ETHOS: A Robotic Encountered-Type Haptic Display for Social Interaction in Virtual Reality

Eric Godden [1], Jacquie Groenewegen [1], and Matthew K.X.J. Pan [1]

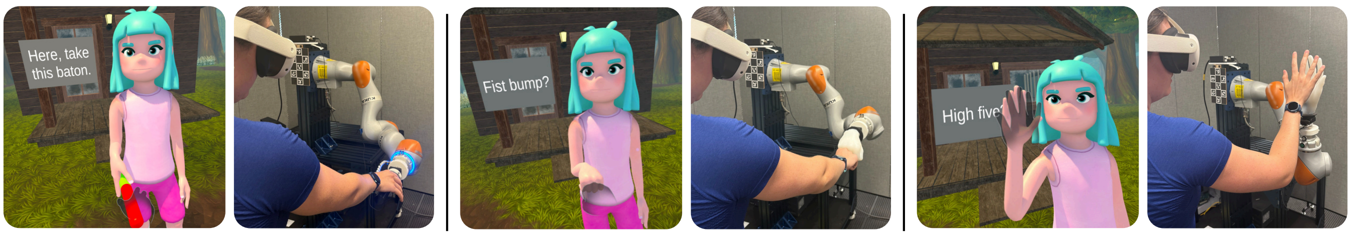

Fig. 1. ETHOS (Encountered-Type Haptics for On-demand Social interaction) enables corresponding virtual and physical renderings of dynamic interpersonal interactions, demonstrated here with an object handover (left), fist bump (centre), and high five (right).

***Abstract*— We present ETHOS (Encountered-Type Haptics for On-demand Social interaction), a dynamic encountered-type haptic display (ETHD) that enables natural physical contact in virtual reality (VR) during social interactions such as handovers, fist bumps, and high-fives. The system integrates a torque-controlled robotic manipulator with interchangeable passive props (silicone hand replicas and a baton), marker-based physical-virtual registration via a ChArUco board, and a safety monitor that gates motion based on the user's head and hand pose. We introduce two control strategies: (i) a *static* mode that presents a stationary prop aligned with its virtual counterpart, consistent with prior ETHD baselines, and (ii) a *dynamic* mode that continuously updates prop position by exponentially blending an initial mid-point trajectory with real-time hand tracking, generating a unique contact point for each interaction. Bench tests show static colocation accuracy of 5.09 ± 0.94 mm, while user interactions achieved temporal alignment with an average contact latency of 28.58 ± 31.21 ms across all interaction and control conditions. These results demonstrate the feasibility of recreating socially meaningful haptics in VR. By incorporating essential safety and control mechanisms, ETHOS establishes a practical foundation for high-fidelity, dynamic interpersonal interactions in virtual environments.**

I. INTRODUCTION

Virtual reality (VR) enables embodied engagement with digital environments and creates immersive experiences that unlock novel affordances. Advances in hardware and content creation over the past decade have driven increasing interest in the field, supporting the adoption of VR across a broad range of domains. Demonstrated successes in training [1], rehabilitation [2], and entertainment [3] highlight the versatility of VR systems in recreating meaningful human experiences.

This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Connected Minds CF-REF.

[1] All authors are with the Ingenuity Labs Research Institute and the Faculty of Electrical and Computer Engineering, Queen's University, Kingston, Canada. {18eg16}{21jng3}{matthew.pan}@queensu.ca

Despite this success, current VR systems are mediated almost entirely by vision and audition. Many everyday experiences, especially social behaviours, depend on brief, precisely timed touch. Incorporating tactile cues can disambiguate these moments and make interactions feel natural. Yet systems for touch in VR (i.e., haptics) remain limited, as the human haptic system integrates diverse receptor signals with distinct sensory characteristics [4], making accurate reproduction inherently complex. Still, augmenting visual feedback with haptics has been shown to increase task performance [5] and presence, the feeling of "being there" in a mediated environment [6]. Importantly, the quality of feedback matters: intuitive cues enhance realism, whereas abstract or incongruent cues disrupt immersion [7].

To address this challenge, we introduce a dynamic encountered-type haptic display (ETHD) that enables realistic, on-demand physical contact in social VR interactions such as handovers, fist bumps, and high-fives. Our system, Encountered-Type Haptics for On-demand Social interaction (ETHOS), integrates a torque-controlled 7-DoF manipulator with interchangeable passive props (silicone hand replicas and a baton), aligns the physical and virtual workspaces via a headset-visible ChArUco board, and applies headset-only safety gating based on user head and hand pose. We implement two control strategies: a *static* mode for stationary alignment to enable consistent baseline contact, and a *dynamic* mode that exponentially blends a mid-point trajectory with live hand tracking to converge smoothly to the user. ETHOS achieves sub-centimetre colocation accuracy (5.09 ± 0.94 mm) and average temporal alignment of contact below perceptual thresholds (28.58 ± 31.21 ms). Together, these contributions outline a practical path toward safe, unencumbered, and socially meaningful haptics in VR.

II. BACKGROUND

A. Haptics in Virtual Reality

Haptics technologies are those that stimulate our sense of touch; the human haptic system integrates signals from *cutaneous* receptors in the skin and *kinesthetic* receptors in muscles, tendons, and joints [4]. Designing displays that address both pathways is hard in practice, so interfaces often target a subset. These focused displays inherently result in a tradeoff between fidelity and versatility; feedback is either highly realistic but application-specific, or abstract but broadly applicable [8].

Wearable and hand-held devices commonly deliver cutaneous cues in VR by stimulating the skin surface (e.g., vibrotactile motors, electrotactile arrays), which are easy to integrate and cost-effective [9], [10], [11]. These solutions preserve mobility but typically provide abstract sensations and limited force/proprioceptive information, which can reduce the naturalness of larger-scale interactions [7]. To render kinesthetic cues, grounded or body-worn kinesthetic devices apply forces and motions to the user. Commercial tabletop devices, such as Touch (3D Systems, USA) and Omega (Force Dimension, Switzerland), offer accurate force feedback over a portion of the workspace but station the user, while wearable kinesthetic systems can constrain free motion and introduce reaction forces at attachment points [12], [13].

Multi-modal approaches combine cutaneous and kinesthetic hardware to improve realism [14], [15], but complexity and scaling become an issue as the breadth of required interactions within an environment grows. This motivates the development of alternatives that keep users unencumbered while still providing physically meaningful contact.

B. Encountered-Type Haptic Displays (ETHDs)

Encountered-type haptic displays (ETHDs) are "*systems capable of positioning all or part of themselves at an encountered location, enabling the user to voluntarily elicit haptic feedback with the environment at the appropriate time and place*" [16]. Thus, ETHDs can dynamically position physical props to facilitate virtual interaction, creating passive haptic feedback. This is advantageous as the interaction with the prop's physical form produces the desired sensory experience and removes the need to actively apply forces or stimulation to the user. The passive nature of this feedback approach simplifies haptic design while naturally recreating physical interaction across both cutaneous and kinesthetic sensory channels.

Dynamic positioning in ETHDs has been realized via human operators [17], drones [18], and most commonly grounded systems. Grounded implementations, typically fixed-base robotic manipulators, offer precise positioning and stability during contact. Their external grounding frees users from cumbersome hardware during free motion while supporting natural rendering at first contact. Despite these advantages, most ETHD systems emphasize *static* surface or object rendering [19], [20]. Some systems increase interactability with interchangeable endpoints [21] or detachable, task-relevant objects [22]. However, the rendering of complex interpersonal interactions, where micro-timing and fine spatial adjustments are critical, remains limited.

In response to this limitation, our work introduces an ETHD designed to extend beyond static object rendering by enabling dynamic contact interactions. Through real-time coordination with user motion, the system generates on-demand contact points that support rich physical engagement with a virtual character, including interactions such as handovers, high-fives, and fist bumps. In doing so, the system advances ETHDs toward high-fidelity, on-demand physical interactions in VR.

III. CONTRIBUTIONS

The work presented in this manuscript makes the following contributions:

1) *Dynamic ETHD design.* We design and implement an ETHD that supports both *static* alignment and *dynamic*, on-demand contact for social VR interactions (handover, high-five, fist bump) using a torque-controlled manipulator and interchangeable passive props.
2) *Control, registration, and safety framework.* We propose two control strategies supporting both static and dynamic interaction conditions, and introduce a headset-visible ChArUco registration pipeline coupled with pose-based safety gating to ensure safe interaction.
3) *Quantitative characterization.* We characterize key performance aspects required for realistic interaction, including physical–virtual colocation (sub-centimetre spatial error) and interaction contact latency (average delays below perceptual thresholds).

IV. SYSTEM

The ETHOS system, developed for enabling physio-social interactions in VR, has physical and virtual subsystems; alignment between these subsystems is achieved through a registration process based on fiducial tracking with the headset's onboard cameras. In parallel, a safety monitoring framework ensures that the user immersed in VR remains protected from unintended collisions with the robot.

A. Physical Subsystem

The physical subsystem comprised interactable objects mounted to the end effector of a robotic manipulator, providing realistic contact surfaces that could be dynamically positioned during interaction. We employed an LBR iiwa 7 R800 robotic manipulator (KUKA, Germany), mounted horizontally on a pedestal. The robot was torque-controlled in real time at a loop rate of 1000 Hz using KUKA's proprietary Fast Robot Interface (FRI) [23], connected to a desktop workstation (Intel Core i7-14700, 4.2 GHz, 32 GB RAM) running Ubuntu 22.04 with a real-time kernel. Interaction behaviours were implemented in software developed on the Robot Operating System 2 (ROS 2), Humble Hawksbill distribution.

A six-axis Mini40 force–torque sensor (ATI Industrial Automation, USA) was integrated with a quick-release camera mount, enabling efficient swapping and secure attachment of interactable props. For each object, the sensor was preconfigured to compensate for internal forces and torques, ensuring accurate monitoring of contact forces during use.

Fig. 2. Interactable props of a baton (left), closed fist (centre), and open hand (right) used to render an object handover, fist bump, and high five interactions, respectively.

With interpersonal interactions as the target application, we developed three custom props to represent either the virtual avatar's hand (in different configurations[1]) or a hand-held object (Figure 2). Their design and fabrication are summarized below:

1) Fist: Cast from a male human hand in a closed configuration using Dragon Skin FX-Pro[2], a platinum silicone material providing natural compliance and skin-like texture. The prop was reinforced with galvanized steel wire to emulate internal bone structure. A closed-fist model was selected to support fist-bump interactions.
2) Open Hand: Fabricated using the same silicone casting process as the fist, but with a modified wire framework to maintain the spread-finger configuration. Because the open hand is structurally less dense, additional reinforcement was added to reproduce the stiffness of outstretched fingers. This prop was designed for high-five interactions.
3) Baton: A simple cylindrical object enabling object handover interactions. The body was 3D printed in PLA and fitted with a metallic plate at one end for attachment to an electromagnet, allowing the robot to securely hold and release the baton during exchanges.

B. Virtual Subsystem

This work leverages the Meta Quest 3 head-mounted display (HMD), selected due to its popularity, influence on the VR community, and precise markerless hand-tracking as characterized in prior work [24]. The Meta Quest 3 features the Qualcomm Snapdragon XR2 Gen 2 processor with 8 GB of RAM and four integrated cameras with a depth sensor to facilitate tracking.

The virtual environment was developed using Unity 6.1 (6000.1.1f1) on a Windows 11 x64 laptop (AMD Ryzen 5800H with Radeon Graphics, 3.2GHz, 16GB RAM, NVidia GeForce RTX 3060 Laptop GPU). The Unity project was built as an Android application and deployed onto the HMD.

The camera rig, passthrough functionality, and hand tracking were implemented using the Meta XR All-in-One SDK (76.0.1)[3]. The Unity project was deployed to the Meta HMD using the Open XR plugin[4], with Meta XR Feature Group enabled to support Meta-specific functionality, alongside the XR Plugin Management Package (4.5.1)[5].

C. Physical-Virtual Coordination Subsystem

Accurate spatial registration of the virtual environment to the physical workspace is essential for seamless interaction. Conventional localization methods, such as motion capture [25] or external depth sensors [26], provide high precision but are rigid, resource-intensive, and often require additional hardware or wearables. Instead, we adopted a lightweight fiducial marker approach, using a ChArUco board[6] and the HMD's built-in cameras. This requires only a single registration procedure, after which a reference transform is stored for alignment.

ChArUco detection and pose estimation were implemented using Quest ArUco Marker Tracking[7] with OpenCV for Unity[8]. We modified the pipeline to support a custom 3×5 ChArUco board and to average board position over a five-second window while discarding outliers, yielding robust registration. The board was mounted on the robot stand, a position with an unobstructed view from the predefined interaction space. All pose data was expressed relative to this board. Headset data (head position and hand tracking) was transmitted at ∼90 Hz via Unity's `MonoBehaviour.Update()` function, matching the HMD's maximum refresh rate. Additional Boolean signals were integrated into the message-passing framework for subsystem synchronization (e.g., sequencing and safety triggers). The robot–board offset was calibrated using a motion capture system: positional differences across multiple end-point poses were averaged and corrected by subtracting the KUKA's known end-effector transform. This process enabled high confidence in the positioning of the ChArUco board and the resulting colocation process, as it serves as a shared anchor point between both virtual and physical worlds.

The robot reported end-effector position to the HMD at 1000 Hz, introducing potential packet loss due to rate mismatch. To mitigate this, important messages were transmitted

[1]https://www.clelia.ca/
[2]https://www.smooth-on.com/products/dragon-skin-fx-pro/
[3]https://developers.meta.com/horizon/downloads/package/meta-xr-sdk-all-in-one-upm/76.0.1
[4]https://docs.unity3d.com/Packages/com.unity.xr.openxr@1.14/manual/index.html
[5]https://docs.unity3d.com/Packages/com.unity.xr.management@4.5/manual/index.html
[6]https://docs.opencv.org/3.4/df/d4a/tutorial_charuco_detection.html
[7]https://github.com/TakashiYoshinaga/QuestArUcoMarkerTracking
[8]https://enoxsoftware.com/opencvforunity/

redundantly over multiple frames. All data was serialized using Google's Protocol Buffer Framework[9] and transmitted via connectionless UDP unicast, prioritizing low latency over guaranteed delivery.

D. Safety Subsystem

Robotic systems operating near humans must prioritize safety, and this requirement is amplified in VR since users cannot reliably perceive or react to the robot's movements. In ETHDs, the burden of safety rests entirely on the system. To enable on-demand haptic contact without disrupting immersion, robot motion must be predictive and ideally imperceptible, so the user remains unaware of the manipulator's configuration or intent. Unlike most robotic platforms that avoid contact altogether, ETHDs intentionally create brief, precisely timed contact events, making robust safety monitoring essential.

Prior ETHDs mitigated risk by constraining the robot's workspace to volumes overlapping only with the user's hands, eliminating the possibility of head or neck collisions for stationary participants [20], [27]. In our system, users are free to move within the environment, so continuous online monitoring is required. The robot halts immediately if the user's head enters its workspace, preventing high-severity collisions. Additionally, before interaction begins, the VR experience guides users to a designated safe zone outside the robot's reach using text prompts and visual markers.

During operation, user position is further monitored to ensure they remain within this zone. If they leave the safe area, the HMD automatically activates passthrough mode, showing the real-world environment via the onboard cameras. Finally, an external operator is present to trigger an emergency stop if needed. Although this human safeguard relies on judgment, it provides a final layer of protection expected to be required only in rare circumstances.

V. INTERPERSONAL INTERACTION DEVELOPMENT

The developed system is first applied to the recreation of physical interpersonal interactions with a virtual avatar - specifically, an object handover, fist bump, and high five as shown in Figure 1. Such interactions are well-suited for dynamic encountered-type haptics, as they require the real-time coordination of two entities with short-duration intermittent contact. These interactions will also be easily recognizable and carry familiar social context for the user to enhance the naturalness of their experience. This initial application aims to recreate meaningful interpersonal experiences in VR and develop the framework for dynamic encountered-type interactions.

A. Interaction Sequencing

Human–human physical interactions are coordinated through a combination of physical and social/cognitive cues [28]. These natural behaviours convey critical information about intent, including the spatial and temporal characteristics of the upcoming interaction. To ensure the ETHD is effective, the virtual experience must properly apply these relevant cues so that users can intuitively anticipate and engage in the interactions.

[9]https://github.com/protocolbuffers/protobuf

Fig. 3. Flowchart depicting the consistent sequencing of user and system actions for all interactions presented in this work: object handover, fist bump, and high five.

The interaction sequence encompasses the predefined steps that guide the user in completing a natural interaction with the ETHD. Mercado et al. [29] applied a pre-defined interaction sequence by indicating to the user, through the highlighting of an object, which interaction they can perform at a given time. While this may limit the perceived naturalness of the interaction, as it removes the user's free will in their environment, it presents a reasonable simplification in the implementation of an ETHD. As the object handover, fist bump, and high five share many of the same physical and social components, a consistent interaction sequence was defined. The interactions are situated within a social greeting scenario, where a virtual avatar engages the user, and follows the sequence outlined in Figure 3.

Throughout the interaction sequence, the system instructs the user with text prompts and/or spatial markers in the virtual environment. To begin, the user is placed in the virtual environment described in Section V-B. The user is directed to stand at a marked location outside the robot's workspace, near the eventual contact point, and must physically move there before the sequence proceeds. Once at the interaction location, the robot moves to a ready position: close to the contact area and at an appropriate position relative to the user for the intended interaction. The virtual avatar then enters the scene and approaches the user, waving and making eye contact as initial indicators of intent. After positioning themselves relative to the interaction mark, the avatar raises their arm to an interaction-appropriate pose that is aligned with the robot's ready position (e.g., for a fist bump, the arm is comfortably outstretched above waist height with a closed fist) as a readiness cue [28]. During this arm movement, the avatar first looks toward the intended contact location and then returns its gaze to the user once in position, following the gaze pattern of Moon et al. [30]. The robot then engages one of two control strategies, static or dynamic, as described in Section V-C. With the avatar indicating readiness, the user responds to complete the interaction. Upon detecting contact via force sensing, the robot retreats, and the avatar exits the scene.

B. Virtual Environment

The virtual interaction scene was implemented in Unity using prefabricated forest and cabin assets from the Unity Asset Store[10]. This setting was chosen to create a natural and aesthetically engaging environment where users could feel comfortable and where an appropriate soundscape can be used to mask sounds emanating from the robot and physical environment. The cabin served as a spatial anchor, allowing the avatar to enter the scene naturally rather than abruptly appearing.

A Mixamo[11] avatar with standard animations (e.g., walking, waving) was used to enhance lifelike motion. To enable realistic interactions, the avatar's arm was additionally rigged with a custom script that modelled human-like motion. Specifically, we integrated the joint angle functions defined by Rasch et al. [31], which describe shoulder and elbow trajectories during handovers. These functions allowed the avatar's arm to move smoothly from a resting position to the interaction location over a predefined motion time, producing a human-like motion. Although based on a human-human handover, the same arm movement was applied across all interactions due to their similar motion patterns.

To further increase immersion, spatial audio was incorporated into the forest environment. Forest soundscapes were delivered through Bose QuietComfort 35 II (Bose Corporation, USA) noise-cancelling headphones, enhancing alignment between the virtual setting and the user's sensory experience.

C. Interaction Control

Our system implements two strategies for robotic control in the phases before and during interaction: static and dynamic. As described in Section V-A, the robot engages one of these control strategies once the avatar has indicated readiness for interaction. In both strategies, the system continuously monitors the force exerted on the end-effector to detect user contact. Pilot studies were completed to establish the force thresholds for natural feeling interactions. The resulting force thresholds, which, when exceeded, indicate the completion of the interaction. In our system, we set these thresholds as 7.5 N, 15.0 N, and 15.0 N for the object handover, fist bump, and high five interactions, respectively. The fist bump and high five force thresholds are higher than in the object handover since these interactions involve strong impacts.

The static strategy matches the capabilities of the state-of-the-art for ETHDs by presenting a stationary object for the user to interact with. In raising their arm after approach, the virtual avatar's arm becomes aligned with the physical prop, enabling the robotic manipulator to remain stationary throughout the rest of the interaction.

The dynamic strategy, in contrast, leverages online pose tracking of the user's hands to create a unique contact point for each interaction. This strategy was inspired by the trajectory generation of Pan et al. [32], with an adjustment of how the weight term updates throughout interaction. At the start of the interaction phase, the prop is positioned at the same pre-defined location as in the static strategy. When the user's hand enters the interaction volume, defined as a 10 × 30 cm rectangle extending 30 cm forward from the prop, the robot computes an initial one-second trajectory that targets the midpoint between the prop and the hand at that instant. The target location is then updated continuously as a weighted average of this initial trajectory and the user's current hand position, as shown in Equation 2. The weight term is defined in Equation 1, where $t_{dynamic}$ denotes the time elapsed since dynamic tracking began. This formulation applies an exponential weighting, prioritizing smooth early motion by favouring the initial trajectory and gradually shifting toward the user's actual hand position as the interaction progresses. The weight is capped at a maximum value of 1, ensuring that if the interaction extends beyond the one-second allocation, the system continues to track the user's current hand position. A visualization of the progression of this tracking is provided in Figure 4.

$$w(t) = \frac{e^{3\alpha(t)} - 1}{e^3 - 1}, \quad \alpha(t) = \min\big(t_{\text{dynamic}}, 1\big) \tag{1}$$

$$x_{target}(t) = (1 - w(t))x_{midTraj}(t) + w(t)x_{hand}(t) \tag{2}$$

VI. PERFORMANCE EVALUATION

Nilsson et al. proposed two criteria for success when using physical props for rendering virtual interaction: similarity and colocation [33]. The criterion of similarity refers to the physical object's material and geometric properties sufficiently matching their virtual counterparts. The criterion of colocation refers to the spatial alignment of virtual and physical form to allow seamless interaction. The focus for this initial system characterization is colocation. Without proper colocation, the interactions may not be feasible at all, leaving the props' physical properties of secondary importance. Although similarity is not directly evaluated here, the props have shown promising initial success in recreating convincing interactions, suggesting their suitability for the present study while warranting further testing in future work.

A. Spatial Alignment

The first component of colocation is spatial alignment under static conditions. This evaluation neglects temporal

[10] https://assetstore.unity.com/
[11] https://www.mixamo.com/#/

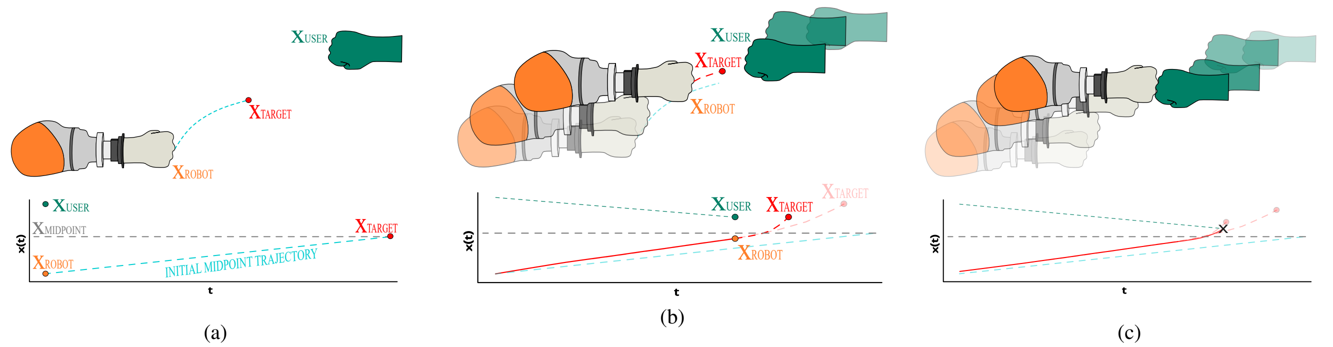

Fig. 4. Visualization of the dynamic control strategy: (a) an initial midpoint trajectory is generated as the user's hand first approaches, (b) the trajectory is blended with the user's current hand position as the interaction progresses, and (c) contact is reached at a unique location for each interaction.

factors, such as message-passing latency, to isolate the performance of our fiducial marker approach to virtual–physical alignment. As the robotic manipulator and its relative pose to the fiducial marker are highly accurate, as the marker's relative placement was determined using motion-capture as described in Section IV-C, the primary source of error in static colocation arises from the registration process. Since this process relies on computer vision tracking of the ChAruCo board, the detected anchor point between the physical and virtual environments can vary across trials, necessitating an evaluation of detection performance.

To provide ground-truth to our measurements, a retroreflective marker-based motion-capture system, Vicon Vero (v2.2)[12], was used. Vicon Tracker Software (3.9.0.128020h)[13] was run on a Windows 11 x64 system (Intel Xeon W-2123 3.6GHz, 16GB RAM) to provide sub-millimetre error tracking of marker constellations aligned with the ChAruCo board and the VR controller.

With markers placed on the ChAruCo board and VR controller, both physical and virtual positions of these objects could be tracked. The spatial alignment could then be determined by registering and recording the error between each measurement frame's relative positions with respect to the objects. This process was completed over 20 registration trials, recording the positional error five times for each trial (100 total measurements). The resulting spatial alignment error was $5.09 \pm 0.94mm$.

B. Interaction Latency

With spatial alignment characterized, temporal factors must also be considered when evaluating colocation. Interaction latency, defined as the time difference between a physical and corresponding virtual event, can significantly influence the realism of haptic interaction. In our system, latency arises from multiple sources, including message-passing delays between subsystems, processing overhead from control and rendering pipelines, and the inherent delay introduced by markerless hand-tracking [24]. While an individual analysis of each component of latency would be ideal, understanding end-to-end latency is of highest priority for the initial system characterization, as this is the final delay presented to the user.

To measure interaction latency, we compared the timing of hand–object collisions in the physical and virtual environments. A physical collision was defined as the moment when the applied force exceeded the predefined threshold, while a virtual collision was marked by the intersection of the user's hand with the object's mesh. Four pilot participants completed 25 static and 25 dynamic trials for each interaction type, producing a total of 600 latency measurements. The aggregated results from these trials are summarized in Table I below.

TABLE I

MEAN INTERACTION LATENCY RESULTS WITH ONE STANDARD DEVIATION FOR HANDOVER, FIST BUMP, AND HIGH FIVE ACROSS STATIC AND DYNAMIC CONTROL CONDITIONS

Interaction Type	**Static Latency (ms)**	**Dynamic Latency (ms)**
Handover	18.83 ± 40.65	25.72 ± 35.48
Fist Bump	36.01 ± 35.45	24.49 ± 18.77
High Five	23.29 ± 14.99	43.11 ± 27.02

VII. DISCUSSION

The development of ETHOS was motivated by the goal of making virtual experiences more meaningful by enabling natural, engaging interactions. Whereas existing ETHDs have largely emphasized static interactions with simple objects, our system advances the field by demonstrating how dynamic interactions can unlock new virtual affordances. To realize this, we designed custom physical props, implemented tailored interaction strategies, and created virtual experiences that support interpersonal interaction with a virtual avatar. We further leveraged a fiducial-based registration method to align the virtual and physical worlds, substantially reducing the tracking hardware required for this type of system. Finally, we conducted an initial system characterization to evaluate ETHOS's suitability for recreating spatially and temporally aligned interactions.

Spatial alignment is of primary importance within our system's operation. For seamless interaction, ETHDs must

[12]https://www.vicon.com/hardware/cameras/vero/

[13]https://help.vicon.com/space/Tracker39

dynamically position physical props to an aligned location with the virtual environment. If the prop's physical location does not sufficiently match its virtual presentation, the desired interaction may become unnatural or even infeasible. As we have leveraged a high-accuracy robotic manipulator, the main source of spatial misalignment within our system comes from the fiducial-based registration process. The colocation error of this registration process was determined to be $5.09 \pm 0.94mm$. This result places our alignment performance well within the threshold of 1 cm identified by Fremerey et al. [34], the maximum displacement for preserving task performance and minimizing user confusion levels in virtual object interaction with physical props. Our alignment performance ensures colocation error within static object interaction will be imperceivable to the user and confirms our fiducial-based registration process is sufficient for connecting the virtual and physical worlds.

Interactions with ETHDs demand more than accurate spatial colocation, they also require contact to occur at the right moment. Delays can disrupt the sense of realism, making precise temporal alignment essential for convincing interactions. In our system, sources of delay such as message passing and hand-tracking latency can shift the timing of contact, creating misalignment between the user's hand and the physical prop. Characterizing these temporal factors is therefore critical to assessing the suitability of our system for recreating natural interactions. Latency perception thresholds (LPTs) define the maximum tolerable delay in a system before user experience or task performance is degraded, and they have long been a focus of VR research. Factors such as task difficulty and interaction speed can significantly influence LPTs [35], with the addition of haptic feedback further heightening sensitivity to delay [36]. As a result, defining a single LPT for ETHD interactions is particularly challenging, as they span phases ranging from free-hand motion to physical contact with a haptic prop, each aligning more closely with different thresholds reported in the literature. For example, an LPT derived from a virtual pointing task (130–170 ms) [35] differs substantially from one based on a tapping task with a haptic feedback device (50 ms) [36]. Despite this variability, the average interaction latencies measured in this work for both static and dynamic tracking conditions remained below all of these reported thresholds. However, when one standard deviation is considered, the most stringent LPT discussed here (50 ms) is exceeded in many trials. This suggests that while mean system performance is sufficient even under highly sensitive conditions, perceptible latency effects may still arise.

As this study represents the first development and characterization of ETHOS, its limitations point directly to future work. Our results show that the system can recreate spatially and temporally aligned interactions under both static and dynamic conditions, demonstrating the feasibility of object handovers, fist bumps, and high-fives. However, subjective evaluation is needed to assess user experience. A dedicated user study, currently underway, will provide critical insight into experiential factors such as naturalness and presence, which underpin our motivation to recreate these interactions in VR. Preliminary impressions already highlight the impact of physicality in enhancing realism ("I felt like the added component of physical touch enhanced my experience of the interaction being real"), enjoyment ("It was fun to physically interact with the avatar"), and connection to the avatar ("My favourite interaction was the [static] fist bump...it made me feel more connected to the virtual character"). These early findings underscore the promise of ETHOS in delivering meaningful interpersonal interactions and reinforce the value of embedding physical touch in virtual environments.

The system incorporates several simplifications that warrant further study. The physical props were designed to approximate real-world counterparts but were not formally characterized, the dynamic interaction control strategy was adapted from object handovers and applied across all interaction types, and the force thresholds were heuristically defined from preliminary pilot observations without systematic justification or consideration of torques. Future work should address these limitations by characterizing prop materials, developing interaction-specific control strategies, and determining representative force–torque profiles. In addition, manipulator compliance should be explored to better reflect the natural compliance of the human arm across the phases of the developed interactions.

VIII. CONCLUSIONS

In this work, we introduced ETHOS (Encountered-Type Haptics for On-demand Social interaction), extending the capabilities of ETHDs beyond static props toward dynamic, socially meaningful contact in VR. By equipping a torque-controlled manipulator with customized physical props, we realized novel interpersonal interactions—handover, fist bump, and high five—that are intuitive, repeatable, and grounded in everyday social behaviour.

Our initial characterization demonstrated sub-centimetre spatial colocation and interaction latencies below perceptual thresholds, establishing that ETHOS can deliver spatially and temporally aligned contact even under dynamic tracking. These results provide the first quantitative evidence that high-fidelity interpersonal interactions are feasible in VR through an ETHD framework.

Looking forward, the technical foundations established here enable a broad set of research directions. Immediate priorities include controlled user studies to evaluate experiential qualities such as naturalness, presence, and trust. Longer term, ETHOS opens pathways toward multi-user encounters, adaptive prop design for richer social gestures, and integration into applications such as immersive training, physical rehabilitation, social VR platforms, and VR theatre.

By demonstrating that unencumbered users can safely and convincingly exchange touch with a virtual character, ETHOS charts a practical path for embedding physicality into virtual experiences. We view this as a critical step toward VR systems that are not only visually and aurally rich, but also socially and physically engaging.

References

[1] M. C. Howard, M. B. Gutworth, and R. R. Jacobs, "A meta-analysis of virtual reality training programs," *Comput. in Human Behav.*, vol. 121, p. 106808, Aug. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S074756322100131X

[2] M. C. Howard, "A meta-analysis and systematic literature review of virtual reality rehabilitation programs," *Comput. in Human Behav.*, vol. 70, pp. 317–327, May 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563217300134

[3] G. Cimolino, T. N. Graham, L. Levin, M. Wales, and M. Wheeler, "You Should Have Stayed Home: How to Captivate an Audience in VR Theatre," in *Extended Abstr. of the Annu. Symp. on Comput.-Human Interact. in Play*. Bremen Germany: ACM, Nov. 2022, pp. 333–338. [Online]. Available: https://dl.acm.org/doi/10.1145/3505270.3558383

[4] K. Hale and K. Stanney, "Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations," vol. 24, no. 2, pp. 33–39, IEEE Comput. Graph. and Appl. [Online]. Available: https://ieeexplore.ieee.org/document/1274059/?arnumber=1274059&tag=1

[5] J. Kreimeier, S. Hammer, D. Friedmann, P. Karg, C. Bühner, L. Bankel, and T. Götzelmann, "Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality," in *Int. Conf. on PErvasive Technol. Related to Assistive Environ.* ACM, pp. 289–298. [Online]. Available: https://dl.acm.org/doi/10.1145/3316782.3321536

[6] J. K. Gibbs, M. Gillies, and X. Pan, "A comparison of the effects of haptic and visual feedback on presence in virtual reality," vol. 157, p. 102717. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S107158192100135X

[7] B. Weber and C. Eichberger, "The benefits of haptic feedback in telesurgery and other teleoperation systems: A meta-analysis," in *Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being*, M. Antona and C. Stephanidis, Eds. Springer International Publishing, vol. 9177, pp. 394–405, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-319-20684-4_39

[8] T. Muender, M. Bonfert, A. V. Reinschluessel, R. Malaka, and T. Döring, "Haptic fidelity framework: Defining the factors of realistic haptic feedback for virtual reality," in *CHI conf. on human factors in comput. syst.*, 2022, pp. 1–17.

[9] C. Pacchierotti, S. Sinclair, M. Solazzi, A. Frisoli, V. Hayward, and D. Prattichizzo, "Wearable haptic systems for the fingertip and the hand: taxonomy, review, and perspectives," *IEEE trans. on haptics*, vol. 10, no. 4, pp. 580–600, 2017.

[10] A. Adilkhanov, M. Rubagotti, and Z. Kappassov, "Haptic devices: Wearability-based taxonomy and literature review," *IEEE Access*, vol. 10, pp. 91 923–91 947, 2022.

[11] W. Guo, Y. Hu, Z. Yin, and H. Wu, "On-skin stimulation devices for haptic feedback and human–machine interfaces," vol. 7, no. 2, p. 2100452, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/admt.202100452. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/admt.202100452

[12] X. Gu, Y. Zhang, W. Sun, Y. Bian, D. Zhou, and P. O. Kristensson, "Dexmo: An inexpensive and lightweight mechanical exoskeleton for motion capture and force feedback in vr," in *CHI Conf. on Human Factors in Comput. Syst.*, 2016, pp. 1991–1995.

[13] I. Choi, E. W. Hawkes, D. L. Christensen, C. J. Ploch, and S. Follmer, "Wolverine: A wearable haptic interface for grasping in virtual reality," in *Int. Conf. on Intell. Robots and Syst. (IROS)*. IEEE, 2016, pp. 986–993.

[14] Z. F. Quek, S. B. Schorr, I. Nisky, W. R. Provancher, and A. M. Okamura, "Sensory substitution using 3-degree-of-freedom tangential and normal skin deformation feedback," in *Haptics Symposium (HAPTICS)*. IEEE, 2014, pp. 27–33.

[15] R. M. Pierce, E. A. Fedalei, and K. J. Kuchenbecker, "A wearable device for controlling a robot gripper with fingertip contact, pressure, vibrotactile, and grip force feedback," in *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, 2014, pp. 19–25.

[16] V. R. Mercado, M. Marchal, and A. Lécuyer, ""haptics on-demand": A survey on encountered-type haptic displays," *IEEE Trans. on Haptics*, vol. 14, no. 3, pp. 449–464, 2021.

[17] L.-P. Cheng, T. Roumen, H. Rantzsch, S. Köhler, P. Schmidt, R. Kovacs, J. Jasper, J. Kemper, and P. Baudisch, "Turkdeck: Physical virtual reality based on people," in *ACM Symp. on User Interface Softw. & Tech.*, 2015, pp. 417–426.

[18] P. Abtahi, B. Landry, J. Yang, M. Pavone, S. Follmer, and J. A. Landay, "Beyond the force: Using quadcopters to appropriate objects and the environment for haptics in virtual reality," in *CHI Conf. on Human Factors in Comput. Syst.*, 2019, pp. 1–13.

[19] S. Mortezapoor, K. Vasylevska, E. Vonach, and H. Kaufmann, "Cobodeck: A large-scale haptic vr system using a collaborative mobile robot," in *Conf. Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2023, pp. 297–307.

[20] E. Vonach, C. Gatterer, and H. Kaufmann, "Vrrobot: Robot actuated props in an infinite virtual environment," in *IEEE Virtual Reality (VR)*, 2017, pp. 74–83.

[21] B. Araujo, R. Jota, V. Perumal, J. X. Yao, K. Singh, and D. Wigdor, "Snake charmer: Physically enabling virtual objects," in *Int. Conf. on Tangible, Embedded, and Embodied Interact.*, ser. TEI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 218–226. [Online]. Available: https://doi.org/10.1145/2839462.2839484

[22] V. R. Mercado, T. Howard, H. Si-Mohammed, F. Argelaguet, and A. Lécuyer, "Alfred: the haptic butler on-demand tangibles for object manipulation in virtual reality using an ethd," in *IEEE World Haptics Conf. (WHC)*, 2021, pp. 373–378.

[23] G. Schreiber, A. Stemmer, and R. Bischoff, "The fast research interface for the kuka lightweight robot," in *IEEE workshop on innovative robot control architectures for demanding (Research) applications how to modify and enhance commercial controllers (ICRA 2010)*, pp. 15–21.

[24] E. Godden, W. Steedman, and M. K. Pan, "Robotic characterization of markerless hand-tracking on meta quest pro and quest 3 virtual reality headsets," *IEEE Trans. on Visualization and Comput. Graph.*, 2025.

[25] M. K. Pan and G. Niemeyer, "Catching a real ball in virtual reality," in *Virtual Reality (VR)*. IEEE, 2017, pp. 269–270.

[26] M. Sra and C. Schmandt, "Bringing real objects, spaces, actions, and interactions into social vr," in *IEEE VR Int. Workshop on Collaborative Virtual Environ. (3DCVE)*, 2016, pp. 16–17.

[27] V. K. Guda, D. Chablat, and C. Chevallereau, "Safety in a human robot interactive: Application to haptic perception," in *Int. Conf. on Human-Comput. Interact.* Springer, 2020, pp. 562–574.

[28] K. Strabala, M. K. Lee, A. Dragan, J. Forlizzi, S. S. Srinivasa, M. Cakmak, and V. Micelli, "Toward seamless human-robot handovers," *J. of Human-Robot Interaction*, vol. 2, no. 1, pp. 112–132, 2013.

[29] V. Mercado, M. Marchal, and A. Lécuyer, "Entropia: Towards infinite surface haptic displays in virtual reality using encountered-type rotating props," *IEEE Trans. on Visualization and Comput. Graph.*, vol. 27, no. 3, pp. 2237–2243, 2021.

[30] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft, "Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing," in *ACM/IEEE Int. Conf. on Human-robot Interact.*, 2014, pp. 334–341.

[31] R. Rasch, S. Wachsmuth, and M. König, "A joint motion model for human-like robot-human handover," in *Int. Conf. on Humanoid Robots (Humanoids)*. IEEE, 2018, pp. 180–187.

[32] M. K. Pan, E. Knoop, M. Bächer, and G. Niemeyer, "Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities," in *IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS)*, 2019, pp. 6735–6741.

[33] N. C. Nilsson, A. Zenner, and A. L. Simeone, "Propping up virtual reality with haptic proxies," vol. 41, no. 5, pp. 104–112, IEEE Comput. Graph. and Appl. [Online]. Available: https://ieeexplore.ieee.org/document/9535175/?arnumber=9535175

[34] S. Fremerey, M. S. Suleman, A. H. A. Paracha, and A. Raake, "Development and evaluation of a test setup to investigate distance differences in immersive virtual environments," in *Int. Conf. on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–4.

[35] S. Yang, K. Yue, H. Gao, M. Guo, Y. Liu, D. Zhang, and Y. Liu, "Dynamic changes of latency perception threshold in virtual reality: Behavioral and eeg evidence," *IEEE Trans. on Visualization and Comput. Graph.*, 2025.

[36] I. Lee and S. Choi, "Discrimination of visual and haptic rendering delays in networked environments," *Int. J. of Control, Automat. and Syst.*, vol. 7, no. 1, pp. 25–31, 2009.