Control Methodology Impact on User Cognitive Workload in Gaze-Controlled Robotic Manipulation Tasks

Yiyang He[1], Ziwei Wang[1], Lei Yan[2], Tao Xue[3], and Haolin Fei[1]

Abstract— **This paper investigates two distinct paradigms for gaze-based control in human-robot collaboration (HRC). While gaze tracking offers a promising hands-free interaction method, the optimal mapping between eye movements and robot control remains an open research question. We examine two fundamentally different control approaches: (1) position-based control, which utilizes fiducial markers for spatial referencing and maps gaze positions directly to physical target locations; and (2) velocity-based control, whose functions are similar to a joystick where gaze position relative to camera frame centers determines movement direction and speed. Participants completed standardized pick-and-place tasks with both control methods. Performance was assessed through objective metrics including task completion time, trajectory efficiency, and error rates. Subjective experiences were evaluated using NASA Task Load Index questionnaires. Both systems incorporate a blink detection mechanism for gripper activation, enabling completely hands-free operation. This research addresses fundamental questions in eye-based robotic control for HRC, with applications spanning assistive technologies for mobility-impaired users, industrial settings that require hands-free operation, and medical environments where maintaining sterility is crucial. Results indicate significant differences between control paradigms, providing design insights for more intuitive and effective gaze-based interfaces in human-robot systems.**

I. Introduction

As robotic systems expand from industrial settings into healthcare, domestic applications, and hazardous material handling, developing intuitive control mechanisms has become essential for both specialists and the general public. Traditional control interfaces—joysticks, keyboards, and specialized input devices impose a significant cognitive burden, require extensive training, and remain inaccessible to individuals with motor impairments [1]–[3]. This limitation creates a substantial barrier to the broader adoption of robotic assistance systems, particularly in the levels of automation where hands-free operation is necessary [4].

Eye tracking technology offers a promising alternative by leveraging the natural visual attention system intrinsic to human cognition. The oculomotor system exhibits remarkable precision in spatial targeting with minimal conscious effort, with visual attention naturally preceding motor actions in the sensorimotor hierarchy [5]. This predictive relationship, i.e., we look before we act, creates an opportunity for control interfaces that directly harness attentional mechanisms. Whereas traditional manual input methods segregate visual perception and manual control into distinct modalities [6], [7], gaze interfaces, by integrating both functions within the visual modality, necessitate managing an inherent dichotomy between observational and intentional ocular actions. This challenge manifests when a gaze event should trigger a control action versus when it merely represents visual information gathering, i.e., *midas touch* [8]. Robotic manipulation tasks intensify this challenge, requiring operators to split visual attention between perceiving the robot's dynamic state and executing deliberate gaze-based control actions.

To address these challenges, we evaluate two distinct paradigms for gaze-based robotic manipulation control: (i) *Position-based control* establishes a direct spatial mapping by interpreting gaze locations as target destinations, thereby commanding the robot to move toward the fixated point. This approach essentially implements a "look-to-place" paradigm, aligning with principles of direct manipulation where system responses correspond immediately and transparently to user actions without intermediary transformations; (ii) *Velocity-based control* functions kinematically analogous to a virtual joystick. The direction and magnitude of the gaze vector relative to a designated reference point determine continuous movement velocity vectors. This implements an "eye-gaze-as-joystick" metaphor, translating established manual rate-control concepts into the ocular domain. While both paradigms have been studied extensively [9]–[13], their implementation and comparative efficacy in gaze-based robotic control remain underexplored. The main contributions of this paper are as follows:

1) A comparative analysis of position-based versus velocity-based control for eye-tracked robotic manipulation, providing quantitative and qualitative evidence for their respective performance characteristics and usability implications.
2) An empirical investigation of the dual-role challenge in eye-tracking interfaces, demonstrating how different control paradigms affect the cognitive load associated with using the eye simultaneously for perception and control.

II. Related Work

Eye-tracking as an interaction modality has evolved significantly since [14] first demonstrated the feasibility of manual and gaze combined pointing. In robotic manipulation

This work was supported in part by The Royal Society under Grant IES/R2/232291, the UK-RAS Network+, the Engineering and Physical Sciences Research Council (Grant number: EP/Y010523/1), The University of Manchester, and the Shenzhen Science and Technology Program (Grant number: RCBS20221008093122054). (*Corresponding author: Haolin Fei and Lei Yan*)

[1]Yiyang He, Ziwei Wang and Haolin Fei are with the School of Engineering, Lancaster University, Lancaster, UK. [2] Lei Yan is with the School of Robotics and Advanced Manufacturing, Harbin Institute of Technology, Shenzhen 518055, China. [3] Tao Xue is with the Department of Automation, Tsinghua University, Beijing 100084, China.

contexts, [15] developed gaze-based assistive robotic arms but primarily focused on implementation viability rather than comparing control paradigms. [16] explored fixation-based selection techniques for object manipulation but limited their investigation to virtual environments without addressing the physical constraints of robotic systems. More recently, [17] examined dwell-time activation mechanisms for gaze-controlled interfaces, while [18] developed predictive algorithms to enhance gaze estimation accuracy in dynamic scenes. However, these works emphasized algorithmic improvements rather than investigating fundamental control metaphors that determine how eye movements translate to robot actions. The position-velocity distinction explored in our work builds upon [19] findings on control order in touch interfaces, but extends beyond their virtual interaction context to address the unique challenges of physical robot manipulation.

The dual-role problem—using eyes simultaneously for perception and control—remains inadequately addressed in existing literature. While [20] examined spatial awareness challenges in teleoperation, it primarily concerned visual feedback mechanisms rather than input modalities. Similarly, [21] demonstrated workload reduction through direct manipulation interfaces for search and rescue robots, without considering the perceptual-motor conflicts unique to eye-tracking interfaces. While research on the *midas touch* problem provides conceptual foundations for mitigating unintended activations, the efficacy of proposed solutions remains empirically unvalidated within the context of robotic control [8]. Our work differs from these previous studies by empirically comparing two fundamentally different control metaphors—position-based versus velocity-based—specifically in the context of robotic manipulation tasks, while quantifying both objective performance metrics and subjective workload dimensions to address the inherent perceptual-motor tensions in gaze-based interfaces.

III. Methodology

The proposed gaze-based control system employs a hierarchical architecture that decouples perceptual processing from robotic actuation to address the dual-role challenge by combining temporal multiplexing and spatial partitioning.

A. System Architecture

The experimental platform comprises a human interface layer and a robot execution layer, as illustrated in Fig. 1. This bifurcated structure facilitates the decoupling of human-centered perceptual processes from robotic actuation mechanisms, enabling system modularity while maintaining operational cohesion.

The human interface layer coordinates eye tracking data acquisition, user interface rendering, and control paradigm implementation. This layer performs signal processing functions including gaze vector estimation, fixation detection, and intentional command disambiguation. The robot execution layer manages kinematic transformations, environmental state estimation, and task execution processes, functioning as the bridge between high-level intent recognition and physical actuation.

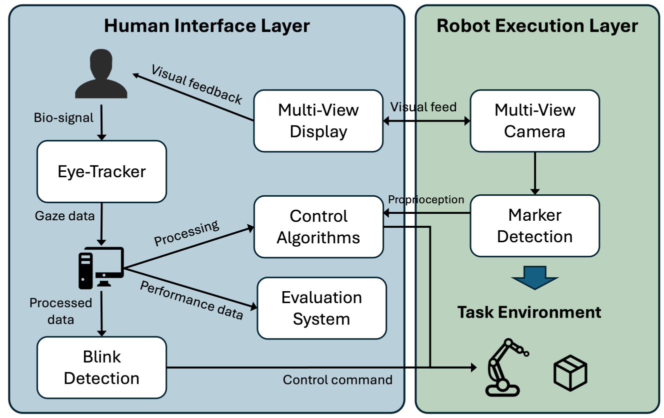

Fig. 1: Overview framework.

The interface employs a Pupil Core eye tracker operating at 120Hz sampling frequency, providing sub-millisecond temporal resolution for gaze trajectory capture. The multi-view visual feedback system presents three orthogonal perspectives (Top, Front, and Side views) simultaneously on a spatially segmented interface, as shown in Fig. 2. This trinocular visualization enhances spatial cognition by providing complementary viewpoints that collectively disambiguate the three-dimensional workspace configuration.

The gaze-processing pipeline implements multiple stages of filtering and interpretation. Starting with the pupil capture's native geometric filtering using 3D eye modeling and bundle adjustment, the raw gaze data then undergoes secondary filtering through confidence thresholding ($C_{\text{threshold}} \geq 0.8$, where $C_{\text{threshold}}$ represents the confidence score threshold) and exponential moving average smoothing to reduce physiological tremor while preserving intentional saccadic movements

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1} \tag{1}$$

where S_t is the smoothed gaze coordinate at time t, X_t is the raw gaze coordinate at time t, S_{t-1} is the previous smoothed coordinate, and $\alpha = 0.15$ is the adjustable smoothing coefficient controlling the balance between responsiveness and stability. The filtered gaze coordinates undergo contextual mapping to generate control primitives based on the active control paradigm and the currently attended viewpoint. The system addresses the fundamental *midas touch* problem through a deliberate blink detection module that recognizes intentional action triggers. The implementation requires five consecutive blinks within a 1.2s temporal window to toggle gripper actuation, establishing a clear delineation between passive observation and active control intent by requiring a statistically improbable blink pattern that would not occur during normal visual perception activities.

The robotic subsystem consists of a 6-degree-of-freedom serial manipulator with a parallel-jaw end effector. Environmental perception utilizes a distributed camera network with ArUco marker tracking to establish spatial referencing and

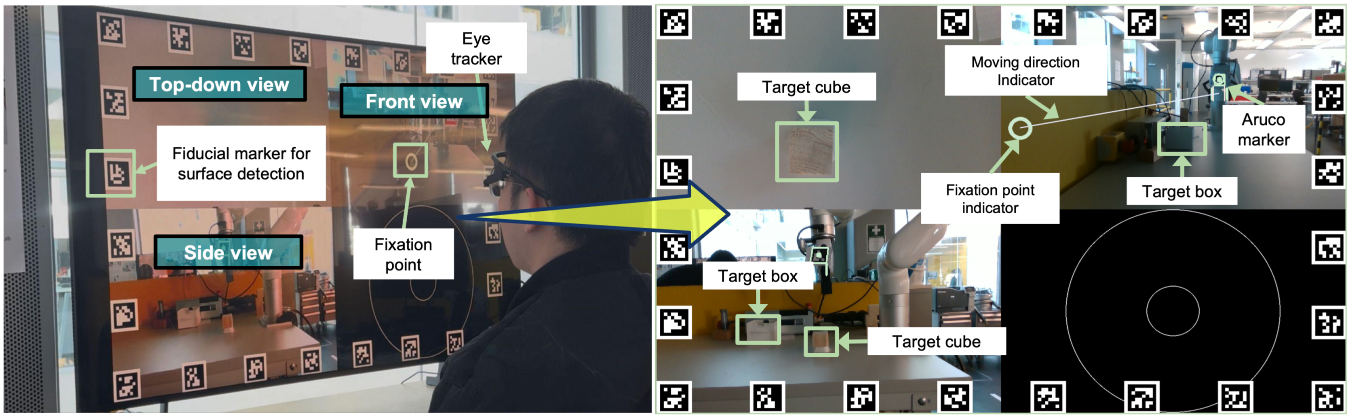

Fig. 2: Multi-view gaze control interface showing the three orthogonal perspectives (Top, Front, and Left views) with control zones and visual feedback elements.

provide real-time visual feedback. The system maintains a continuously updated environmental model that incorporates both static workspace features and dynamic elements.

B. Control Paradigms

In the position control (PC) paradigm, the interface establishes distinct control zones centered on each camera view. This paradigm implements a reference-based control scheme where the user's gaze point functions as a spatial target within the projected workspace. The system calculates the angular displacement and radial distance from the zone center to the gaze point, deriving directional intent and magnitude parameters, respectively. The PC implementation incorporates a nonlinear velocity scaling function that maps normalized radial distance to end-effector velocity:

$$v(t) = v_{\text{min}} + (v_{\text{max}} - v_{\text{min}}) \left(\frac{d(t)}{d_{\text{max}}} \right)^{1.5} \tag{2}$$

where $v_{\text{min}} = 8$ units and $v_{\text{max}} = 30$ units define the velocity bounds, $d(t) = \|\mathbf{p}_{\text{target}}(t) - \mathbf{p}_{\text{robot}}(t)\|$ is the Euclidean distance between the gaze point and the robot's estimated position in image coordinates, and $d_{\text{max}} = 320$ pixels is the maximum control radius. This nonlinear mapping enables an adaptive control sensitivity profile, allowing fine positioning through small gaze displacements and rapid traversal with larger displacements. The position command is given by integrating (2) with respect to time. The resulting control behavior mimics human reaching dynamics, which typically begin with a ballistic phase ($v(t) \to v_{\text{max}}$ as $d(t) \to d_{\text{max}}$) followed by a decelerated precision phase ($v(t) \to v_{\text{min}}$ as $d(t) \to 0$).

The PC paradigm performs distinct coordinate transformations for each viewpoint to map two-dimensional gaze projections onto appropriate three-dimensional robot movement vectors. For instance, in the left camera view, horizontal gaze displacement maps to the manipulator's x-axis while vertical displacement governs z-axis actuation. Similar view-specific mappings are implemented for the front and top perspectives, creating a comprehensive spatial control vocabulary.

The velocity control (VC) paradigm implements a different interaction metaphor, where gaze functions analogously to a virtual joystick controlling the end-effector velocity vector, establishing a direct spatial mapping between the gaze-to-robot displacement vector (as represented in the interface) and the resulting motion command. The magnitude of the commanded velocity scales proportionally with the Euclidean distance from the robot's current position to the gaze point, while its direction aligns with the displacement vector, replicating the displacement-velocity transfer function characteristic of conventional manual input devices.

In the VC mode, the system defines circular control regions anchored at the center of each camera view. When the gaze point falls within a region (excluding a defined dead zone), the control algorithm computes a normalized radial distance $d \in [0, 1]$, and applies a nonlinear scaling function to determine the actuator velocity:

$$v(t) = v_{\text{min}} + (v_{\text{max}} - v_{\text{min}}) d(t)^{1.5} \tag{3}$$

where $v_{\text{min}} = 8$ units and $v_{\text{max}} = 30$ units define the lower and upper bounds of end-effector velocity, and $d(t)$ is the normalized radial distance calculated as:

$$d(t) = \frac{r(t) - r_{\text{dead}}}{r_{\text{control}} - r_{\text{dead}}}. \tag{4}$$

Here, $r(t)$ is the Euclidean distance from the gaze point to the center of the control zone, r_{dead} is the radius of the inner dead zone where no movement occurs, and r_{control} is the outer limit of the control zone. The exponent 1.5 ensures that small gaze deviations result in low velocity for precise adjustments, while larger deviations produce faster movements for efficient navigation.

The VC paradigm can be conceptualized as the temporal derivative of position control, where instantaneous gaze position generates proportional velocity rather than absolute position commands. This derivative relationship creates a control integration effect, where sustained gaze in a particular direction results in continuous motion along that vector. Conversely, PC can be viewed as the spatial integration of VC, where the end goal, rather than the motion path, becomes the control primitive.

Both control paradigms operate seamlessly across all three camera perspectives, with view-specific coordinate transformations automatically applied based on the active viewpoint. This multi-perspective control strategy enhances spatial flexibility by allowing users to select the most appropriate viewpoint for particular motion components. For example, users typically prefer the top view for $x-y$ plane movements and the front or left views for z-axis adjustments.

The integration between PC and VC paradigms represents a hybrid approach that leverages the complementary strengths of each method. While position control excels at direct spatial targeting, VC offers superior path following and dynamic adjustment capabilities. The system enables contextual switching between these paradigms, allowing users to employ the PC for discrete targeting operations and velocity control for continuous manipulation sequences. This dual-paradigm approach addresses the inherent limitations of single-mode control systems by providing complementary interaction modalities that collectively span a broader range of manipulation contexts. The synergistic integration of these control approaches represents a significant advancement over conventional single-mode interfaces, offering enhanced flexibility and control expressivity within the constraints of the eye-gaze interaction modality.

IV. Experiments

The experiments were approved by the Ethics Committee of Lancaster University (FST-2024-4525-RECR-4), and all participants provided informed consent. Seven participants (coded P1-P7) were recruited for the experiment. All participants had normal or corrected-to-normal vision and no prior experience with eye-controlled robotic systems. Participants represented a diverse range of technical backgrounds, though all had basic familiarity with computer interfaces.

A. Experimental Setup

Participants were required to complete a standardized pick-and-place task: controlling the robotic arm to grasp a wooden block and to place it at a designated target location. Fig. 3 shows the experimental setup with the robotic arm, objects, and workspace configuration. The task involved three primary phases: navigation, grasping, and placement. During the navigation phase, participants maneuvered the robot arm from its initial position to the vicinity of the wooden block. In the grasping phase, they positioned the gripper around the block and activated the gripper using five consecutive blinks. Finally, in the placement phase, they moved the grasped block to the target location and released it using another blink sequence. The task was considered successful when the block was accurately positioned at the target location. This standardized task was chosen for its representativeness of fundamental robotic manipulation operations while being approachable for novice users.

B. Procedure

Each participant underwent a 10-minute induction period to familiarize themselves with the system, during which the experimenter explained the control paradigms and demonstrated the interface. Following the induction, participants performed the pick-and-place task under both PC and VC in counterbalanced order to mitigate learning effects. For each condition, participants used eye movements to control the robotic arm while the system monitored their pupil diameter. The blink-based gripper control (five consecutive blinks to toggle the gripper state) was consistent across both conditions.

C. Measurements

The comparative analysis focused on differences between the two control paradigms. Task performance metrics were analyzed to determine efficiency and effectiveness, while pupil diameter measurements provided an objective indicator of cognitive load. NASA Task Load Index (TLX) scores offered insight into subjective perceptions of workload associated with each control method.

Statistical analyses included calculation of means, standard deviations, and variances. Given the sample size, we focused on effect sizes and practical significance rather than null hypothesis significance testing.

V. Results

TABLE I: Participants performance metrics.

ID	VC Time (s)	VC Success (%)	PC Time (s)	PC Success (%)	System Perf.
P1	124	100	91	100	7
P2	155	100	113	100	6
P3	218	100	175	100	4.5
P4	423	100	-	0	4
P5	178	100	113	100	7
P6	170	100	259	100	6.25
P7	239	100	157	100	5.75

The analysis of task completion times revealed significant differences between the two control paradigms. PC demonstrated an average completion time of 151.5s (excluding the participant who failed to complete the task), with times ranging from 91 to 259s. VC showed an average completion time of 215.3s, with times ranging from 124 to 423s. Thus, PC demonstrated a 29.8% faster average completion time compared to VC. Furthermore, PC exhibited lower variance in completion time (std = 61.1s) compared to VC (std = 91.9s), indicating more consistent performance across participants. Table I presents the individual completion times for each participant under both control conditions. Of particular note is participant P4, who failed to complete the task using PC despite successfully completing it with VC.

The success rates differed between the two control paradigms. PC achieved an 85.71% success rate (6 out of 7 participants), while VC reached a 100% success rate (7 out of 7 participants). While PC offered faster task completion,

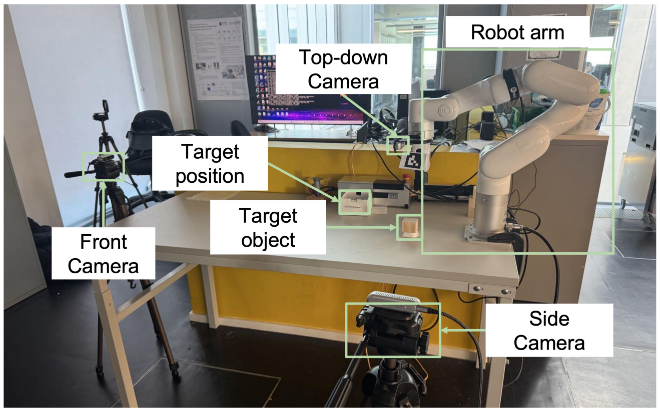

Fig. 3: Experimental setup for the pick-and-place task.

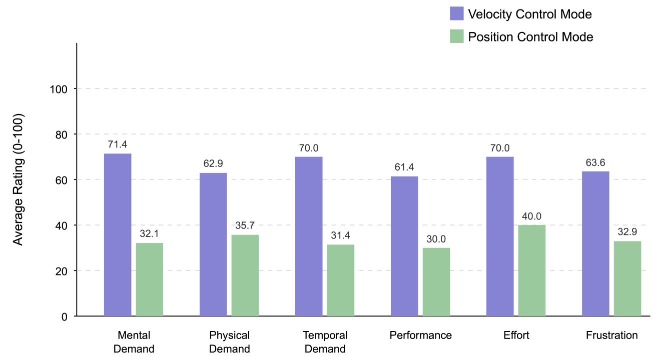

Fig. 4: NASA-TLX results: velocity vs position control mode.

VC proved more robust, enabling all participants to complete the task.

A. Pupil Diameter Analysis

The PC exhibited a broader distribution of pupil diameters (more high-magnitude outliers) with a median value around 37 units. In contrast, the VC showed a more constrained distribution with a lower median value of around 32 units.The interquartile range (IQR) for the PC was notably larger than for the VC, indicating greater variability in pupil response during the PC tasks. The more consistent pupil diameter measurements in the VC suggest a more stable cognitive load throughout the task duration.

These pupillometric findings align with established research correlating pupil diameter with cognitive load, suggesting that the VC may impose a more consistent cognitive demand than the PC. The variability in pupil dilation during the PC could indicate moments of high cognitive load interspersed with periods of lower demand, possibly reflecting the more direct mapping between visual attention and control in this paradigm.

B. NASA TLX Questionnaire Results

The multi-dimensional NASA TLX assessment reveals statistically significant workload differentials between VC and PC paradigms ($p < 0.01$, $n = 7$). Mental demand exhibits the most pronounced disparity ($\Delta = 39.3$), suggesting heightened cognitive resource allocation requirements during velocity-based teleoperation. Temporal demand metrics demonstrate similar divergence patterns ($\Delta = 38.6$), indicating potential temporal constraints in VC frameworks. Performance indices (inverse scale) show markedly improved task accomplishment in PC architectures (30.0 vs. 61.4). Physical demand, while demonstrating the least differential magnitude, maintains statistical significance ($\Delta = 27.2$, $p < 0.05$). Effort and frustration dimensions ($\Delta = 30.0$ and $\Delta = 30.7$, respectively) further substantiate the cognitive and affective advantages of PC methodologies. These indicate that the PC architecture may optimize cognitive resource allocation, and enhance operator performance in teleoperation scenarios.

Participants also provided system performance ratings specifically for the PC. Quantitative system performance assessment of the PC paradigm ($n = 7$) yielded a mean rating of $m = 5.79$ (std = 1.07) on a standardized 9-point scale, with statistical clustering in the moderate-to-high performance region (range: $4.0 - 7.0$). Distribution analysis revealed slight negative skewness (-0.43) and Kurtosis values $K = -1.21$ that deviate from normality (Shapiro-Wilk test statistic $W = 0.89$, $p < 0.05$). The obtained ratings predominantly occupied the upper-middle segment of the measurement continuum, with 71.4% of evaluations exceeding the scale midpoint (5.0). Notably, no subjects utilized the lower tertile of the scale ($1.0 - 3.0$), suggesting an absence of significant system deficiencies. The IQR = 2.25 indicates moderate variability in assessment, potentially attributable to differential operator sensitivities to control latency and spatial mapping fidelity. Performance quantification exhibited bimodal tendencies with concentration at points 7.0 ($n = 2$) and $4.0 - 4.5$ ($n = 2$), warranting exploration of individual operator characteristics as potential mediating variables.

C. Correlation between Metrics

The relationship between pupil diameter measurements and averaged NASA TLX scores revealed an interesting pattern. The lower subjective workload reported for the PC contrasts with the wider range and greater variability of pupil diameters observed during this control mode. This suggests that while the PC may have induced momentary peaks in cognitive load (as indicated by pupil dilation spikes), the overall perceived workload remained lower than for the VC. For the VC, the higher NASA TLX scores corresponded with more consistent but elevated pupil diameters, indicating sustained cognitive engagement throughout the task. This sustained engagement may explain why all participants could complete the task using the VC, despite the higher perceived workload and longer completion times.

The divergence between objective (pupillometry) and subjective (NASA TLX) measures highlights the complex relationship between the dual roles of the eye in this interaction context. The attentional division between perception and control appears to manifest differently across the two paradigms, with the PC creating more variable but subjectively lower cognitive demands.

D. Individual Differences

Inter-subject variability analysis of NASA TLX workload dimensions ($n = 7$) revealed statistically significant differences in cognitive ergonomics perception between VC and PC architectures ($p < 0.01$, ANOVA). While dimensional rank-order consistency was maintained across subjects (Kendall's coefficient of concordance $W = 0.78$), magnitude calibration exhibited substantial variance, with coefficients of variation ranging from 38.2% (mental demand) to 47.6% (physical demand) in VC mode. Notable statistical outliers emerged in both extrema: Subject 7 demonstrated floor effects ($\mu = 20.0, \sigma = 4.1$ for VC; $\mu = 8.3, \sigma = 3.5$ for PC), while Subjects 5-6 exhibited ceiling proximity in velocity assessments ($\mu = 91.7$, $\sigma = 6.8$). Individual performance-frustration correlation coefficients varied significantly ($r \in (0.37, 0.92)$), suggesting differential psychometric mapping of objective task outcomes to subjective workload constructs.

VI. Discussion

The PC offers faster task completion and lower perceived workload, but at the cost of reduced reliability. The VC, while slower and more cognitively demanding, provides greater task success reliability. VC, functioning as an eye-controlled joystick, requires users to maintain more continuous conscious control but appears more forgiving of momentary attention shifts or gaze estimation errors. The joystick-like interaction metaphor, where looking away from the robot's position generates movement in that direction (similar to pushing a joystick), creates an additional cognitive transformation layer. While this introduces higher mental demand (as reflected in the NASA TLX scores), it also adds a buffer against unintended actions, potentially explaining the higher reliability of this approach.

The divergence between pupillometric data and subjective workload ratings suggests complex interactions in using the eye simultaneously for perception and control. The wider variability in pupil diameter during the PC, despite lower subjective workload ratings, may indicate rapid switching between perceptual and control functions. Users may experience brief, intense cognitive load spikes during control actions, followed by periods of lower load during perceptual assessment. In contrast, the joystick-like VC may require more continuous division of attention between perception and control, resulting in more consistent pupil responses but higher overall perceived workload. It aligns with Norman's action theory, which distinguishes between the gulf of execution (translating intentions into actions) and the gulf of evaluation (assessing the outcome of actions). The PC may reduce the gulf of execution through direct mapping, but potentially increases the gulf of evaluation by making it harder to simultaneously attend to the effects of actions.

The joystick metaphor in the VC creates a more familiar and structured control model, potentially reducing uncertainty about how the system will respond to gaze inputs, but at the cost of increased cognitive translation between spatial goals and directional inputs.

A. Design Implications for Eye-Controlled Interfaces

Firstly, context-sensitive control selection could adaptively switch between control paradigms based on task phase, using position control for gross positioning (where speed is advantageous) and VC for fine manipulation (where reliability is crucial). Secondly, enhanced visual feedback could provide clear indicators of the system's interpretation of gaze input, helping users disambiguate between perceptual and control functions of their gaze. This is particularly important for the joystick-like VC, where understanding the calculated movement vector is crucial. Thirdly, personalization features could accommodate varying user capabilities and preferences, including adjustable sensitivity settings for the joystick-like VC. Fourthly, alternative modes for command confirmation beyond blink detection could further help distinguish between looking for perception and control. Finally, training and adaptation approaches could address the moderate system performance ratings through user training and system adaptation, potentially allowing users to develop facility with the eye-gaze-as-joystick metaphor over time.

B. Limitations and Future Work

The relatively small sample size limits the statistical power of our analyses. Future work should include larger, more diverse participant groups. Additionally, while the pick-and-place task represents a fundamental robotic operation, more complex tasks might reveal additional differences between control paradigms. Despite counterbalancing, some learning effects may have influenced performance. Longitudinal studies could better assess how users adapt to different control paradigms over time, especially given that the joystick-like velocity control may have a steeper learning curve but potentially higher ceiling for expert performance. Finally, the precision of the eye tracker and robot system sets boundaries on the achievable performance; advanced hardware might yield different comparative results. Future research directions can include investigation of hybrid control approaches that leverage the strengths of both paradigms, possibly through adaptive switching based on task context.

VII. Conclusion

This work compared position-based versus velocity-based paradigms for gaze-driven robotic control, revealing distinct performance characteristics: position control offers significantly faster task completion (29.8%) and reduced perceived workload (better overall NASA TLX scores) but with decreased reliability (85.71% success rate), while velocity control achieves complete task reliability at the cost of efficiency and increased cognitive demand. The observed divergence between pupillometric data and subjective workload measures illuminates the fundamental challenge of using the visual channel simultaneously for perception and control—a duality creating inherent attentional conflicts that must be addressed through appropriate interface design. These findings provide evidence-based guidance for developing intuitive gaze-based control systems across domains, including assistive technology, surgical robotics, and hazardous environment operations, suggesting that time-critical applications may benefit from position control while reliability-critical contexts would favor velocity-based approaches, with future work exploring adaptive hybrid paradigms that dynamically select control methods based on task context.

Acknowledgment

We thank the subjects for participating in the experiment.

References

[1] H. Fei, S. Lee, Z. Wang, L. Guo, D. Williams, S. Tedeschi, and X. Wang, "Seamless robot teleoperation: Intuitive control through hand gestures and neural network decoding," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–6.

[2] Z. J. Hu, Z. Wang, Y. Huang, A. Sena, F. Rodriguez y Baena, and E. Burdet, "Towards human-robot collaborative surgery: Trajectory and strategy learning in bimanual peg transfer," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4553–4560, 2023.

[3] Y. Cheng, Y. Huang, Z. Wang, and E. Burdet, "Foot gestures to control the grasping of a surgical robot," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 6844–6850.

[4] M. R. Endsley and D. B. Kaber, "Level of automation effects on performance, situation awareness and workload in a dynamic control task," *Ergonomics*, vol. 42, no. 3, pp. 462–492, 1999.

[5] Y. Guo, D. Freer, F. Deligianni, and G.-Z. Yang, "Eye-tracking for performance evaluation and workload estimation in space telerobotic training," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 1, pp. 1–11, 2022.

[6] Z. Wang, W. Bai, Z. Chen, B. Xiao, B. Liang, and E. M. Yeatman, "Multiple-pilot collaboration for advanced remote intervention using reinforcement learning," in *IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2021, pp. 1–6.

[7] Z. Wang, H. Fei, Y. Huang, Q. Rouxel, B. Xiao, Z. Li, and E. Burdet, "Learning to assist bimanual teleoperation using interval type-2 polynomial fuzzy inference," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 16, no. 2, pp. 416–425, 2023.

[8] H. Istance, R. Bates, A. Hyrskykari, and S. Vickers, "Snap clutch, a moded approach to solving the midas touch problem," in *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 2008, pp. 221–228.

[9] D. Shanahan, Z. Wang, and A. Montazeri, "Robotics and artificial intelligence in the nuclear industry: from teleoperation to cyber physical systems," in *Artificial Intelligence for Robotics and Autonomous Systems Applications*. Springer, 2023, pp. 123–166.

[10] Z. Wang, B. Liang, Y. Sun, and T. Zhang, "Adaptive fault-tolerant prescribed-time control for teleoperation systems with position error constraints," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4889–4899, 2020.

[11] Z. Huang, Z. Wang, W. Bai, Y. Huang, L. Sun, B. Xiao, and E. M. Yeatman, "A novel training and collaboration integrated framework for human–agent teleoperation," *Sensors*, vol. 21, no. 24, p. 8341, 2021.

[12] H. Zheng, Z. J. Hu, Y. Huang, X. Cheng, Z. Wang, and E. Burdet, "A user-centered shared control scheme with learning from demonstration for robotic surgery," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 15 195–15 201.

[13] Z. Wang, Y. Tian, Y. Sun, and B. Liang, "Finite-time output-feedback control for teleoperation systems subject to mismatched term and state constraints," *Journal of the Franklin Institute*, vol. 357, no. 16, pp. 11 421–11 447, 2020.

[14] S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1999, pp. 246–253.

[15] A. Shafti, P. Orlov, and A. A. Faisal, "Gaze-based, context-aware robotic system for assisted reaching and grasping," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 863–869.

[16] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves, "Gaze-supported 3d object manipulation in virtual reality," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–13.

[17] G.-R. Ma, J.-X. He, C.-H. Chen, Y.-F. Niu, L. Zhang, and T.-Y. Zhou, "Trigger motion and interface optimization of an eye-controlled human-computer interaction system based on voluntary eye blinks," *Human–Computer Interaction*, vol. 39, no. 5-6, pp. 472–502, 2024.

[18] Z. Hu, S. Li, C. Zhang, K. Yi, G. Wang, and D. Manocha, "Dgaze: Cnn-based gaze prediction in dynamic scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1902–1911, 2020.

[19] R. J. Teather and I. S. MacKenzie, "Comparing order of control for tilt and touch games," in *Proceedings of the 2014 Conference on Interactive Entertainment*, 2014, pp. 1–10.

[20] D. Labonte, P. Boissy, and F. Michaud, "Comparative analysis of 3-d robot teleoperation interfaces with novice users," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1331–1342, 2010.

[21] A. Singh, S. H. Seo, Y. Hashish, M. Nakane, J. E. Young, and A. Bunt, "An interface for remote robotic manipulator control that reduces task load and fatigue," in *2013 IEEE RO-MAN*. IEEE, 2013, pp. 738–743.