Exploring Pupil Dilation as an Indicator of Performance in Gaze-Based Robot Navigation for Assistive Technology

Parastoo Azizinezhad
University of Essex
Colchester, United Kingdom
p.azizinezhad@essex.ac.uk

Anirban Chowdhury
University of Essex
Colchester, United Kingdom
a.chowdhury@essex.ac.uk

***Abstract*—Human-robot interaction (HRI) based assistive devices play a crucial role for individuals with severe disability, significantly impacting their quality of life. A pivotal step towards creating a more human-centric HRI involves gaining a thorough understanding of the user's mental load such as cognitive load, stress, and fatigue, which can influence the performance of the system. Previous studies have found pupil dilation as a potential candidate for exploring mental workload. This paper explores the impact of pupil diameter variation on performance during an eye-tracking-based robot navigation task. Nineteen healthy individuals participated in the experiment where they used eye-gaze to activate different navigational buttons on a computer screen to control the movement of a mobile robot on a predefined trajectory for two rounds. The variation of pupil diameter is correlated to various performance parameters such as lap completion time and number of commands. Results show that the difference between the Gaussian means of the pupil diameter distribution during round1 and round2 is significantly correlated (ρ = 0.5, $p-$value $=$ 0.03) with the lap completion time while the correlation with the number of commands is also found to be strong (ρ = 0.45, $p-$value $=$ 0.05). These quantifications of pupil diameter variations with performance measures have the potential to play a vital role in advancing the HRI systems as they can be used to predict the performance variation in real-time so that the HRI can be more responsive to the user's changing mental states, a key requirement for the practical usability and acceptability of such systems as assistive technologies.**

***Index Terms*—Pupil diameter, Eye-tracking, Assistive device**

I. INTRODUCTION

This paper presents a novel user interface based on eye-tracking technology designed for individuals facing severe mobility challenges, particularly those with limitations in both upper and lower body mobility. The primary objective is to offer a control mechanism for their assistive devices, such as powered wheelchairs or wheelchair-mounted robotic arms, to facilitate their participation in daily activities. Besides the physical capabilities of the assistive systems, the user interface and the interaction experience with the technology play crucial roles in shaping the overall user experience [1].

Despite considerable research into autonomous assistive technologies, their widespread adoption among the target users remains limited. Recognizing the importance of understanding dynamic changes in mental states during interactions with assistive technologies, it is vital to explore the potential for adapting to variations in the user's mental state. A deeper understanding of these dynamic states can significantly improve the effectiveness and responsiveness of assistive technologies [2].

Various assistive and autonomous wheelchairs have been introduced, employing a variety of methods to interact and receive commands from users. Common methods include electroencephalogram (EEG), electromyogram (EMG), joystick, eye-tracking, and graphical user interfaces (GUIs) [3]. Some systems utilize a single method, while others combine multiple methods to enhance performance and accessibility. RoboChair [4], introduced in 1997, is an autonomous wheelchair equipped with both a joystick for motion control and a GUI. A fusion of eye-tracking and inertial-measurement-units (IMUs) was used in a previous study by [5] for movement intention detection for safer operation of lower-limb exoskeletons for mobility assistance while Meena *et al.* developed a gaze-controlled virtual keyboard for assisted learning of children with dyslexia using eye-tracker among other modalities [6].

EEG signals are also used as a control modality in numerous autonomous wheelchair projects, such as in [7]. Another notable project, [8] utilized a hybrid brain-computer interface (BCI) method incorporating EEG and electrooculogram (EOG) to control a wheelchair-mounted robotic arm system. However, the need for repeated calibration, higher user fatigue, and inter-subject variability limits its practical uses.

Therefore it is more practical to use modalities such as an eye-tracker for controlling mobility assistance devices for people with both upper and lower limb disability. In [9], a wheelchair-mounted robotic arm controlled by a screen-based eye-tracker was introduced. The interface includes buttons and options for controlling both the wheelchair's movements and the robotic arm. Similarly, [10] features a similar combination of a robotic system, though with different control systems.

Moreover, eye-tracking finds various applications, particularly in psychological studies, offering insights into human cognition and points of interest [11], [12]. Mental workload estimation has also been studied concerning pupil dilation in many works [13], [14]. Guo *et al.* highlighted the influence of time pressure and latency on mental workload, aside from factors such as task specificity, which is crucial in the context of assistive devices. It was also noted that gaze information, including pupil diameter, plays a significant role in such mental workload estimation [15]. Rather than using raw pupil diameter, researchers have also experimented with the relationship between power and frequency at the mean frequency to calculate cognitive load [16].

Mental activities are known to influence pupil diameter; for example, increased mental load typically corresponds with pupil dilation [17], while fatigue can lead to pupil constriction [18]. Additionally, pupil diameter has been used to differentiate cognitive processes, such as decision-making and sustained focus, highlighting its role in understanding user engagement and cognitive state [19].

Therefore, in this study, we asked whether the performance variation of a user during a human-robot interaction (HRI) task can be captured and quantified through the variation in pupil diameter as an indirect estimate of mental workload, so that future HRI systems can be responsive to the changing mental state of the user. We investigate the variation in pupil diameter as an indirect measure of mental workload during human-robot interaction tasks. We analyze how this variation correlates with various performance parameters, such as lap completion time, number of commands issued, and average time taken to select commands. Additionally, subjective measures of user mental workload during task performance, as well as the usability and acceptability of the system, are evaluated using standard scales. The changes in pupil diameter as the user controlled the motion of a mobile robot on a pre-defined trajectory through their eye-gaze have been discussed in this paper concerning various performance parameters such as lap completion time, number of commands, and average time to select the commands. Subjective measures of the mental workload of the users performing the task and the usability and acceptability of the system were also taken using standard scales.

Following this introduction, the methodology section provides an in-depth look at the system design and experimental protocol that underpin this study. The subsequent sections delve into the results and discussion, highlighting key findings through detailed graphs and analyses. The paper concludes with a summary of insights and considerations for future work.

II. Methodology

A. System Description

The primary equipment utilized in this project is the Tobii Pro nano eye-tracking device, boasting a 60Hz sampling rate. This eyetracker is monitor-mounted and connected to a PC for seamless integration. The User Interface displayed on the monitor (Fig. 1) includes a live video stream at the center of the screen, providing an enriched view of the path. Additionally, the interface incorporates four control buttons dedicated to robot movements. The inclusion of this live video feed enhances user awareness of the path and task without causing distraction from the available control options.

The system operates by moving the mouse cursor to the corresponding gaze point. Upon hovering over a button, a timer initiates. If the user maintains their gaze on the button for the dwell time, currently set to 2 seconds, a click is registered. Moving the cursor before the completion of the dwell time terminates the timer. This mechanism enables the generation of continuous commands, allowing the timer to make a click at regular intervals as long as the cursor remains on the button. This feature proves advantageous, particularly for issuing continuous commands, especially during the navigation of long, straight routes. We chose this dwell time of 2 seconds to align with the average duration of our robot's actions, aiming to prevent conflicts between user commands and robot actions while ensuring a smooth interaction experience.

Throughout the experiment, the system records the gaze point on the screen and the pupil diameter for both eyes at a 60Hz sampling rate, with corresponding timestamps. Additionally, user commands and their timestamps are recorded, providing a comprehensive dataset for analysis.

The experiment employs a Turtlebot Burger as the mobile robot, establishing wireless communication with the PC. The robot publishes its motor speed based on generated commands, enabling seamless movement along the predefined path. This path, printed on A0 papers and covering approximately 2 square meters, provides intentional guidance with a balanced distribution of right and left turns, including 45 and 90-degree turns tailored to match the robot's capabilities. The robot's movements are predefined as steps, with each command instructing the robot to move 10 cm or turn 45 degrees.

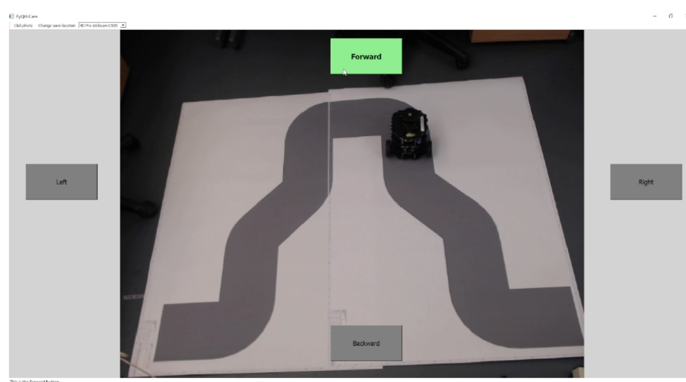

Fig. 1. A screenshot of the user interface and the path design.

B. Experiment Protocol

The experiment was conducted indoors under controlled lighting conditions. Participants were seated in adjustable computer chairs and instructed to customize the chair and monitor settings based on their preferences, following the eye-tracker's posture guide. Before each round, the system underwent calibration for each user, ensuring an optimal distance and posture with the eye-tracker. Then participants were advised to maintain the prescribed posture throughout the experiment recording.

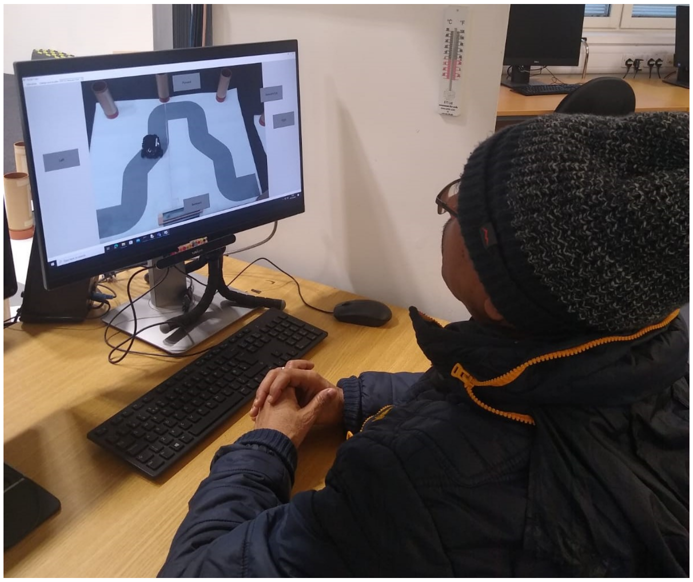

Fig. 2. A participant using the device during the experiment.

Participants were instructed to complete two rounds of the designated path, with all eye-tracker readings, time stamps, and commands recorded in a file. Performance analysis was conducted based on these recordings. The first measure considered was the number of corrections made during the experiment, calculated by counting direct changes from left to right, forward to backward, and vice versa. Lap completion time and the average command selection were also measured to assess performance. Additionally, changes in pupil diameter were recorded during the experiment at a 60Hz sampling rate for both eyes.

During participant recruitment, factors such as experience with eye-tracking software and biologically controlled software were controlled by ensuring a diverse pool of participants with varying levels of familiarity. The NASA Task Load Index (NASA TLX) [20] served as a crucial tool to capture participants' self-reflective feedback on task load, complemented by additional questions about their overall system experience. Participants rated their task load across six domains: frustration, effort, performance, temporal demand, physical demand, and mental demand.

Furthermore, in conjunction with the NASA TLX, participants provided insights into the user acceptability and usability of the system. To evaluate this aspect, the Questbec User Evaluation of Satisfaction with Assistive Technology (Quest 2.0) and Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaires were administered. The experiment also incorporated acceptability and usability questionnaires to gather feedback on the interface and its potential enhancements. Throughout the experimental phase, valuable recommendations and suggestions were gathered to guide future improvements in the adaptive system.

C. *Participants information*

At this stage, data were collected from 19 participants, involving healthy adults across diverse age groups who were invited to participate in the experiment. Approximately 21% of the participants were over the age of 60. 42% of participants wore glasses during the experiment. Ethical approval for the experiment was granted by the University of Essex Ethics Subcommittee 3 (ERAMS Reference code: ETH2223-2300) and all subjects gave informed consent.

III. Results and Discussion

To gain a comprehensive understanding of the interaction and its outcomes, we are focused on analyzing the overall changes and characteristics of pupil dilation, along with their corresponding performance measures, for each participant. Additionally, we are examining the questionnaire results.

Performance measures include total lap competition time, which is the total time that participants took to complete the lap from the start to finish. Another performance measure is the Average selection time, which is the average of duration between the commands generated by the user in the lap. The number of commands issued during the lap is also used as a performance indicator throughout the experiment. The number of commands serves as a critical metric reflecting the efficiency, cognitive load, error rate, and time efficiency of user interactions with the system. Higher command counts may indicate challenges in task completion, increased cognitive load, higher error rates, and longer completion times.

The Gaussian distribution reveals that 50% of participants exhibit more dilation in the left pupil rather than the right, while approximately 37% show dilation in the right pupil during the recordings. Additionally, there were instances where participants displayed overlapping or similar pupil diameters.

Table. I provides an overall view of the data collected during the experiment. The 'Minimum' column displays the minimum pupil value recorded among the participants, and similarly, the 'Maximum' column shows the maximum value. The 'Mean' columns present the average mean values for each eye during the rounds, and the 'Standard Deviation' column represents the average of standard deviations.

TABLE I
Minimum and Maximum pupil in the data, along with Average Mean and average Standard deviation of each pupil in the rounds.

	Minimum	Maximum	Mean	Standard deviation
Round 1 Left Pupil	0.8998	5.0078	3.2477	0.1808
Round 1 Right Pupil	0.9412	5.2929	3.1883	0.1871
Round 2 Left Pupil	0.9951	5.0728	3.2	0.1748
Round 2 Right Pupil	0.8621	5.2542	3.1533	0.1799

Fig. 3 illustrates key performance metrics across different rounds and this figure presents the Lap completion time, the total number of commands issued, and the average selection time in each case. The box plots reveal the distribution and variance within each round and provide insight into trends or changes in performance as rounds progress. By comparing

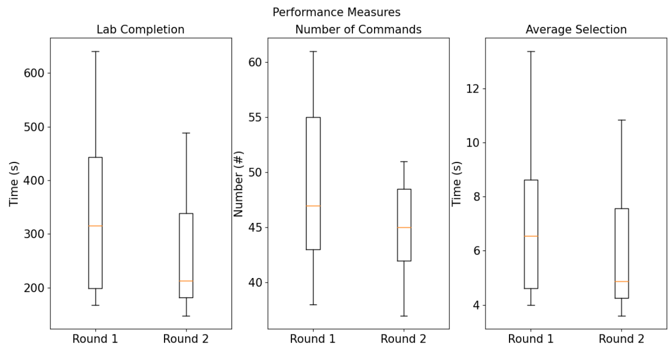

Fig. 3. The box plots representing the Lap completion time, total number of commands, and average selection time for each round.

these metrics, we can observe variations in user efficiency and responsiveness, and it indicates potential improvements or fatigue factors that might affect task completion over time. After the first round, participants demonstrated noticeable improvement across all key performance metrics—Lap completion time, total number of commands, and average selection time— which is indicating an adaptation period where they became familiar with the system's functionality. This learning effect suggests that initial exposure to the interface helps users internalize the command structure and interaction flow and as a result, it leads to more efficient task performance in subsequent rounds, faster lap completions, fewer command inputs, and reduced time for each selection. This trend highlights the intuitive nature of the system, as users can quickly gain proficiency with minimal rounds, a positive sign of usability and user-friendly design.

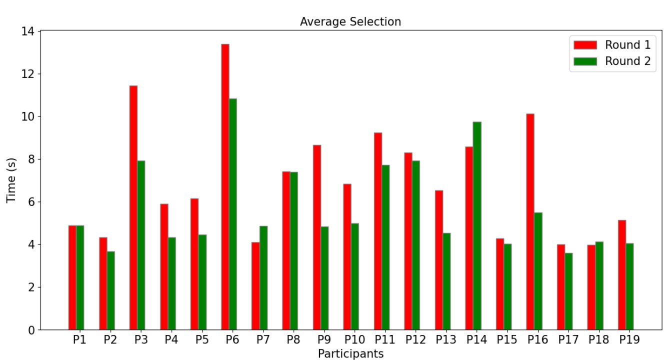

Fig. 4. The average selection time for each round for each participants.

Fig. 4 shows the average selection time during the rounds for each participant. The difference the time between the rounds has been used as a performance measure in Fig. 10. The performance measures, as illustrated in Figure 3, exhibit overall improvement in Round 2. A noticeable decrease is observed in lap completion time, average selection time, and the number of commands.

However, for around 15% of participants there has been a slight increase in the average selection time (Fig. 4) while for lap completion time (Fig. 5) the figure is around 10% of

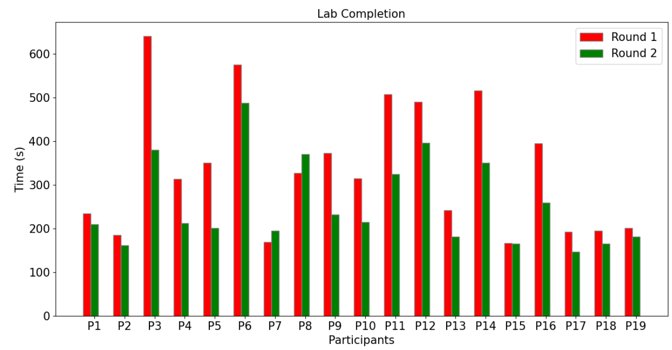

Fig. 5. The total lap completion time for each round for each participants.

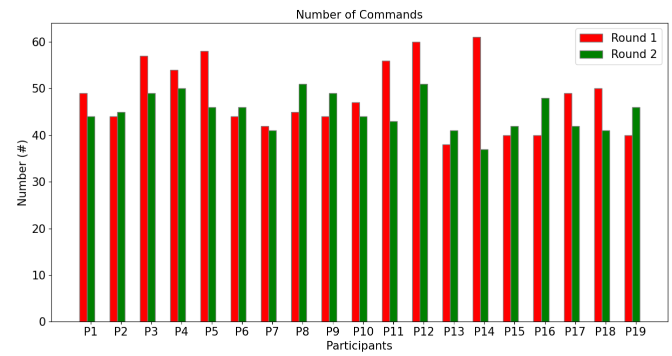

Fig. 6. The graphs shows the total number of commands generated by participants during each round.

participants. Around 42% of participants show a slight increase in the total number of commands in round 2 however the improvement for the rest of the participants has been more significant (shown in Fig. 6).

Fig. 6 shows the total number of commands that have been produced by participants during the rounds. The difference between the value in round 2 and round 1 has been used as an indication of performance measure in Fig. 8.

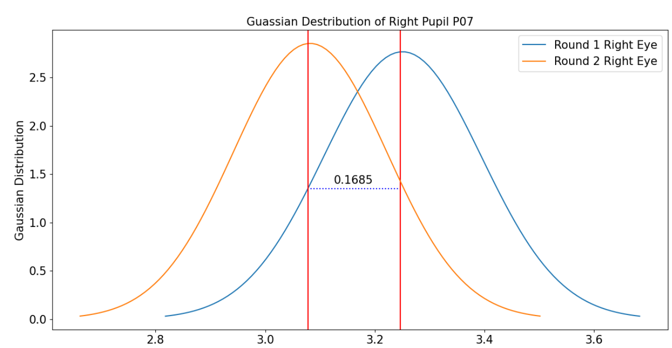

Fig. 7. This plot shows the gaussian distribution of P07's Right pupil during round 1 and 2. The difference between the mean values in the Gaussian distribution is shown in the plot.

Fig. 7 shows an example of gaussian distribution in right pupil data for participant number 7. The red lines show the peaks of the Gaussian distribution (mean). The difference between these lines is used to calculate the difference between means which has been used in Figs. 8-9.

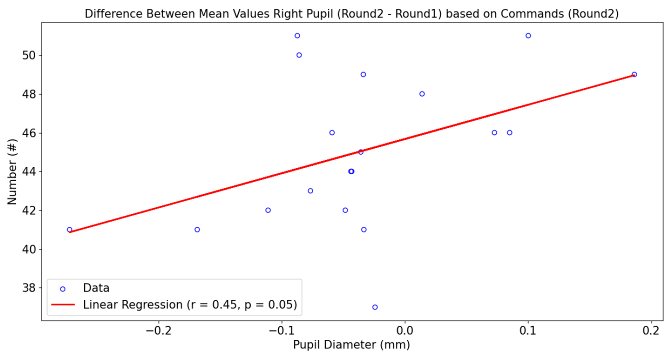

Fig. 8. scatter plot and linear regression analysis will be conducted to examine the difference between the mean pupil sizes of Round 2 and Round 1, correlating these differences with number of commands selected in round 2.

Fig. 9 depicts a linear relationship between the difference in the mean of the Gaussian distribution for right pupil diameter from Round 1 to Round 2 and the lap completion time of Round 2. The graph illustrates that individuals exhibiting increased dilation in their right eye tend to spend more time completing the second lap. This finding underscores a potential correlation between pupillary responses and task performance.

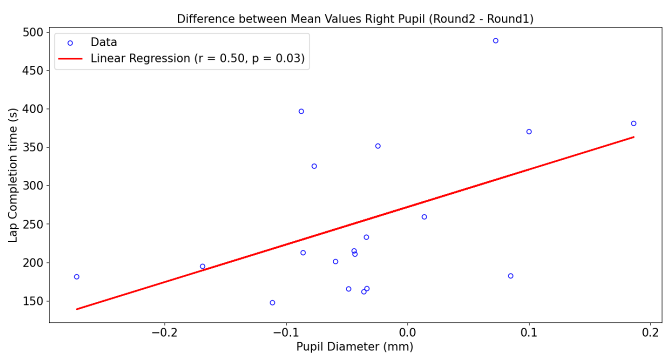

Fig. 9. A scatter plot and linear regression analysis is conducted to examine the difference between the mean pupil sizes of Round 2 and Round 1, correlating these differences with Lap 2 completion time.

Similarly, a comparable outcome is evident in Fig. 8, wherein the difference in mean right pupil diameter is examined concerning the average selection time between commands. The graphical representation suggests that participants with an elevated right pupil diameter exhibit longer pauses between issuing commands. This observation may provide further insights into the potential role of pupillary dynamics in influencing the temporal aspects of participants' interactions with the system.

It is interesting to note that approximately 10% of participants experienced an increase in lap completion time during Round 2 compared to Round 1. Additionally, around 21% of participants demonstrated an extended average duration between their commands in Round 2 as opposed to their performance in Round 1. These findings underscore the variability in participants' task performance, shedding light on individual differences in adapting to the system or task demands.

Analysis of the collected data from each round reveals an interesting trend which is the participants with a lower mean pupil diameter tend to spend more time completing the lap, while those with a higher mean pupil diameter complete the lap sooner. However, these participants didn't report a high mental load or fatigue in the questionnaires.

In Fig. 10, the difference between the average time spent between each command and the variation in the total number of commands between Round 2 and Round 1 is illustrated. Importantly, participants who demonstrated a slower pace in issuing commands during the second round tend to generate fewer commands in Round 2 as opposed to Round 1. This observation implies a potential relationship between the participants' command generation speed and the overall number of commands executed, emphasizing the increased effort and precision invested in decision-making during the second lap.

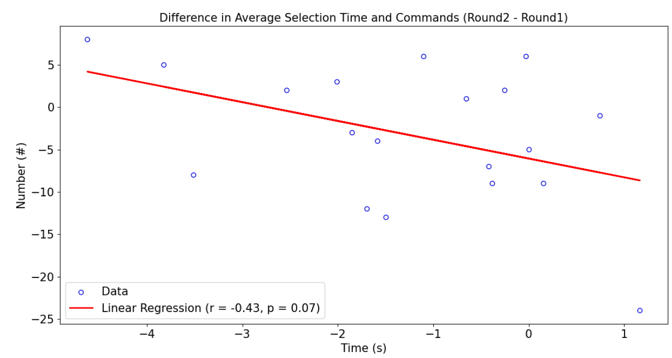

Fig. 10. A scatter plot projecting the relation between The difference between the average selection time and the difference in the number of total commands generation during R2 and R1.

A. *Subjective measures*

In terms of participant feedback, a predominant concern has emerged regarding the predefined steps for robot movement and the use of a mobile robot instead of an actual wheelchair. Some participants have expressed that the fixed dwell time of 2 seconds is perceived as excessively long. Additionally, participants reported blinking less while concentrating on the buttons, leading to tiredness.

While all demands (physical, mental, and temporal) were reported to be low, physical demand was perceived as harder by most participants. The primary causes of high physical demand were reported as eye tiredness and maintaining posture. The challenge of distinguishing the robot's direction was attributed to mental demand. These valuable insights will be considered while changing the mobile robot paradigm into the actual powered wheelchair paradigm with an adaptive HRI interface. Additionally, we will implement a new adaptive setting to directly address these concerns and improve the overall user experience.

Regarding the interpretation of the results, tiredness and increased mental demand contribute to decreased performance outcomes. However, an alternative perspective is proposed based on observations of lower button selection time, total completion time, number of commands, and mean pupil diameter in Round 2. This suggests that participants may have become more familiar with the task and its requirements, resulting in reduced mental workload and improved performance. The decrease in lap time and pupil diameter could be attributed to participants' increased efficiency and reduced cognitive effort in Round 2, possibly due to learning effects or task familiarity developed during Round 1. Therefore, while tiredness and mental demand may still play a role, the influence of participants' familiarity with the task should also be considered in interpreting the results.

Consistent with previous findings, this experiment revealed similar trends in pupil diameter. During Round 1, when most participants interacted with the system for the first time, pupil diameter was generally larger, indicating a higher cognitive load. By Round 2, the diameter decreased, likely reflecting reduced cognitive load due to increased familiarity with the system and some degree of fatigue. Interestingly, participants who improved their performance metrics tended to show increased pupil diameter in subsequent rounds, whereas those who reported feeling fatigued exhibited lower pupil diameters.

B. Future Work

In our forthcoming study, we plan to delve deeper into the analysis of command generation and its correlation with pupil response during decision-making, alongside examining blinking frequency. Moreover, the future study will encompass a more comprehensive analysis of the data, with a specific focus on scrutinizing the intervals between each command to elucidate the intricacies of the decision-making process. Additionally, the investigation will center on the detailed examination of pupil dilation changes at each step, providing a nuanced understanding of the cognitive dynamics involved.

IV. Conclusion

The paper primarily highlights the impact of the pupil diameter variation on the performance of an HRI system. The results show for the first time that there is a strong correlation between the difference in the Gaussian mean of the pupil diameter between the two rounds of the same HRI task on the various performance parameters of the HRI. The importance of the study lies in the fact that such measures of pupil diameter variation may act as the indirect measure of the users' mental state and can be used to adapt the HRI system parameters in real-time for optimal performance which can play a major role in advancing the usability and acceptability of HRI based assistive technologies in the society.

References

[1] C. Guo and E. Sharlin, "Exploring the use of tangible user interfaces for human-robot interaction: a comparative study," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 121–130.

[2] M. F. Hinss, A. M. Brock, and R. N. Roy, "Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces," *Frontiers in Neuroergonomics*, vol. 3, 2022.

[3] J. Leaman and H. M. La, "A comprehensive review of smart wheelchairs: Past, present, and future," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 486–499, 2017.

[4] G. Pires, N. Honorio, C. Lopes, U. Nunes, and A. Almeida, "Autonomous wheelchair for disabled people," in *ISIE'97 Proceeding of the IEEE International Symposium on Industrial Electronics*. IEEE, 1997, pp. 797–801.

[5] Y.-X. Liu, Z.-Y. Wan, R. Wang, and E. M. Gutierrez-Farewik, "A method of detecting human movement intentions in real environments," in *2023 International Conference on Rehabilitation Robotics (ICORR)*, 2023, pp. 1–6.

[6] Y. K. Meena, H. Cecotti, B. Bhushan, A. Dutta, and G. Prasad, "Detection of dyslexic children using machine learning and multimodal hindi language eye-gaze-assisted learning system," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 1, pp. 122–131, 2023.

[7] A. Úbeda, E. Iáñez, and J. M. Azorín, "Shared control architecture based on rfid to control a robot arm using a spontaneous brain–machine interface," *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 768–774, 2013.

[8] Q. Huang, Z. Zhang, T. Yu, S. He, and Y. Li, "An eeg-/eog-based hybrid brain-computer interface: application on controlling an integrated wheelchair robotic arm system," *Frontiers in neuroscience*, vol. 13, p. 1243, 2019.

[9] M. S. H. Sunny, M. I. I. Zarif, I. Rulik, J. Sanjuan, M. H. Rahman, S. I. Ahamed, I. Wang, K. Schultz, and B. Brahmi, "Eye-gaze control of a wheelchair mounted 6dof assistive robot for activities of daily living," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, pp. 1–12, 2021.

[10] M. Iskandar, G. Quere, A. Hagengruber, A. Dietrich, and J. Vogel, "Employing whole-body control in assistive robotics," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5643–5650.

[11] P. Majaranta and A. Bulling, "Eye tracking and eye-based human–computer interaction," *Advances in physiological computing*, pp. 39–65, 2014.

[12] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2176–2184.

[13] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman, "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the 2010 symposium on eye-tracking research & applications*, 2010, pp. 141–144.

[14] M. I. Ahmad, I. Keller, D. A. Robb, and K. S. Lohan, "A framework to estimate cognitive load using physiological data," *Personal and Ubiquitous Computing*, pp. 1–15, 2020.

[15] 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.

[16] R. Gavas, D. Chatterjee, and A. Sinha, "Estimation of cognitive load based on the pupil size dilation," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 1499–1504.

[17] P. van der Wel and H. van Steenbergen, "Pupil dilation as an index of effort in cognitive control tasks: A review," *Psychonomic Bulletin & Review*, vol. 25, no. 6, pp. 2005–2015, 2018.

[18] Y. Wang, G. Naylor, S. E. Kramer, A. A. Zekveld, D. Wendt, B. Ohlenforst, and T. Lunner, "Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task," *Ear Hear*, vol. 39, no. 3, pp. 573–582, May/Jun 2018.

[19] P. Azizinezhad, H. Ghonchi, and A. Chowdhury, "Pupil diameter classification using machine learning during human-computer interaction," in *2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*. IEEE, 2024, pp. 1–6.

[20] [Online]. Available: https://humansystems.arc.nasa.gov/groups/TLX/