X was capable of rendering fine details, particularly the ridges and grooves of the wooden board, which experts appreciated, resulting in high overall ratings. Better maneuverability of the Touch X (only 31% of experts mentioned maneuverability issues), leading experts to try higher accelerations, could also result in higher force errors. The Falcon's poor overall ratings suggest that the planar force error metric alone may not fully capture texture-rendering quality. Additional metrics, such as normal force error and stickiness, could provide a more comprehensive assessment. Future work will analyze the other four active benchmarks and the interactions with unpowered devices.

Acknowledgments

The authors thank Lijuan Wang, Alborz Aghamalekisarvestani, 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.

[2] H. Seifi, F. Fazlollahi, M. Oppermann, J. A. Sastrillo, 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 *Proc. of the ACM CHI Conference on Human Factors in Computing Systems, ACM.*, 2019.

[3] 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 *Proc. of the ACM CHI Conference on Human Factors in Computing Systems, ACM.*, 2020.

[4] 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, 2023.

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