Reliability in Focus: Trust, Agency, Ownership, and Gaze Behavior in a VR Prosthesis Simulator

Fabio Egle, *Graduate Student Member, IEEE*, Constantin Kleinbeck, Liv Herzer, Philipp Beckerle, *Senior Member, IEEE*, Daniel Roth, *Member, IEEE*, and Claudio Castellini, *Member, IEEE*

***Abstract*— Psychological factors such as ownership, agency, and trust are critical to the acceptance and effective use of prosthetic devices, yet their relationship to control reliability remains underexplored. We investigated how induced delays and artificial malfunctions influence these factors during prosthesis simulator use in a fully immersive virtual reality environment. A Pasta Box Task was implemented in Unity with MuJoCo physics simulation, using surface electromyography myocontrol and integrated eye tracking to measure subjective and visuomotor responses. Thirty non-disabled participants completed six within-participant conditions crossing two control delay and three artificial malfunction levels. Validated questionnaires assessed ownership, agency, and trust, while gaze metrics quantified fixation percent, target-locking strategy, and eye arrival and leaving latencies. Both delay and malfunction significantly reduced psychometric scores, with artificial malfunctions exerting the largest overall effect, while delay particularly diminished agency. Artificial malfunctions also increased fixations on the prosthesis and altered gaze strategies, suggesting compensatory behavior. Delay primarily affected eye–arrival latency and the number of fixations, whereas artificial malfunctions influenced target-locking strategy and eye–leaving latency, indicating distinct visuomotor adaptations to each reliability factor. Weak but significant correlations emerged between gaze behavior and psychometric measures. The results of the experiment highlight the value of immersive, physics-accurate virtual reality as an early-stage platform for the controlled evaluation of myocontrol and prosthesis behavior and for capturing relevant psychometric and visuomotor indicators relevant to user-centered design.**

***Index Terms*— Prosthesis, trust, embodiment, virtual reality, eye tracking, myoelectric control, human–machine interaction.**

Received 23 August 2025; revised 2 February 2026; accepted 12 March 2026. Date of publication 23 March 2026; date of current version 10 April 2026. Associate Editor: Jonathon W. Sensinger. This work was supported in part by European Commission's Horizon Europe Framework Program with the project IntelliMan (AI-Powered Manipulation System for Advanced Robotic Service, Manufacturing and Prosthetics) under Grant 101070136 and in part by German Federal Ministry of Education and Research (BMBF) through the Robotics Institute Germany (RIG). The work of Daniel Roth and Constantin Kleinbeck were supported in part by the German Federal Ministry of Research, Technology and Space, "Robotics Institute Germany (RIG)" under Grant #16ME0997K and in part by the Munich Institute of Robotics and Machine Intelligence (MIRMI). *(Corresponding author: Fabio Egle.)*

Fabio Egle and Liv Herzer are with The Assistive Intelligent Robotics Laboratory, Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany (e-mail: fabio.egle@fau.de).

Constantin Kleinbeck and Daniel Roth are with the Human-Centered Computing and Extended Reality Lab and the Munich Institute of Robotics and Machine Intelligence of the Technical University of Munich, 81675 München, Germany, and also with the TUM University Hospital, Clinic for Orthopedics and Sports Orthopedics, Technical University of Munich, 81675 München, Germany.

Philipp Beckerle is with the Chair of Autonomous Systems and Mechatronics, the Department of Electrical Engineering, and the Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany.

Claudio Castellini is with The Assistive Intelligent Robotics Laboratory, Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany, and also with the Institute of Robotics and Mechatronics, German Aerospace Center, 82234 Weßling, Germany.

Digital Object Identifier 10.1109/TNSRE.2026.3676703

I. Introduction

RESEARCH on myocontrol methods and prosthesis design has made enormous progress in the last decades [1]. Still, current prosthetic systems in clinical settings face a very high abandonment rate [2], [3]. To overcome the issues at hand and achieve a more rapid translation from research to the clinics, we believe that myocontrol development and early assessment need to align closer with clinical reality and activities of daily living. This requires easily accessible, standardized, and realistic human-in-the-loop (HIL) tests that can be carried out by research groups with minimal effort and investment, facilitating the collection of both functional and psychosocial parameters.

In recent years more studies have focused on prosthetic embodiment [4], [5], [6]. However, only a few studies have investigated the concept of trust in prosthesis, despite findings that upper- and lower-limb prosthesis users link trust to system reliability, responsiveness, predictability, and past experience [7], [8]. Research in the field of robotics and automation shows that trust is influenced by reliability and can change dynamically over time [9], [10], [11], [12], [13], with reduced reliability leading to measurable decreases in trust [14], [15].

In this work, we present an experiment to assess ownership, agency, and trust during prosthesis use inside a fully immersive virtual reality (VR) environment. The setup implemented a physics-based version of the Pasta Box Task (PBT) [16] utilizing MuJoCo [17] within Unity [18], allowing systematic manipulation of control reliability via artificial delays and malfunctions. Built-in eye-tracking enabled us to capture detailed gaze metrics, alongside subjective measures. We investigated how six combinations of additional control delay (none vs. 300 ms) and artificial malfunction frequency (none, medium, high) affect ownership, agency, trust, and gaze behavior, with the eye-tracking analysis serving as the main behavioral layer for future translational prosthesis assessment.

Based on previous research on embodiment in prosthetics, latency effects in VR, and trust in automation, we anticipated that both artificial malfunctions and delays would influence participants' subjective and gaze-based responses. Specifically, we expected malfunction frequency to be associated with lower ratings of ownership, agency, and trust, and delays to particularly affect agency. For gaze behavior, we anticipated that reduced control reliability would lead to increased fixations on the prosthesis and reduced target-locking.

II. Related Work

In the evaluation of upper-limb prosthetics, standardized functional protocols such as the Southampton Hand Assessment Procedure (SHAP) and the Assessment of Capacity for Myoelectric Control (ACMC) quantify task performance and control capacity at different levels of abstraction of activities of daily living (ADLs) [19], [20]. Among prosthetic assessments that incorporate ADLs, the Pasta Box Task (PBT) falls somewhere in between. First published in [16], it has been used in multiple studies investigating motor planning and gaze assessment, including studies with prosthesis users [21], [22], [23], [24], [25], [26], [27]. The PBT imitates a kitchen setting by requiring participants to move a pasta box between shelves of different heights. Participants must therefore move on various planes of motion, including *Reach*, *Grasp*, *Transport*, and *Release* phases.

Although the described assessment tests allow the functionality of the prosthesis and/or the investigation of musculoskeletal parameters to be evaluated, they are limited in their ability to represent participants' subjective perceptions and, consequently, the psychosocial outcomes of prosthetic control by itself. Many past studies have focused on or included the topic of prosthetic embodiment and often relied on explicit measures through questionnaires. Recently, however, an increasing number of prosthetic studies have employed implicit embodiment assessments, including intentional binding as a measure of agency (e.g. [4]) and the proprioception drift as a measure of ownership (e.g. [4], [5], [6])

VR researchers have long explored the idea of virtual embodiment, based on body ownership and embodiment in the real world [28]. It is commonly described as the perception of something not originally part of the user's body being experienced and processed as one [29]. Factors like motor synchrony and appearance are attributed to its emergence [30]. However, the literature identifies multiple pathways by which body ownership can be induced, including visuo-tactile synchrony [31], visuo-motor synchrony [32], [33], visuo-proprioceptive congruence [34], [35], and first-person perspective alignment [33], [36]. These mechanisms have been shown to generate ownership even in the absence of tactile feedback, particularly in immersive VR scenarios [37]. Embodiment is often explored with full body avatars, but just as relevant when just looking at specific body parts, as shown in the foundational rubber hand experiment [31]. Via body ownership and agency, embodiment is influenced by how well they feel their virtual self responds to their input. Here, ownership describes that something virtual is perceived as part of our body. Agency, on the other hand, is the experience of initiating and controlling virtual actions. It occurs when expected outcomes of actions align with what we perceive, reinforcing the belief that we caused the result.

In the field of prosthetic research, embodiment is also broken down to the concept of ownership and agency. Induced through synchronized sensory input, prosthetics can be experienced as natural extensions of the body, therefore fostering an update of the body's representation in the brain. This experience enables researchers to capture in detail how users of the system perceive their virtual prosthesis [38], [39]. Ownership and agency are influenced by how well individuals feel their prosthesis responds to their input. Control fidelity [30] and end-to-end latency can affect user experience and performance in immersive VR [40]. Visuo-motor delays have also been shown to modulate ownership and agency in immersive VR and related hand-illusion paradigms [41], [42]. Additionally, embodiment can amplify affective responses to virtual stimuli [43] and may shape attitudes toward system reliability. Therefore, prosthesis control behavior is paramount.

While the concept of embodiment has been incorporated into the scientific debate on prostheses, the role of trust is still comparatively underrepresented. Qualitative reports from prosthesis users indicated that trust is closely tied to perceived reliability of mechatronic design and control, and that improved sensory feedback can support this trust [7]. Lower-limb prosthesis users further reported that higher trust reduced the need to continuously monitor the device and attributed trust to knowledge of capabilities, past experiences, responsiveness, predictability, and the level of prosthesis technology [8]. In robotics and automation, trust has been studied more extensively and is commonly described as a multidimensional construct capturing perceptions of system *purpose*, *process*, and *performance* as well as *ability*, *integrity*, and *benevolence* [9], [10], [11], [12]. Importantly, trust is dynamic and can evolve through phases of *formation*, *dissolution*, and *restoration* [9], [13]. It is not an objective evaluation of functionality alone, but integrates perceived design and performance on a subjective level and can therefore be informative about future reliance and use behavior [44]. Trust is strongly related to perceived reliability and system faults [9], and decreases have been observed under reduced reliability and malfunctioning control in prior human-robot interaction (HRI) and teleoperation settings [14], [15]. Consequently, we included trust as a primary outcome alongside ownership and agency in our reliability manipulation.

Human gaze has long been used to study visuomotor control in musculoskeletal and prosthetic research. In pick-and-place actions, gaze typically dwells on key landmarks: it is directed to the object before and during *Grasp* and shifts

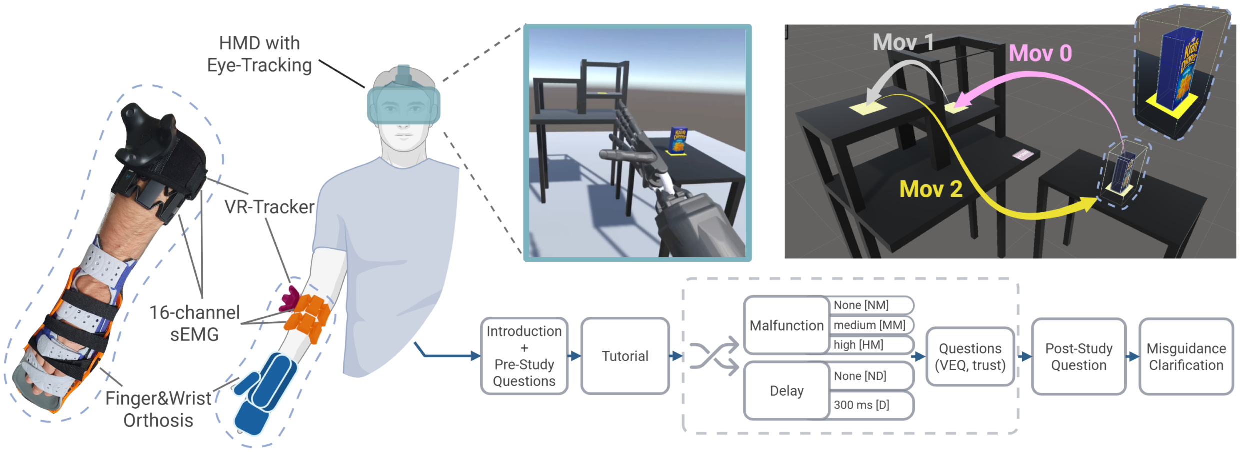

Fig. 1. **Experimental Setup and Protocol**. On the left, the sensor setup and attachment are depicted. The left breakout view shows the real-life attachment. On the right, the virtual reality environment, i.e., the pasta box setup with the order of movements, is displayed. The right breakout highlights the grasp collider surrounding the pasta box used to define the *Reach*, *Grasp*, *Transport*, and *Release* phases. Below, the study protocol of the within-participant study is shown. Abbreviations: sEMG = surface electromyography; VR = virtual reality; HMD = head-mounted display (with eye-tracking); VEQ = Virtual Embodiment Questionnaire; Mov0/Mov1/Mov2 = first, second and third movement in the Pasta Box Task; NM/MM/HM = no, medium and high malfunction level; ND/D = no delay and 300ms delay.

to the target during *Transport* [45]. In functional prosthesis tasks, this pattern can change, with users showing prolonged monitoring of the prosthetic hand during *Reach* and *Transport* [27]. Accordingly, several myoelectric-control studies use gaze features (e.g., hand/target fixation and gaze switching) as objective markers of motor planning demands and control reliability [22], [23], [24], [25], [46]. For example, a position-aware controller reduced fixations on the prosthetic hand [23], whereas increased unpredictability and delays have been linked to more hand monitoring, reduced target fixations, and more gaze switches [47].

Immersive VR enables standardized, instrumented HIL studies with integrated sensing (e.g., eye tracking) and experimental control. Prior AR/VR prosthesis training shows mixed skill-transfer results [48], [49], [50], [51], but positive transfer is more often reported in more realistic prosthetic scenarios [52], [53], [54]. To increase interaction fidelity, physics engines such as MuJoCo have been adopted in both non-immersive and immersive prosthesis simulation [17], [55], [56], [57].

Furthermore, earlier work has already dealt with the VR-based assessment or training of prosthesis embodiment [5], [6], [58], [59]. In [5], different numbers of surface electromyography (sEMG) sensors and induced delays in the myocontrol, and their influence on prosthesis embodiment, were investigated. While a higher number of electrodes led to a significantly increased embodiment score, the delay itself did not have a significant effect on the embodiment score.

III. Methods

A. Experimental Setup

1) Hardware: The participants wore two sEMG armbands (Myo Thalmic Labs) with eight channels each on their right forearm, an HTC Vive Eye Pro head-mounted display (HMD) and tracker, and an adapted finger and wrist orthosis (ManuLoc Long Plus, Bauerfeind), as shown in Figure 1. If necessary, the wrist and fingers were further blocked with an additional 3D-printed brace and straps to induce isometric muscle contractions. The armbands stream via Bluetooth Low Energy (BLE) with a sampling frequency f_s = 200 Hz and an 8-bit resolution. At the host PC, both BLE streams were combined within MOSAIC, a C#-based signal processing framework [60], where further signal processing and model learning were carried out (Section III-A.3). The HTC Vive Eye Pro includes eye-tracking capability with a sampling rate of up to 120 Hz and describes combined gaze accuracy of 1.08 deg [61].

2) Virtual Environment: The developed setup is based on the PBT first published in [16] and furthermore used for prosthetic research in [21], [22], [23], [24], [25], [26], and [27]. Designed to imitate a kitchen setting, the PBT requires participants to move a pasta box between shelves of varying heights, as shown in Figure 1. Therefore, participants must move through different planes of motion, including *Reach*, *Grasp*, *Transport*, and *Release* phases. The diameter and weight distribution of the pasta box (PB) were modified to simplify the interaction with the prosthesis. Furthermore, the phases of interaction can be clearly divided into *Reach*, *Grasp*, *Transport*, and *Release*, allowing the analysis of single phases of the movement. In contrast to earlier implementations of this test, the segmentation of interaction phases was implemented only based on the interaction with (invisible) colliders in the VR environment:

1) Reach: Started when the *Home* button was pressed.
2) Grasp: Started when the hand first collides with the *grasp collider box* (see Figure 1), which is 70 mm around the pasta box from all sides.
3) Transport: Started when the *grasp collider box* no longer has a collision with the table.
4) Release: Started when the *grasp collider box* has a collision with the next target location (yellow sticker) and ends when the hand leaves the *grasp collider box*.

During the Release phase, the loss of contact was assessed to determine whether the pasta box (PB) was successfully placed. Placement (PB fully inside target, PB touching only the target, or PB outside the target) and orientation (up, sideways, upside down) were derived from collision boundaries and the angle between the global y-axis and the PB's local y-axis. As this study did not primarily assess functional myocontrol performance, we defined a *semi-success* as grasping and transporting the PB to the target location without a hand-disconnection/drop. The final placement criterion was relaxed such that touching the target and not being sideways was sufficient. After each movement ended (independent of success or failure), the PB turned grayscale to indicate it should no longer be manipulated; the Home button became active and, when pressed, respawned the PB at the next start position in the correct orientation, after which the Home button returned to grayscale.

Based on the position and orientation of the VR tracker in the real world, the position of the virtual prosthesis was adjusted inside the VR environment. To minimize the influence of the body representation inside VR, we decided only to render the robotic hand and forearm.

In contrast to previous implementations of this test, it was developed using the physical simulation engine MuJoCo [17] inside Unity. Therefore, the robotic hand used is fully simulated at each timestep. This was chosen to increase the realism in robotic interaction in VR. The robotic hand used in this transradial prosthesis scenario was the Shadow Hand, for which a MuJoCo version was available [62] and then adapted to the task. The adaptation included a MuJoCo weld connection to the position of the Vive tracker, rescaling to an anthropomorphic size, and adjusting the ranges of motion for the degrees of freedom (DOFs).[1] The VR environment was created using Unity [18].

3) Signal Processing and Myocontrol: The signal processing and myocontrol were consistent with the steps of the position control in [63], limited to two actuated DOFs, i.e., hand-opening-closing (HOC) and wrist-flexion-extension (WFE). The streamed signals from the sEMG sensors were (i) amplified and rectified, (ii) lowpass filtered (2nd order Butterworth [64], cutoff frequency $f_c = 0.6$ Hz, (ii) subsampled at 25 Hz. Afterwards, the Ridge Regression [65] was trained with four control actions (rest, hand-closing, wrist-flexion, and wrist-extension) in four rounds of training and two different poses (elbow bent at approximately 90 deg and arm extended). Each training repetition consisted of a 3 s sEMG data recording, in which the participant was asked to perform a medium (60 −70 % maximum force) contraction of the specified action. The actuation of the DOFs θ_{HOC} and θ_{WFE} was computed with the corrected regressor predictions $\bar{y}_C, \bar{y}_E, \bar{y}_F$ (see [66]) for hand closing, wrist extension, and flexion with

$$\theta_{HOC} = \bar{y}_C$$
$$\theta_{WFE} = \frac{\bar{y}_E - \bar{y}_F + 1}{2} \quad (1)$$

The communication between Unity and the signal processing suite was implemented through a User Datagram Protocol (UDP) stream using the Streamlined Input Manager (SIM) [67].

4) Myocontrol Manipulation: To test our approach of assessing psychological factors through biosignals caused by interaction with our environment, we first had to induce different levels of the same psychological factors. To this end, the myocontrol was manipulated with two different additional delays and three levels of artificial malfunction, resulting in six conditions for the within-subject study (Figure 1) The additional delay of the prosthesis control was either *None* [ND] or *300 ms* [D]. In accordance with [40] and [68], this was assumed to be in a range where it would have a measurable impact. This delay was in addition to the inherent control delay. Control delay was defined as the latency between measured EMG activity and the virtual prosthesis reaching the new target position. The dominant contribution arose from causal EMG envelope filtering; its effective delay was estimated via spectrally weighted group-delay analysis and time-domain alignment of causal and zero-phase filtering, yielding ca. 370 ms. Downstream scheduling (envelope subsampling and Unity update timing) contributed ca. 25 ms, while MuJoCo actuator dynamics added ca. 50 ms, depending on actuation step size. Thus, the total externally imposed control delay was approximately ca. 445 ms in baseline conditions and ca. 745 ms when the experimentally induced delay was applied. Importantly, this describes the latency until a new target position is reached, not just movement onset. This latency is reduced by the advance of the EMG signal before force/movement onset (electromechanical delay) by 30 −100 ms [69]. Additionally, the usage of isometric contractions reduced the mechanical travel time to reach a certain force level. We consider the latency caused by other processes to be negligible in this context.

The artificial malfunctions were categorized into three levels: none, medium, and high. In this context, 'none' meant no malfunctions, medium meant 10/60 movements, and high meant 20/60 movements had planned malfunctions. These malfunctions were planned for the transport phases in each condition.

To induce malfunctions that resemble intermittent myocontrol failures, we injected a brief, externally imposed opening component during Transport at pre-randomized onset positions within the transport trajectory. The injected disturbance was smoothly blended in and out and was tuned to reliably reduce controllability and cause object drops despite ongoing user input.

B. Measurements and Feature Calculations

1) Questionnaires: In the context of our study, we use the following definitions for Ownership, Agency, and Trust. Ownership refers to the feeling that a body (part) belongs to oneself, is one's own, and is the source of all sensory stimulation [68], [70], [71]. Whereas, Agency refers to the feeling of perceiving (motor) control over a body and, through that, control over events in the environment [68], [72], [73]. Trust, we refer to as "the attitude that an agent will help achieve an individual's goals in a situation characterized

[1] A video highlighting the capabilities of the environment can be found here: https://doi.org/10.5281/zenodo.16933419

by uncertainty and vulnerability" [10, p.54] consistent with standard models from automation and human–robot interaction literature [74]. The subjective measures used in this work are questionnaires selected for their relevance to our research questions, validation, and widespread use. We used two questionnaires: *Ownership* and *Agency* were captured via the body ownership and agency sub-scales of the Virtual Embodiment Questionnaire (VEQ) [68]. The subscales can be used on their own. We did not capture the third dimension "Change", as it relates more to full virtual avatars as opposed to prostheses. Although originally validated under visual latency manipulations [68], the VEQ has since been applied in contexts involving other embodiment mechanisms, including studies testing visuo-tactile feedback (e.g., [75]). *Trust* was measured via the Trust in Automated Systems scale [74]. Although the scale did not originally consider prosthesis, we believe it to be general and adaptable enough to fit the use case. Compared to other scales evaluating trust in automated systems and robotics, it does not overly anthropomorphize the prosthesis and force participants into inconsistent interpretations of attributed ascribed [76]. Both scales were slightly adjusted in their wording to refer to the prosthesis as the object of relevance, and were used otherwise unchanged. Compared to previous work, we specifically adapted and used established instruments in place of using custom and non-validated questions.

2) Eye-Tracking: Eye-tracking analyses were based on fixations to three predefined regions of interest (ROIs): hand, object, and target. The hand ROI refers to the colliders of the hand and wrist parts of the virtual prosthetic hand. The object ROI referred exclusively to the pasta box colliders. The target ROI denoted task-relevant target locations and was phase-dependent, comprising the pickup location during *Reach* and *Grasp*, and the drop-off location during *Transport* and *Release*. These definitions were applied consistently across all eye-tracking metrics, except for TLS. Here, the object is assigned to the hand or the target ROI, depending on the phase (see below). A fixation on an ROI was defined when a spherical gaze ray (diameter d_r = 5 cm) intersected the corresponding ROI collider. Invalid samples were treated as "no fixation", and brief gaps (‘!100 ms) were bridged to account for blinks [21]. Fixations shorter than 120 ms were discarded as they are unlikely to reflect information processing [77]. Multiple ROIs could be fixated simultaneously (e.g., spatial overlap); in such cases, overlapping fixations within the same ROI were merged for fixation-percent computation.

a) Fixation percent: The fixation percent describes the duration during which an ROI was fixated (processed as described above) in a phase of a movement, normalized by the duration of the phase [77].

b) Number of fixations: The number of fixations (NOF) describes the number of times any ROI was newly fixated (processed as described above) [21].

c) Target-locking strategy (TLS): This metric is defined as the difference between fixation percent on the target ROI and fixation percent on the hand ROI [77], [78]. A more positive score indicates more time spent looking at the target than the hand, which is theoretically associated with greater confidence in the control of the prosthesis. For TLS only, the object ROI (i.e. the PB) is grouped with the target ROI during Reach/Grasp, and grouped with the hand ROI during Transport/Release. When fixations on multiple objects belonging to the same ROI (i.e., hand or target) temporally overlapped, these fixations were merged and treated as a single fixation for the purpose of computing the TLS.

d) Eye-arrival latency (EAL) and eye-leaving latency (ELL): These metrics measure the extent to which the fixations precede or lag behind the actions at the beginning and at the end of the Transport phase. Our implementation is based on the definitions by Lavoie et al. [21]. For the EAL at pickup and the ELL at drop-off, we considered the first fixation on the pickup or last fixation on the drop-off location and PB, respectively. The EAL at drop-off used the first fixation on the drop-off site. EAL calculations utilized fixations during *Reach*, *Grasp*, and *Transport*. Analogously, the phases *Transport* and *Release* were considered for drop-off.

C. Experimental Protocol

1) Study Design: The repeated-measures experiment included six (2 × 3) conditions, crossing two levels of delay with three levels of artificial malfunction frequency. The conditions are denoted as follows:

1) ND-NM: No Delay–No Malfunction
2) ND-MM: No Delay–Medium Malfunction
3) ND-HM: No Delay–High Malfunction
4) D-NM: 300 ms Delay–No Malfunction
5) D-MM: 300 ms Delay–Medium Malfunction
6) D-HM: 300 ms Delay–High Malfunction

The order of conditions for each participant was generated using the balanced Latin square method to minimize carry-over effects, and therefore, repeating after six participants. All participants performed all conditions.

The specific round and movement in which artificial malfunctions occurred were pre-randomized prior to data collection, such that the 10 (medium malfunction) or 20 (high malfunction) events were assigned to unique round–movement combinations for all participants. For each of these events, the onset time of the malfunction was additionally pre-randomized as a relative position r_i within the corresponding *Transport* phase, expressed as a fraction of the transport distance. Thus, malfunctions occurred only during transport phases selected by the pre-randomization and at different, unpredictable positions within those phases (Section III-A.4). To reduce carry-over effects and reinforce the impression of different prostheses across conditions, prosthesis color was randomized for each condition.

2) Study Procedure: Participants were informed about the study procedure, provided informed consent, and completed the participant information and data privacy form. To limit the influence of the study content on the participants' behavior, the participants were told they would test six different prostheses in VR and evaluate their functionality. The omission of the study purpose was fully disclosed at the end of the session. Participants then donned the two Myos and the hand–wrist orthosis. The orientation and position of the Myos were standardized across participants, and the VR tracker was fastened around or clipped onto the Myo closer to the elbow. Next, the myocontrol model was trained on participant-specific sEMG data in four rounds (Section III-A.3). A brief pre-study questionnaire assessed prior VR experience and current physical condition (with a focus on nausea and discomfort); the physical-condition question was repeated at the end of the study.

Participants then wore the VR HMD, completed eye-tracking calibration, and familiarized themselves with the PBT environment (2 min exploration and 2 min free control practice; Figure 1). At the start of each condition, eye tracking was checked and calibrated if necessary. Each participant completed all six within-subject conditions; each condition comprised 20 rounds with three movements per round (Figure 1). After each movement (successful placement during *Release* or a drop during *Grasp*, *Transport*, or *Release*), participants pressed the Home button to respawn the pasta box for the next movement. After completing all rounds of a condition, participants removed the HMD and completed questionnaires on *ownership*, *agency*, and *trust*. In rare cases of instruction non-adherence or VR-environment issues, the affected round was restarted.

3) Participants and Exclusion: The study was conducted with 36 non-disabled participants, six of whom (one for each condition order) were regarded as spares in case of exclusions. Participants were excluded from the study if their myocontrol performance was compromised by more than 25% of movements where the object disconnected from their hand during interaction, despite no artificial malfunction being induced. Additionally, participants were excluded if their eye-tracking validity fell below 80% in a substantial number of rounds, rendering the condition unusable. One participant was excluded due to low eye-tracking validity, and two were excluded due to low myocontrol performance, suggesting that their control was not functioning properly. These participants were replaced by spares with the same condition order. Therefore, the total number of participants analyzed was reduced to 30 to achieve a balanced number of condition orders. If a problem with the developed system was encountered, the specific round was repeated immediately, and the second attempt was included in the analysis. The included participants consisted of 18 male and 12 female with age (mean ± SD) 28.86 ± 5, 14 of which 25 were right-, 3 were left- and 2 both-handed. The experiments were conducted at FAU Erlangen-Nürnberg, approved by the university's ethics committee prior to the start of the study (No.: 22-131-S) and conducted in accordance with the Declaration of Helsinki.

D. Data and Statistical Analysis

The results of the pasta box experiment were calculated for each movement. The percentage of *semi-success* (see Section III-A.2) was calculated as a score per condition for each participant. Durations were aggregated for each participant and condition, considering the movements with semi-success. Fixations were aggregated as the mean for each phase and movement within a condition of a participant. For fixations, only movements with semi-success and without hand-object-disconnects in the second half of the condition were considered to minimize the influence of the actual malfunction and to account for the time needed to adjust to the new setting. Since a large variability was found between fixations in the transport phase, depending on which movement was performed, fixations were subtracted by the overall movement's mean before aggregation. Questionnaires were analyzed as defined in the instructions. No further aggregation was necessary as they were given out once per condition.

Before any statistical tests were conducted, outliers were capped with the interquartile range (IQR) method. Data normality was evaluated with the Shapiro–Wilk test. If normality was confirmed, a two-way repeated-measures ANOVA (rmANOVA) on the factors delay and malfunction was conducted as an Omnibus test, accounting for sphericity violations with the Greenhouse-Geisser correction (denoted with p_{GG} in applicable cases). If the rmANOVA yielded a significant influence of a factor, paired t-tests with Benjamini-Hochberg correction (denoted with p_{BH}) were performed as a post-hoc test. If data normality was not confirmed, a Friedman ANOVA was applied to each factor, and in case a significance was found, a pairwise Wilcoxon signed-rank test was conducted. Differences were considered statistically significant for p-values below the threshold of $\alpha = 0.05$. Furthermore, to account for the high variability between movements and participants before calculating correlations, the fixations on the hand and object during the transport phase were robustly z-normalized by their median and IQR.

The VR environment, gaze analysis code, and dataset with EMG, kinematic, and gaze data are available here.[2]

IV. Results

In this section, we present the overall results of the study, as well as the influence of artificial malfunctions and delays on the questionnaire results and gaze measurements.

A. Pasta Box Task Results

Because artificial malfunctions were scheduled at pre-randomized round-movement combinations, they could not be executed if a movement terminated early due to an unsuccessful user interaction. Accordingly, Figure 2A shows the inter-participant variability in the number of executed artificial malfunctions within each condition. To account for variability in the number of artificial malfunctions within each condition, we additionally ran a fixed-effects regression for all reported results, treating malfunction count as a continuous predictor. The results were consistent with the reported rmANOVA.

Across study order, task performance and timing showed a learning trend, with semi-success increasing and movement durations decreasing over repeated trials (Figure 2C–D). Despite this overall improvement, participants differed markedly in their movement-time variability, with pronounced inter-individual differences in duration variability (Figure 2E).

[2]The VR environment, gaze analysis code, and dataset are available on Zenodo

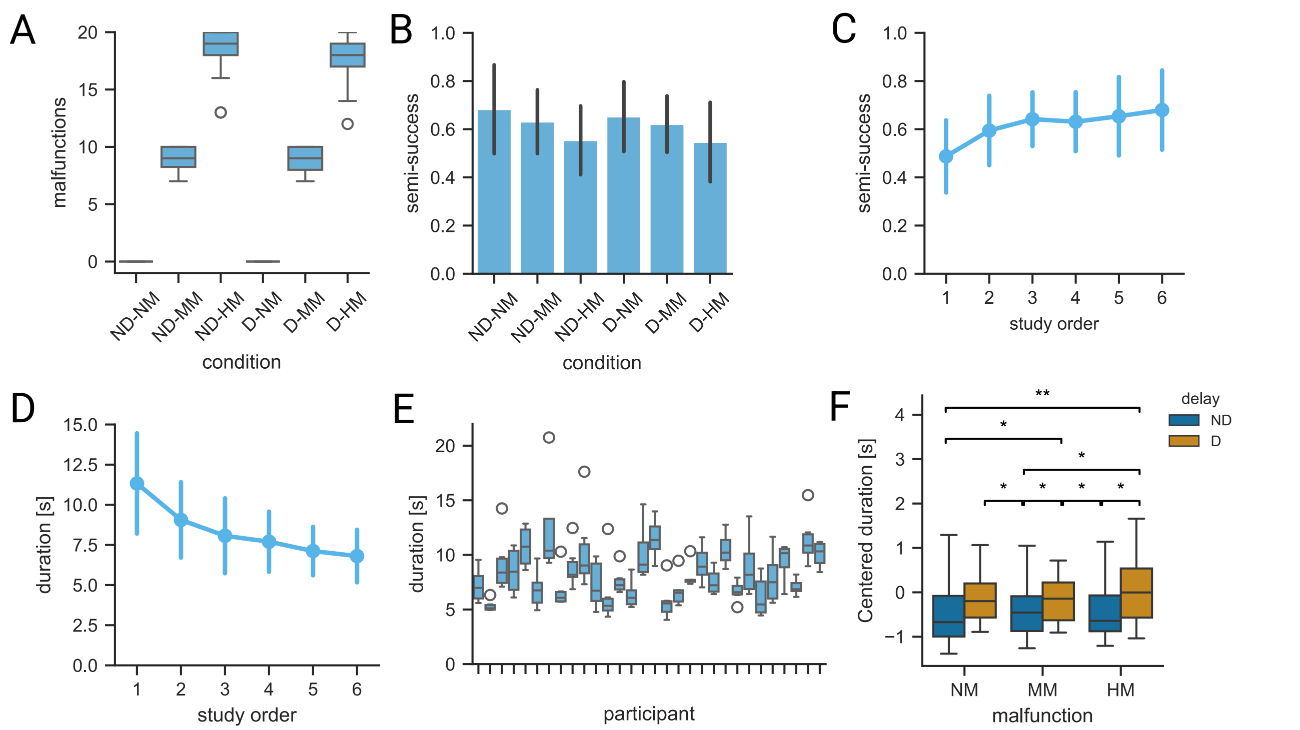

Fig. 2. **Pasta box and duration results -** Panel A displays the distribution of artificial malfunctions that were triggered by the system per person and condition as a boxplot with median, IQR, whiskers, and outliers. These artificial malfunctions do not include any unsuccessful user interactions. Panel B shows the semi-success rate per condition in barplots, mean and SD errorbars. Panels C and D display the change in semi-success and duration depending on the study order in movements where there was no artificial malfunction, as mean and SD, indicating a learning effect across the experiment. Panel E shows the movement durations per person, illustrating the high individual variability. Panel F shows the centered duration (subtracted by the respective movement's mean duration) needed during the *Grasp* phase, depending on the condition, as a median, IQR, and whiskers. Outliers in panel F are capped. Asterisks indicate the significance levels ($*$: p_{BH} < .05, $**$: p_{BH} < .01) from the post-hoc Wilcoxon signed rank test. All variables were aggregated by participant and condition. *Abbreviations:* ND = no delay; D = 300 ms delay; NM/MM/HM = no/medium/high malfunction; IQR = interquartile range; SD = standard deviation; p_{BH} = Benjamini–Hochberg-corrected p-value.

Condition-wise effects on semi-success and phase durations are summarized in Figure 2B,F.

We also investigated how the durations of the phases change depending on the condition. In the phase *Grasp*, the Friedman ANOVA showed a significant influence of the factor delay ($\chi^2(1) = 8.53, p_{unc} = .004, W = .28$) with a significant difference between ND-NM and D-HM ($p_{BH} = 0.016$) in the post-hoc Analysis. The results from the post-hoc Wilcoxon signed rank tests can be seen in Figure 2F, indicating multiple significant differences between delay conditions but not within a delay group. Furthermore, in the phase *Transport*, the Friedman ANOVA indicated a significant influence of the factor malfunction ($\chi^2(2) = 14.47, p_{unc} < .001, W = .24$).

B. Questionnaires

Four participants stated that they had no experience with VR, 20 participants had a little experience (less than five times), and six participants reported occasional use. The pre- and post-questionnaires on the physical condition of the participant indicated an increase in discomfort from 10.3 ± 15.9 to 14.5 ± 17.0 on a scale $[0, 100]$. A Wilcoxon signed rank test indicated no significant difference between measurements ($W = 67.0, p = .095, r = -0.42$).

Next, we analyzed the results of the subjective questions to determine whether the expected effect of delay and malfunctions was evident. Figure 3 depicts the results of the ownership, agency, and trust questions. Across all three measures, scores tended to be highest in the ND–NM condition and progressively lower with increasing level of malfunction (except for agency and trust in condition D-MM). Delay conditions generally showed reduced scores compared to their counterpart. The lowest median scores in all questionnaires were observed in the D-HM condition. The participant answered the questionnaire after each condition; therefore, no further aggregation steps were necessary. Two conditions of the ownership questionnaire showed minor normality violations, which the rmANOVA was deemed robust against. The conducted rmANOVAs confirmed a significant influence of the factor malfunction on the ownership ($F_{2,58} = 9.65, p_{GG} < .001, \eta^2 = .041$), agency ($F_{2,58} = 9.86, p_{GG} < .001, \eta^2 = .055$), and trust ($F_{2,58} = 9.41, p < .001, \eta^2 = .051$). Furthermore, the factor delay also had a significant influence on the ownership ($F_{1,29} = 7.07, p = .013, \eta^2 = .015$), agency ($F_{1,29} = 8.46, p = .007, \eta^2 = .041$), and trust ($F_{1,29} = 5.09, p = .032, \eta^2 = .017$). The post-hoc paired t-tests with applied Benjamini-Hochberg (BH) correction revealed the significant differences depicted in Figure 3. The interaction between the factors malfunction and delay had no significant influence on the subjective scores.

C. Eye-Tracking and Correlations

The eye tracking results comprise the analysis of ROI fixations, NOF, TLS, and EAL/ELLs. To minimize the effect

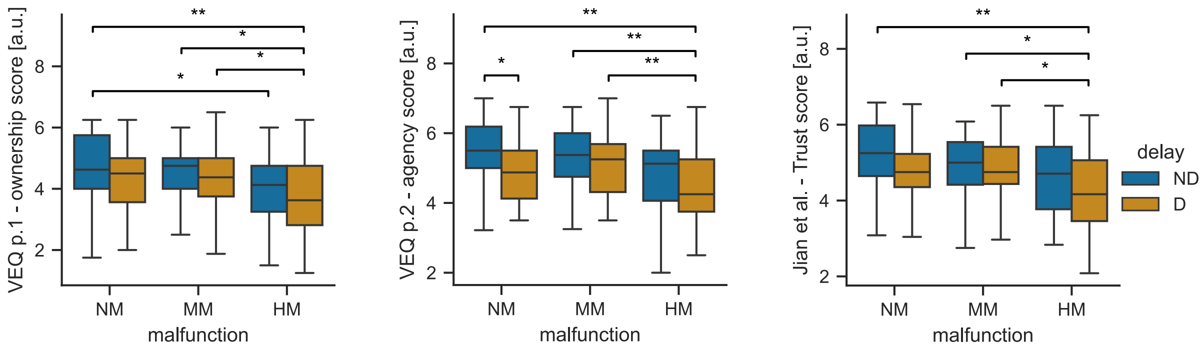

Fig. 3. **Questionnaire results -** Panel A shows boxplots with medians and IQRs of the ownership, agency, and trust scores for each condition. Asterisks illustrate the significance levels ($*: p_{BH} < .05, ** : p_{BH} < .01$) from the post-hoc paired t-test. Abbreviations: VEQ = Virtual Embodiment Questionnaire; NM/MM/HM = no/medium/high malfunction; ND/D = no delay/delay; IQR = interquartile range; p_{BH} = Benjamini–Hochberg-corrected p-value.

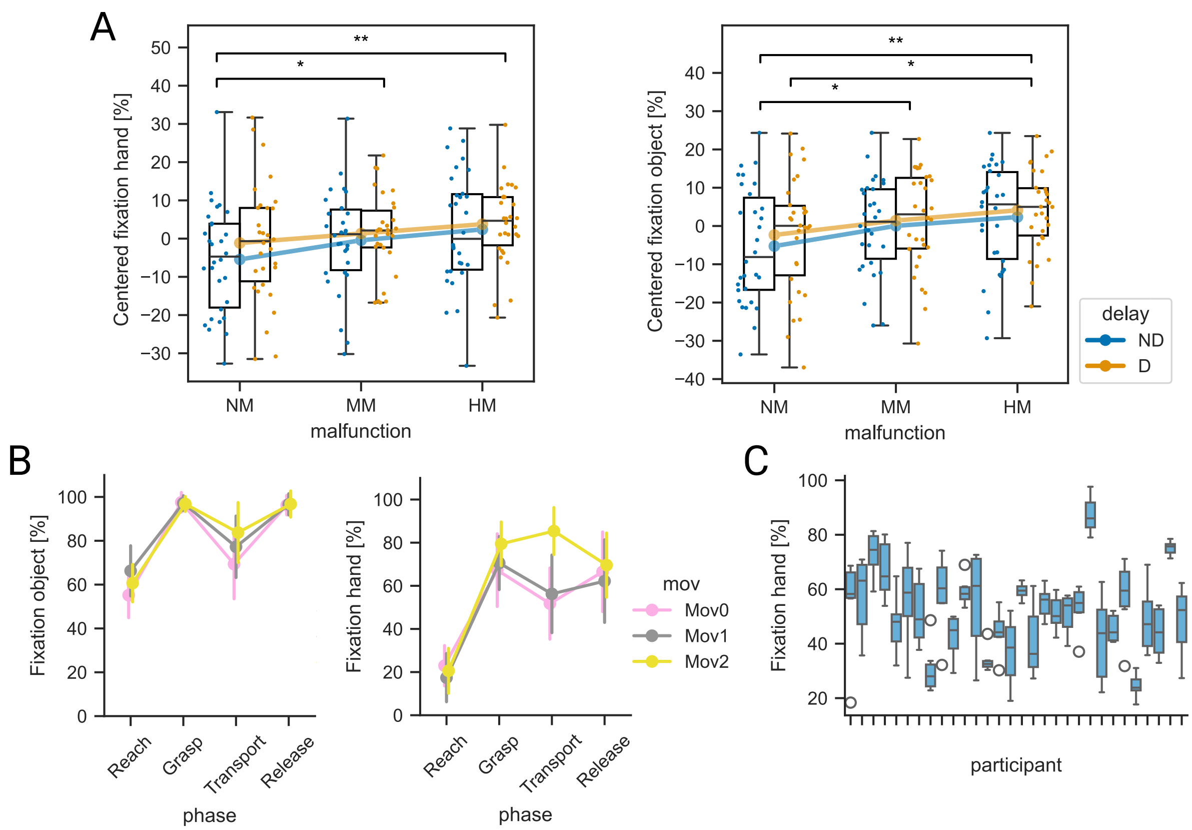

Fig. 4. **Region of Interest (ROI) fixation results -** The figure compares hand and object ROI fixations across conditions and participants. Panel A shows the centered fixation percent on hand ROI (left) and object ROI (right) over conditions as boxplots with median and IQR, single participant values, and overlaid mean. Asterisks illustrate the significance levels ($* : p_{BH} < .05, ** : p_{BH} < .01$) from the post-hoc paired t-test. Panel B illustrates the ROI fixation percent in the different movements and phases of the Pasta Box Task with mean and SD. Panel C shows the differences in hand ROI fixation per participant. Abbreviations: ND = no delay; D = 300 ms delay; NM/MM/HM = no/medium/high malfunction; Mov0/1/2 = first/second/third movement; IQR = interquartile range; SD = standard deviation; p_{BH} = Benjamini–Hochberg-corrected p-value.

of the acute malfunction on the gaze, only measurements from semi-successful movements without artificial malfunctions from the second half of the condition were processed as defined in Section III-D. Figure 4A shows the different overall fixation on the hand and object ROI per movement phase, while Figure 4B and C illustrate the differences in ROI fixation between movements, phases, and participants. Interestingly, the ROI fixations behave very similarly for the different movements and phases, except for the *Mov2 Transport* phase. During *Reach* participants generally focused more on the object (60.71 ± 11.12 %) than the hand (20.33 ± 10.57 %). The same applied to the next phases *Grasp* (hand: 72.44 ± 14.34 %, object: 97.16 ± 3.82 %), *Transport* (hand: 64.47 ± 21.46 %, object: 76.79 ± 15.69 %) and *Release* (hand: 66.09 ± 17.84 %, object: 96.53 ± 5.14 %). The ROI fixations showed a very different behavior between people, visible in Figure 4C. Because of the large differences in ROI fixations between

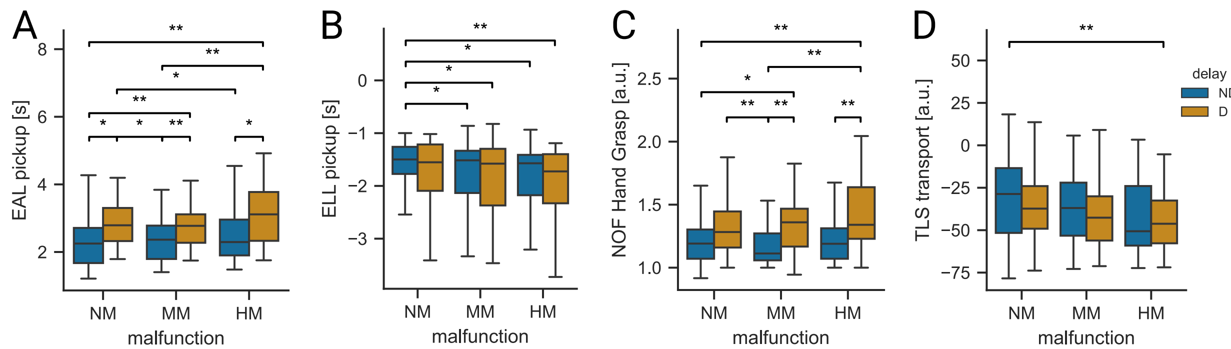

Fig. 5. **Other gaze feature results -** This panel shows a selection of other gaze features with significant results illustrated over conditions as boxplots with median and IQR. Panel A displays the eye-arrival latency during pickup, Panel B the eye-leaving latency during pickup, Panel C the number of fixations on the hand ROI during *Grasp*, and Panel D the target-locking strategy during *Transport*. Asterisks illustrate the significance levels ($*$: $p_{BH} < .05$, $**$: $p_{BH} < .01$) from the post-hoc paired t-test or Wilcoxon signed rank tests. Abbreviations: ND = no delay; D = 300 ms delay; NM/MM/HM = no/medium/high malfunction; ROI = region of interest; EAL = eye-arrival latency; ELL = eye-leaving latency; NOF = number of fixations; TLS = target-locking strategy; IQR = interquartile range; p_{BH} = Benjamini–Hochberg-corrected p-value.

movements in the *Transport* phase, we centered all ROI fixations by their movement's mean before further aggregating over movements. Therefore, Figure 4A shows the fixation percent relative to the movement's mean. The conducted rmANOVA confirmed that the factor malfunction had a significant influence on the hand ROI fixation ($F_{2,58} = 8.12, p_{GG} < .001, \eta^2 = .038$) and object ROI fixation ($F_{2,58} = 15.73, p_{GG} < .001, \eta^2 = .046$). The factor delay had just a significant influence on the hand ROI fixation ($F_{1,29} = 4.19, p = .050, \eta^2 = .009$) and no significant influence on the object ROI fixation ($F_{1,29} = 2.88, p = .10, \eta^2 = .006$). The results of the post-hoc paired t-tests with BH corrections are visible in Figure 4A. Figure 5 depicts the behavior of the other fixation-based features with significant omnibus tests and their respective post-hoc test results. TLS results during transport (Figure 5D) indicated a clear downward trend across conditions. With minor normality violation, a rmANOVA was conducted, confirming that malfunction had a significant influence ($F_{2,58} = 11.76, p_{GG} < .001, \eta^2 = .032$), but delay did not ($F_{1,29} = 2.34, p = .137, \eta^2 = .004$). For the EAL during pickup, due to non-normality, two separate Friedman tests were conducted, each assessing one of the factors independently. The Friedman test showed a statistically significant influence of the factor delay ($\chi^2(1) = 10.78, p_{unc} = .001, W = 0.36$), but not the factor malfunction ($\chi^2(2) = 4.07, p_{unc} = .131, W = 0.07$) Interestingly, the ELL during pickup was only significantly influenced by the factor malfunction ($\chi^2(2) = 15.20, p_{unc} < .001, W = 0.25$) and not delay ($\chi^2(1) = 1.20, p_{unc} = .273, W = .04$). The Friedman ANOVAs for the NOF during *Grasp* yielded a significant influence of the factor delay on the number of hand ($\chi^2(1) = 19.2, p_{unc} < 0.001, W = 0.64$) and object fixations ($\chi^2(1) = 8.53, p_{unc} = .003, W = 0.28$). In contrast, the factor malfunction had no significant influence on the number of hand ($\chi^2(2) = 2.64, p_{unc} = .267, W = 0.04$) and object fixations ($\chi^2(2) = 1.69, p_{unc} = .430, W = 0.03$).

Preliminary analysis indicated that other phases of ROI fixations, EAL, ELL, NOF, and TLS were not affected by the experimental conditions. Given the lack of variation in these variables, they were not included in subsequent statistical analyses.

TABLE I
CORRELATION TABLE BETWEEN FIXATION PERCENT DURING TRANSPORT AND PSYCHOMETRIC SCORES

	Hand ROI fixation	Object ROI fixation
Ownership	$\rho = -0.228, p = .002$	$\rho = -0.224, p = .003$
Agency	$\rho = -0.252, p = .001$	$\rho = -0.217, p = .003$
Trust	$\rho = -0.228, p = .002$	$\rho = -0.230, p = .002$

Furthermore, the relationship between the ROI fixations during transport and subjective scores was investigated. Due to the large differences in fixation depending on the movement and participant, the ROI data were further processed as described in Section III-D. Although the ROI fixations were robust-z transformed to account for the high variance between participants and movements, there is still only a weak correlation found with the subjective scores, see Table I.

V. DISCUSSION

This study primarily investigates the effects of control manipulations, artificial malfunctions, and delays on ownership, agency, trust, and gaze behavior in a fully immersive prosthesis environment simulated using MuJoCo within VR. Our results confirmed that both malfunctions and delays have significant effects on psychometric measurements and visuomotor behavior.

A. Pasta Box Task Results

Trends in semi-success, phase durations, and gaze indicated that the environment supported systematic and repeatable measurement. However, the number of executed malfunctions varied across conditions because scheduled events could be skipped when movements terminated early (Figure 2A). Semi-success differed across conditions (Figure 2B) and might have diverged further had all scheduled malfunctions been executed, although participants sometimes recovered via compensatory movements. Nonetheless, malfunctions still reduced controllability irrespective of the final outcome. A learning trend was visible over the ca. 90 min protocol (Figure 2C-D). Discomfort increased slightly but with high variability, suggesting minimal and inconsistent physical strain or VR-related discomfort. Grasp duration was affected by delay but not malfunction (Figure 2F), consistent with the imposed latency, whereas Transport duration was affected by malfunction, plausibly reflecting corrective or hesitant behavior under reduced reliability. Unlike in [79], EAL/ELL at pickup were determined based on the start of Transport and at drop-off based on the end of Transport, meaning that the values are not directly comparable.

B. Ownership, Agency and Trust

Both delay and malfunction significantly reduced ownership, agency, and trust compared to the baseline condition. Malfunction had a similar effect on ownership, agency, and trust. Agency was the only measure influenced to a similar extent by both factors, whereas ownership and trust were more strongly affected by malfunction. Compared to the findings reported in previous works investigating avatar and modified avatar embodiment, our ownership and agency scores showed a close alignment [80]. Specifically, the mean and standard deviation observed in our study fall within the same range, suggesting consistency and comparability despite differences in sample composition and virtual body. Compared to another work investigating the effect of latency on ownership and agency, we found smaller effect sizes [68].

The smaller delay-related effect sizes compared to [68] may stem from our higher baseline control latency. In prosthetics, controller delay is often reported as EMG-to-movement-onset latency, with delays below ~100–200 ms generally tolerated and performance degradation with increasing values [81]. In immersive VR with full-body avatars, ownership and agency begin to decline at delays above ~125 ms and deteriorate further beyond ~300 ms [41], [81], whereas in a hand illusion, the critical window below ~200 ms and no illusion at 590 ms were observed [42]. Within this broader context, our measured total control delays (ca. 445 ms baseline and ca. 745 ms with added delay) may appear large relative to the values cited above. In our position-control pipeline, the reported delays quantify latency until the prosthesis reaches a new target position, not merely movement onset, and include strong smoothing required for stable (non-jittery) position commands. Alternative pipelines with lower delay could be achieved by using a different filter type (e.g., adaptive filters) and controllers that are more robust to input noise (e.g., velocity control). It remains unclear how the position control paradigm influences the user's perceived delay. However, perceived sensorimotor delay is likely shorter because EMG precedes force/movement onset by 30–100 ms [69], and isometric contractions reduce mechanical travel time. Finally, only distal DOFs (hand/wrist) were delayed, while first-person viewpoint and forearm tracking remained fully synchronized.

Similar to [68], delay influenced both ownership and agency, which is plausibly explained by their strong correlation reported in prior embodiment work. The more substantial effect of delay on agency suggests that sustained sensorimotor incongruence primarily diminishes a user's sense of control, whereas the unpredictability of malfunctions disrupts ownership, agency, and trust to a more comparable extent. Trust followed a similar degradation pattern to agency and ownership and showed significant influences of both delay and malfunction, consistent with expectations from robotic automation literature that changes in control reliability shape trust levels [9], [10], [14].

We further note that ownership in our setup was supported through multiple pathways: (1) visuo-motor synchrony established through the continuous Shadow Hand forearm control, (2) visuo-proprioceptive synchrony arising from muscle force cues during isometric contractions, and (3) consistent first-person visual alignment between the virtual and physical forearm. While visuo-tactile feedback was absent by design, cutaneous pressure from finger contact with the orthosis provided additional force-related input. We acknowledge that this limits transferability to people with transradial amputation.

C. Gaze Behavior and Correlations

The eye-tracking results confirmed that manipulating the myocontrol significantly altered visual attention patterns. Delay significantly increased fixation on the prosthetic hand, whereas malfunction increased fixation to both the prosthetic hand and object during transport, aligning with prior findings by Cheng et al. [23] and Chadwell et al. [47], indicating a visual compensatory behavior in response to the reduced system predictability or functionality. Interestingly, the other significant fixation features, i.e., EAL, ELL, NOF, and TLS, were either significantly influenced by either delay or malfunction. Similar to the phase duration, the EAL during pickup was significantly influenced by the delay. This again could be due to the actual delay in the grasping represented in this difference. Additionally, delay had a significant influence on the NOF on the hand (Figure 5C) and object, while malfunction did not. Interestingly, in [47], the number of gaze switches was correlated with undesired activation and electromechanical delay, of which only the latter could be found in our study. On the other hand, both ELL during pickup (Figure 5B) and TLS in the transport phase (Figure 5D) were only significantly influenced by the factor malfunction. This suggests that some of the time-dependent visual-motor features during *Grasp*, such as EAL at pickup (longer grasping results in longer EAL) and NOF (longer grasping results in a higher NOF), simply represented the induced delay. Whereas, features which are based on fixations during *Transport*, such as the fixation percent, ELL at pickup, and TLS, are influenced by the factor malfunctions, or both factors.

While correlations between ROI fixations and subjective scores were statistically significant, they were quite weak ($\rho \approx -0.2$) in strength. This could be, among other factors, due to the high variability between participants visible in Figure 4. Nevertheless, in conditions where participants looked more at the prosthesis or object, they felt less ownership, agency, and trust.

A limitation is that gaze strategies in controller-mediated VR can differ from real-world interaction [79]. Potentially, this is due to the lack of haptic feedback. However, reduced haptic feedback in VR may partially resemble prosthesis-use constraints [27]. Future work should directly replicate key conditions with users with limb loss to quantify the validity.

D. VR Environment and Dataset

We developed a physics-based, fully immersive VR prosthesis simulation (Unity with MuJoCo) with integrated eye tracking that enables controlled manipulation of delay and malfunction parameters while capturing synchronized psychometric, gaze, and task-performance responses. While previous research suggests that a high level of realism inside VR leads to better skill transfer in a real prosthesis setting [49], [50], [52], [53], [54], the comparability of psychometric scores between VR and real-world prosthesis use remains to be validated. Our contribution extends earlier prosthesis simulations using MuJoCo [56], [57] by embedding it into a task-oriented VR setup (PBT) with simultaneous eye tracking and subjective measurement, enabling reproducible reliability manipulations. To support replication and follow-up work, the VR environment, analysis code, and dataset are made openly available. The accompanying multimodal dataset (EMG, kinematics, gaze) can further be used to study signatures of control breakdowns and to develop detection methods that could trigger model updates or shared-autonomy interventions during myoelectric control.

E. Limitations

The study presents several limitations. First, the artificial malfunction implementation, although randomized in timing and behavior, only led to malfunctions in HOC and did not simulate errors in WFE. Second, although great care was taken to ensure malfunctions appeared natural, their perceived realism was not directly assessed. In our study, we utilized the robotic Shadow Hand, which is not an actual prosthetic hand. Future VR studies should consider using MuJoCo models of actual prostheses to allow for a closer comparison. Furthermore, while the MuJoCo plugin provided a highly realistic interaction, it still lacks the haptic and sensory fidelity of physical prosthetic use, limiting the transferability of our findings. Additionally, the impact of EAL/ELL at drop-off and TLS may have been diminished due to the limited size of the target locations and interaction with the gaze raycast. This requires methodological refinement in the next development iteration. Furthermore, the color-coding of prosthesis conditions in our study (to simulate a *new* prosthesis each time) aimed to disrupt carry-over trust biases, but it is unclear to what extent this worked.

Our interpretation of ownership outcomes remains cautious, acknowledging the current validation scope of the VEQ and the need for future studies incorporating ownership-specific manipulations, such as visuo-tactile mismatches. However, alternative validated instruments for measuring embodiment are limited [37], and the VEQ has gained broad acceptance in the VR and HRI communities [82], [83]. Recent applications have demonstrated its responsiveness to visuo-tactile synchrony (e.g., [75]).

Finally, our design aimed to take a step towards approximating characteristics of a transradial amputation using an orthosis for wrist and finger immobilization, i.e., isometric contractions. However, the presence of cutaneous feedback, particularly from finger pressure on the orthosis, as well as many other confounding factors, limits direct translation. Additionally, Zbinden et al. [39] caution that embodiment paradigms developed for non-disabled individuals may have limited validity for people with amputation due to differences in sensory input and body representation. Furthermore, it is unclear to what extent results from virtual HRI studies compare to real-world scenarios [84]. Future studies should investigate embodiment and gaze responses in and outside VR among participants who have undergone transradial amputation to further assess clinical generalizability.

VI. Conclusion

This study investigated the effects of induced delays and artificial malfunctions on ownership, agency, trust, and gaze behavior in a fully immersive prosthesis simulation. By manipulating delay and malfunction levels across six conditions, we found that higher malfunction severity consistently reduced ownership, agency, and trust, while added latency primarily diminished agency more than the other psychometric measurements. The largest declines in all three measures occurred under high malfunction combined with delay, whereas the no-delay, no-malfunction (ND-NM) condition yielded the highest scores. Trust, as expected, dropped with malfunction and delay, supporting its role as a user-centered measure of system reliability. Importantly, our eye-tracking analysis complements these subjective findings with objective, task-embedded indicators of how participants adapt their visual monitoring under reduced control reliability. The factors impacted different eye-tracking metrics like hand and object fixation percent, number of fixations, eye-arrival latency, eye-leaving latency, and target-locking strategy separately, highlighting the translational potential of gaze-based measures for future prosthesis evaluation.

To conduct this experiment, we developed a VR environment that enables controlled manipulation of prosthesis behavior while capturing detailed gaze and performance data. This environment will be made openly available to support further research. The dataset collected, including sEMG, kinematic, and gaze signals, offers opportunities for investigating automatic detection of control issues and informing adaptive or shared-autonomy myocontrol strategies.

Acknowledgment

- Figure 1 was created in BioRender by Egle, F. [https://BioRender.com/rc6keex].
- The python framework *Tiny Pipelines for Complex Problems (TPCP)* [85] was used facilitating end-to-end biosignal and study data analysis.
- The Unity package *mj-unity-extensions* from the project *modular-agents* [86] was used inside the PBT environment.

References

[1] A. Marinelli et al., "Active upper limb prostheses: A review on current state and upcoming breakthroughs," *Prog. Biomed. Eng.*, vol. 5, no. 1, Jan. 2022, Art. no. 012001.

[2] E. A. Biddiss and T. T. Chau, “Upper limb prosthesis use and abandonment: A survey of the last 25 years,” *Prosthetics Orthotics Int.*, vol. 31, no. 3, pp. 236–257, 2007.

[3] S. Salminger et al., “Current rates of prosthetic usage in upper-limb amputees – have innovations had an impact on device acceptance?,” *Disability Rehabil.*, vol. 44, no. 14, pp. 1–12, Dec. 2020, doi: 10.1080/09638288.2020.1866684.

[4] J. S. Schofield, C. E. Shell, Z. C. Thumser, D. T. Beckler, R. Nataraj, and P. D. Marasco, “Characterization of the sense of agency over the actions of neural-machine interface-operated prostheses,” *J. Visualized Exp.*, no. 143, p. 58702, Jan. 2019, doi: 10.3791/58702.

[5] T. S. Castañeda, M. Connan, P. Capsi-Morales, P. Beckerle, C. Castellini, and C. Piazza, “Experimental evaluation of the impact of sEMG interfaces in enhancing embodiment of virtual myoelectric prostheses,” *J. NeuroEngineering Rehabil.*, vol. 21, no. 1, p. 57, Apr. 2024.

[6] A. Visconti et al., “Enhancing upper limb prosthetic embodiment through virtual reality with eye-tracking and vibrotactile feedback,” in *Proc. IEEE 3rd Int. Conf. Intell. Reality (ICIR)*, Dec. 2024, pp. 1–8.

[7] L. Jabban, B. W. Metcalfe, J. Raines, D. Zhang, and B. Ainsworth, “Experience of adults with upper-limb difference and their views on sensory feedback for prostheses: A mixed methods study,” *J. NeuroEngineering Rehabil.*, vol. 19, no. 1, p. 80, Jul. 2022.

[8] C. L. McDonald, T. Bhargava, E. G. Halsne, S. M. Van Der Veen, B. J. Hafner, and B. J. Darter, “Perceptions of prosthetic attention among lower limb prosthesis users: A focus group study,” *Disability Rehabil.*, vol. 46, no. 19, pp. 4553–4566, Sep. 2024, doi: 10.1080/09638288.2023.2280066.

[9] H. A. Abbass, J. Scholz, and D. J. Reid, *Foundations of Trusted Autonomy*, vol. 117. Cham, Switzerland: Springer, 2018.

[10] J. D. Lee and K. A. See, “Trust in automation: Designing for appropriate reliance,” *Human Factors: J. Human Factors Ergonom. Soc.*, vol. 46, no. 1, pp. 50–80, Mar. 2004.

[11] B. Barber, *The Logic and Limits of Trust*. New Brunswick, NJ, USA: Rutgers Univ. Press, 1983.

[12] R. C. Mayer, J. H. Davis, and F. D. Schoorman, “An integrative model of organizational trust,” *Acad. Manage. Rev.*, vol. 20, no. 3, pp. 709–734, Jul. 1995.

[13] B. M. Muir and N. Moray, “Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation,” *Ergonomics*, vol. 39, no. 3, pp. 429–460, Mar. 1996.

[14] M. Desai et al., “Effects of changing reliability on trust of robot systems,” in *Proc. 7th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, Mar. 2012, pp. 73–80.

[15] M. A. Abd, I. Gonzalez, C. Ades, M. Nojoumian, and E. D. Engeberg, “Simulated robotic device malfunctions resembling malicious cyberattacks impact human perception of trust, satisfaction, and frustration,” *Int. J. Adv. Robotic Syst.*, vol. 16, no. 5, Sep. 2019, Art. no. 1729881419874962.

[16] A. M. Valevicius et al., “Characterization of normative hand movements during two functional upper limb tasks,” *PLoS ONE*, vol. 13, no. 6, Jun. 2018, Art. no. e0199549.

[17] E. Todorov, T. Erez, and Y. Tassa, “MuJoCo: A physics engine for model-based control,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 5026–5033.

[18] *Unity Game Engine*, Unity Technologies, Unity Software Inc., San Francisco, CA, USA, Mar. 2023.

[19] C. M. Light, P. H. Chappell, and P. J. Kyberd, “Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity,” *Arch. Phys. Med. Rehabil.*, vol. 83, no. 6, pp. 776–783, Jun. 2002.

[20] L. Hermansson, A. Fisher, B. Bernspång, and A.-C. Eliasson, “Assessment of capacity for myoelectric control: A new rasch-built measure of prosthetic hand control,” *J. Rehabil. Med.*, vols. –1, no. 1, p. 1, Jan. 2004.

[21] E. B. Lavoie et al., “Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks,” *J. Vis.*, vol. 18, no. 6, p. 18, Jun. 2018.

[22] K. Y. Cheng, C. S. Chapman, and J. S. Hebert, “Spatiotemporal coupling of hand and eye movements when using a myoelectric prosthetic hand,” in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2022, pp. 1–6.

[23] K. Y. Cheng, H. E. Williams, A. W. Shehata, P. M. Pilarski, C. S. Chapman, and J. S. Hebert, “Exploring the impact of myoelectric prosthesis controllers on visuomotor behavior,” *J. NeuroEngineering Rehabil.*, vol. 22, no. 1, p. 57, Mar. 2025.

[24] H. E. Williams, C. S. Chapman, P. M. Pilarski, A. H. Vette, and J. S. Hebert, “Gaze and movement assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol,” *PLoS ONE*, vol. 14, no. 12, Dec. 2019, Art. no. e0219333.

[25] H. E. Williams, C. S. Chapman, P. M. Pilarski, A. H. Vette, and J. S. Hebert, “Myoelectric prosthesis users and non-disabled individuals wearing a simulated prosthesis exhibit similar compensatory movement strategies,” *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 72, May 2021.

[26] A. M. Valevicius, Q. A. Boser, C. S. Chapman, P. M. Pilarski, A. H. Vette, and J. S. Hebert, “Compensatory strategies of body-powered prosthesis users reveal primary reliance on trunk motion and relation to skill level,” *Clin. Biomechanics*, vol. 72, pp. 122–129, Feb. 2020.

[27] J. S. Hebert et al., “Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users,” *JAMA Netw. Open*, vol. 2, no. 9, Sep. 2019, Art. no. e1911197.

[28] C. Lopez, P. Halje, and O. Blanke, “Body ownership and embodiment: Vestibular and multisensory mechanisms,” *Neurophysiologie Clinique/Clin. Neurophysiology*, vol. 38, no. 3, pp. 149–161, Jun. 2008.

[29] K. Kilteni, R. Groten, and M. Slater, “The sense of embodiment in virtual reality,” *Presence*, vol. 21, no. 4, pp. 373–387, Nov. 2012.

[30] R. Fribourg, F. Argelaguet, A. Lecuyer, and L. Hoyet, “Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view,” *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 2062–2072, May 2020.

[31] M. Botvinick and J. Cohen, “Rubber hands ‘feel’ touch that eyes see,” *Nature*, vol. 391, no. 6669, p. 756, Feb. 1998.

[32] M. V. Sanchez-Vives, B. Spanlang, A. Frisoli, M. Bergamasco, and M. Slater, “Virtual hand illusion induced by visuomotor correlations,” *PLoS ONE*, vol. 5, no. 4, Apr. 2010, Art. no. e10381.

[33] A. Maselli and M. Slater, “The building blocks of the full body ownership illusion,” *Frontiers Human Neurosci.*, vol. 7, p. 83, Mar. 2013.

[34] L. D. Walsh, G. L. Moseley, J. L. Taylor, and S. C. Gandevia, “Proprioceptive signals contribute to the sense of body ownership,” *J. Physiol.*, vol. 589, no. 12, pp. 3009–3021, Jun. 2011.

[35] S. Keenaghan, L. Bowles, G. Crawfurd, S. Thurlbeck, R. W. Kentridge, and D. Cowie, “My body until proven otherwise: Exploring the time course of the full body illusion,” *Consciousness Cognition*, vol. 78, Feb. 2020, Art. no. 102882.

[36] E. Kokkinara, K. Kilteni, K. J. Blom, and M. Slater, “First person perspective of seated participants over a walking virtual body leads to illusory agency over the walking,” *Sci. Rep.*, vol. 6, no. 1, p. 28879, Jul. 2016.

[37] A. Mottelson, A. Muresan, K. Hornbæk, and G. Makransky, “A systematic review and meta-analysis of the effectiveness of body ownership illusions in virtual reality,” *ACM Trans. Comput.-Hum. Interact.*, vol. 30, no. 5, pp. 1–42, Sep. 2023.

[38] J. S. Schofield, M. A. Battraw, A. S. R. Parker, P. M. Pilarski, J. W. Sensinger, and P. D. Marasco, “Embodied cooperation to promote forgiving interactions with autonomous machines,” *Frontiers Neurorobotics*, vol. 15, pp. 1–7, Apr. 2021.

[39] J. Zbinden, E. Lendaro, and M. Ortiz-Catalan, “Prosthetic embodiment: Systematic review on definitions, measures, and experimental paradigms,” *J. NeuroEngineering Rehabil.*, vol. 19, no. 1, p. 37, Mar. 2022.

[40] P. Caserman, M. Martinussen, and S. Göbel, “Effects of end-to-end latency on user experience and performance in immersive virtual reality applications,” in *Entertainment Computing and Serious Games*, vol. 11863. Cham, Switzerland: Springer, 2019, pp. 57–69.

[41] T. Waltemate et al., “The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality,” in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol.*, NY, NY, USA, Nov. 2016, pp. 27–35.

[42] M. A. F. Ismail and S. Shimada, “‘Robot’ hand illusion under delayed visual feedback: Relationship between the senses of ownership and agency,” *PLoS ONE*, vol. 11, no. 7, Jul. 2016, Art. no. e0159619.

[43] D. Gall, D. Roth, J.-P. Stauffert, J. Zarges, and M. E. Latoschik, “Embodiment in virtual reality intensifies emotional responses to virtual stimuli,” *Frontiers Psychol.*, vol. 12, Sep. 2021, Art. no. 674179.

[44] T. Sanders, A. Kaplan, R. Koch, M. Schwartz, and P. A. Hancock, “The relationship between trust and use choice in human–robot interaction,” *Human Factors*, vol. 61, no. 4, pp. 614–626, Jun. 2019.

[45] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan, “Eye–hand coordination in object manipulation,” *J. Neurosci.*, vol. 21, no. 17, pp. 6917–6932, Sep. 2001.

[46] A. Chadwell et al., "Upper limb activity in myoelectric prosthesis users is biased towards the intact limb and appears unrelated to goal-directed task performance," *Sci. Rep.*, vol. 8, no. 1, p. 11084, Jul. 2018.

[47] A. Chadwell, L. Kenney, S. Thies, J. Head, A. Galpin, and R. Baker, "Addressing unpredictability may be the key to improving performance with current clinically prescribed myoelectric prostheses," *Sci. Rep.*, vol. 11, no. 1, p. 3300, Feb. 2021.

[48] D. Chappell et al., "Virtual reality pre-prosthetic hand training with physics simulation and robotic force interaction," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4550–4557, Apr. 2022.

[49] A. Boschmann, D. Neuhaus, S. Vogt, C. Kaltschmidt, M. Platzner, and S. Dosen, "Immersive augmented reality system for the training of pattern classification control with a myoelectric prosthesis," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 25, Feb. 2021.

[50] L. V. Dijk, C. K. V. D. Sluis, H. W. V. Dijk, and R. M. Bongers, "Learning an EMG controlled game: task-specific adaptations and transfer," *PLoS ONE*, vol. 11, no. 8, Aug. 2016, Art. no. e0160817.

[51] F. E. R. Mattioli, E. A. Lamounier, A. Cardoso, A. B. Soares, and A. O. Andrade, "Classification of EMG signals using artificial neural networks for virtual hand prosthesis control," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 7254–7257.

[52] C. A. Garske, M. Dyson, S. Dupan, G. Morgan, and K. Nazarpour, "Serious games are not serious enough for myoelectric prosthetics," *JMIR Serious Games*, vol. 9, no. 4, Nov. 2021, Art. no. e28079.

[53] B. Maas, J. Tchimino, B. V. Dijk, A. Murgia, C. K. V. D. Sluis, and R. M. Bongers, "The effectiveness of virtual reality training for arm prosthesis control compared with prosthesis simulator training," in *Proc. Myoelectric Controls Symp.*, Aug. 2024, pp. 1–4.

[54] B. Maas, J. Tchimino, B. van Dijk, A. Murgia, C. K. van der Sluis, and R. M. Bongers, "Skill transfer of upper limb prosthesis control after training in a virtual reality environment," *Technol. Disability*, vol. 37, no. 2, pp. 137–148, Jan. 2025.

[55] D. T. Kluger et al., "Virtual reality provides an effective platform for functional evaluations of closed-loop neuromyoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 876–886, May 2019.

[56] V. Kumar and E. Todorov, "MuJoCo HAPTIX: A virtual reality system for hand manipulation," in *Proc. IEEE-RAS 15th Int. Conf. Humanoid Robots (Humanoids)*, Nov. 2015, pp. 657–663.

[57] J. Xie and X. Hu, "Virtual reality for evaluating prosthetic hand control strategies: A preliminary report," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Nov. 2021, pp. 6263–6266.

[58] J. Happold, D. Morais, P. Capsi-Morales, and C. Piazza, "Preliminary validation of an immersive virtual reality framework for prosthetic embodiment quantification," in *Proc. 10th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Sep. 2024, pp. 68–74.

[59] G. Barresi et al., "Exploring the embodiment of a virtual hand in a spatially augmented respiratory biofeedback setting," *Frontiers Neurorobotics*, vol. 15, pp. 1–17, Aug. 2021.

[60] H. Braun et al., "Introducing MOSAIC: A modular, open-source suite for assistive intelligent control," *IEEE Access*, vol. 13, pp. 213713–213733, 2025.

[61] I. Schuetz and K. Fiehler, "Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability," *J. Eye Movement Res.*, vol. 15, no. 3, p. 10, Sep. 2022.

[62] K. Zakka and Y. Tassa, "MuJoCo Menagerie: A collection of high-quality simulation models for MuJoCo," MuJoCo Menagerie Contributors, Tech. Rep., 2022.

[63] D. Di Domenico et al., "Towards bioinspired co-adaptive control of a 3-DoFs upper-limb prosthesis via incremental learning," Tech. Rep., Mar. 2025.

[64] S. Butterworth, "On the theory of filter amplifiers," *Experim. Wireless*, vol. 7, no. 6, pp. 536–541, Oct. 1930.

[65] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970.

[66] D. D. Domenico et al., "Co-adaptive velocity and position control of 3-DoFs prosthesis via incremental learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 34, pp. 856–867, 2026.

[67] F. Egle and C. Bäker, "AIROB-lab/SIM—Streamlined-input-manager-UDP: V0.1.2-beta," Zenodo, FAU Erlangen-Nürnberg, Erlangen, Germany, Tech. Rep., Jun. 2025.

[68] D. Roth and M. E. Latoschik, "Construction of the virtual embodiment questionnaire (VEQ)," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 12, pp. 3546–3556, Dec. 2020.

[69] P. R. Cavanagh and P. V. Komi, "Electromechanical delay in human skeletal muscle under concentric and eccentric contractions," *Eur. J. Appl. Physiol. Occupational Physiol.*, vol. 42, no. 3, pp. 159–163, Nov. 1979.

[70] S. Gallagher, "Philosophical conceptions of the self: Implications for cognitive science," *Trends Cognit. Sci.*, vol. 4, no. 1, pp. 14–21, Jan. 2000.

[71] M. Tsakiris, "My body in the brain: A neurocognitive model of body-ownership," *Neuropsychologia*, vol. 48, no. 3, pp. 703–712, Feb. 2010.

[72] M. Tsakiris, S. Schütz-Bosbach, and S. Gallagher, "On agency and body-ownership: Phenomenological and neurocognitive reflections," *Consciousness cognition*, vol. 16, no. 3, pp. 645–660, 2007.

[73] M. Tsakiris, G. Prabhu, and P. Haggard, "Having a body versus moving your body: How agency structures body-ownership," *Consciousness Cognition*, vol. 15, no. 2, pp. 423–432, Jun. 2006.

[74] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cognit. Ergonom.*, vol. 4, pp. 53–71, Mar. 2000.

[75] Z. Li, M. Bujić, O. Buruk, E. Bampouni, S. Järvelä, and J. Hamari, "Haptics-mediated virtual embodiment: Impact of a wearable avatar-controlling system with kinesthetic gloves on embodiment in VR," *Frontiers Virtual Reality*, vol. 5, pp. 1–12, Sep. 2024.

[76] M. Chita-Tegmark, T. Law, N. Rabb, and M. Scheutz, "Can you trust your trust measure?," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, Mar. 2021, pp. 92–100.

[77] M. Wilson, J. McGrath, S. Vine, J. Brewer, D. Defriend, and R. Masters, "Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts," *Surgical Endoscopy*, vol. 24, no. 10, pp. 2458–2464, Oct. 2010.

[78] J. V. V. Parr, S. J. Vine, N. Harrison, and G. Wood, "Examining the spatiotemporal disruption to gaze when using a myoelectric prosthetic hand," *J. Motor Behav.*, vol. 50, no. 4, pp. 416–425, Jul. 2017.

[79] E. Lavoie, J. S. Hebert, and C. S. Chapman, "Comparing eye–hand coordination between controller-mediated virtual reality, and a real-world object interaction task," *J. Vis.*, vol. 24, no. 2, p. 9, Feb. 2024.

[80] C. Kleinbeck, H. Schieber, J. Kreimeier, A. Martin-Gomez, M. Unberath, and D. Roth, "Injured avatars: The impact of embodied anatomies and virtual injuries on well-being and performance," *IEEE Trans. Vis. Comput. Graphics*, vol. 29, no. 11, pp. 4503–4513, Nov. 2023.

[81] T. R. Farrell and R. F. Weir, "The optimal controller delay for myoelectric prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 111–118, Mar. 2007.

[82] L. Bareišytė et al., "Questionnaires for evaluating virtual reality: A systematic scoping review," *Comput. Human Behav. Rep.*, vol. 16, Dec. 2024, Art. no. 100505.

[83] M. Guy, J. Normand, C. Jeunet-Kelway, and G. Moreau, "The sense of embodiment in virtual reality and its assessment methods," *Frontiers Virtual Reality*, vol. 4, pp. 1–23, Dec. 2023.

[84] T. Mielke, M. Allgaier, D. Schott, C. Hansen, and F. Heinrich, "Virtual studies, real results? Assessing the impact of virtualization on human–robot interaction," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, NY, NY, USA, Apr. 2025, pp. 1–8.

[85] A. Küderle, R. Richer, R. C. Sîmpetru, and B. M. Eskofier, "Tpcp: Tiny pipelines for complex problems–A set of framework independent helpers for algorithms development and evaluation," *J. Open Source Softw.*, vol. 8, no. 82, p. 4953, Feb. 2023.

[86] B. Hodossy and J. Llobera, "Balint-H/modular-agents," Imperial College, London, U.K., Tech. Rep., Jun. 2025.