# A.Eye Drive: Gaze-based semi-autonomous wheelchair interface

Mahendran Subramanian[+], Noyan Songur[+], Darrell Adjei, Pavel Orlov, & A Aldo Faisal[*], *Members, IEEE*

***Abstract*— Existing wheelchair control interfaces, such as sip & puff or screen based gaze-controlled cursors, are challenging for the severely disabled to navigate safely and independently as users continuously need to interact with an interface during navigation. This puts a significant cognitive load on users and prevents them from interacting with the environment in other forms during navigation. We have combined eyetracking/ gaze-contingent intention decoding with computer vision context-aware algorithms and autonomous navigation drawn from self-driving vehicles to allow paralysed users to drive by eye, simply by decoding natural gaze about where the user wants to go: A.Eye Drive. Our "Zero UI" driving platform allows users to look and interact visually with at an object or destination of interest in their visual scene, and the wheelchair autonomously takes the user to the intended destination, while continuously updating the computed path for static and dynamic obstacles. This intention decoding technology empowers the end-user by promising more independence through their own agency.**

## I. Introduction

Existing wheelchair control interfaces developed for severely-disabled patients, do not yet provide sufficient independence in urban mobility. Sip & puff, screen-based gaze controlled cursors or the EEG-based navigation systems require the user to interact with an interface constantly, while only allowing the user to provide low-level directional commands to navigate in urban settings. This creates a need for an interface less system capable of higher-level intention decoding while reducing cognitive load to enable severely-disabled to move in the urban continuum independently. Decoding natural human behaviour to predict future actions for prosthetic, orthotic and assistive technology has been shown to operate at benchmark levels or above [1-4].

Eye-tracking has been widely studied as an interface for enabling technologies for the severely-disabled [5,6]. Previously proposed systems allowed the user to steer the wheelchair using only their eyes, providing continuous directional commands through their gaze [7-9]. Some of these systems consist of on-screen buttons on a video scene feed, with the user gazing at the buttons to navigate [8,10]. We have previously proposed a gaze-driven control system based on natural gaze behaviour [1,11], where the user's intention is decoded in real-time from natural eye-movements. However, to eliminate the need for constant interaction with the interface, advanced navigation techniques mediated by inputs derived from natural user behaviour is the optimal solution for this problem [11]. Taking into account of the 3D endpoint decoding advances [4,12-16] here, we take a Human-in-the-AI loop approach by incorporating autonomous driving technology with gaze-based intention decoding. By doing so, our user driving platform can input the destination commands via decoding their natural gaze, while the AI algorithms take care of the navigation. This significantly reduces the cognitive load and eliminates the need to interact with a screen-based user interface (hence, our labelling of Zero UI).

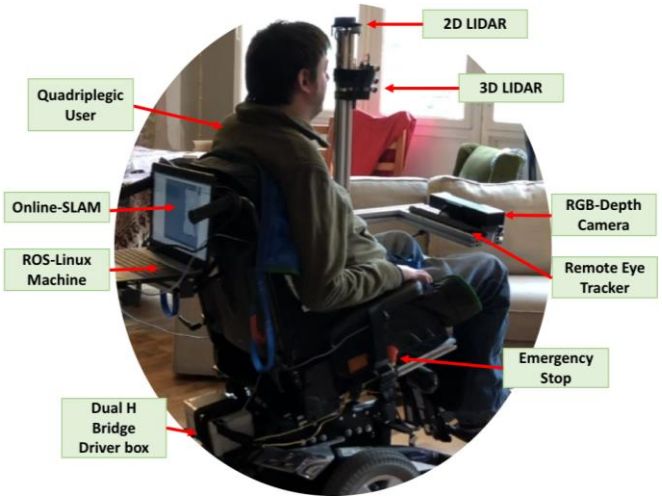


**Figure 1. A.Eye Drive wheelchair with a quadriplegic user.**

## II. Material and Methods

### A. System Architecture

Our system consists of an RGB-D camera (Kinect v2 – 512 x424 @ 30fps; 0.5-4.5 m range; 70/60° angle of view), eye-trackers (Tobii EyeX Controller and SMI), and a 2D lidar and a homebuilt 3D lidar fitted on an electric wheelchair (Invacare, UK) using Bosch-Rexroth bars as shown in Fig. 1. The wheelchair control system was replaced with a regenerative dual channel motor driver (Dimension Engineering, U.S.A) to control the wheelchair's motors. The driver board was connected to the Linux machine via USB A to micro USB B.

### B. Gazeinformatics-based intention decoding

Differences in gaze fixations between interactions and viewing were studied to build an intention decoding engine. This AI module decodes the users' high-level intention. During the experimental session, each subject performs tasks with and without interactive intentions, and the gaze tracking setup records the corresponding eye movements. This data was used to train a Machine Learning based classifier. Our intention

*This work was supported by EPSRC (EP/N509486/1) and a Toyota Mobility Foundation Discovery Prize. We acknowledge the support of our end-user volunteers TN and PM.

All authors are at the Brain & Behaviour Lab, Imperial College London, London, UK (Corresponding author aldo.faisal@imperial.ac.uk). + denotes equal contribution.

decoding engine is explained in [4], and the experiments are shown in [1].

### C. 3D Gaze-based destination

To determine the 3D coordinates of the user's intended destination, we combine remote eye-trackers with an RGB-D camera. The remote eye-trackers, placed at a distance of 60 cm from the user, can track users gaze and provide 2D screen coordinates of the gaze on a 60 cm x 34 cm display at a rate of 60Hz. To convert this 2D information into 3D coordinates, we overlay 2D gaze-point of the user on the 3D point cloud map of the environment reconstructed by the RGB-D camera. The alignment of these two-different field-of-views is carried out by two calibration steps. In the first calibration step, a calibration plane (60 cm x 34 cm) consisting of 7 calibration points is placed 1 cm above the eye-trackers. The calibration is carried out using the software provided by the sensor manufacturer. During the process, the user fixates on the seven calibration points in order and calibration is validated using nine points defined on the calibration plane. In the second calibration step, the 2D scene of the eye-trackers is overlaid on the depth image from the RGB-D camera. Calibration plane consisting of equally spaced 12 ping-pong balls glued on the calibration plane (120 cm x 60 cm). The calibration plane was placed approximately 2m from the RGB-D sensor. The subjects were told to fixate on each ball in order, while the coordinates of the gaze-point and the ball coordinates in the depth image were recorded. Using these points, the projective transformation matrix was calculated using the fitgeotrans function in MATLAB (MathWorks R2016b). The calibration quality was tested using the absolute pixel error between the reference coordinates on the depth image and the transformed 2D gaze coordinates. To test the accuracy of the 3D gaze-point estimation, a grid of equally spaced 18 points, covering a workspace of 1m (width) x 1m (height) x 3.5m (depth) were predefined. The grid contained two height layers (15 cm and 75 cm above the ground). These heights were selected as they were assumed to represent the most manipulative areas in daily life activities (standard dinner table height ~ 76 cm) and wheelchair navigation (floor conditions). Five subjects were used to test the 3D gaze-point estimation. The calibration process was carried out for each subject according to the steps as mentioned above. The calibration was repeated until the average pixel error during the alignment step was less than 5 pixels. The subjects were told to fixate on a 2 cm radius filled circle drawn on a 60 cm x 30 cm box, which was placed at predefined locations in randomised order. The 3D coordinates at the gaze point were recorded for 2 seconds while the user fixated on the filled circle on the box. A total of 30 testing points for each subject (15 locations at each height level) were recorded and analysed. Absolute errors in three dimensions, as well as the Euclidean distance error, are reported in the results section.

### D. Autonomous navigation and obstacle avoidance architecture

The AI architecture (Fig. 2.) was built on a 2D online-SLAM algorithm to both create the map and perform localisation in real-time. The wheelchair localisation was improved by using the odometry obtained from the online-SLAM instead of the wheel odometry. The loop-closure technique employed by the deployed mapping and localisation algorithm provides a 2D SLAM that is computationally inexpensive but efficient [17]. This also allows the wheelchair to operate using a single 2D lidar. The nodes entirely native for this architecture are `cmd_to_wheelchair_drive`; `natural_gaze_intention_decoder`; and the `gaze_monitor`. We developed these nodes using C, Matlab, and Python respectively for ROS. Configuration files were then written for the remaining nodes to integrate them with the hardware and system architecture. `Navigation_stack` was utilised for obstacle detection and path planning. The `costmap_common_params` was configured to have a maximum `obstacle_range` of 2.5m (the furthest that wheelchair could be from an obstacle in the testing environment) and an `inflation_radius` of 0.7 m to incorporate the size of the wheelchair base.

### E. Integration of Autonomous Navigation architecture with Gazepoint estimation

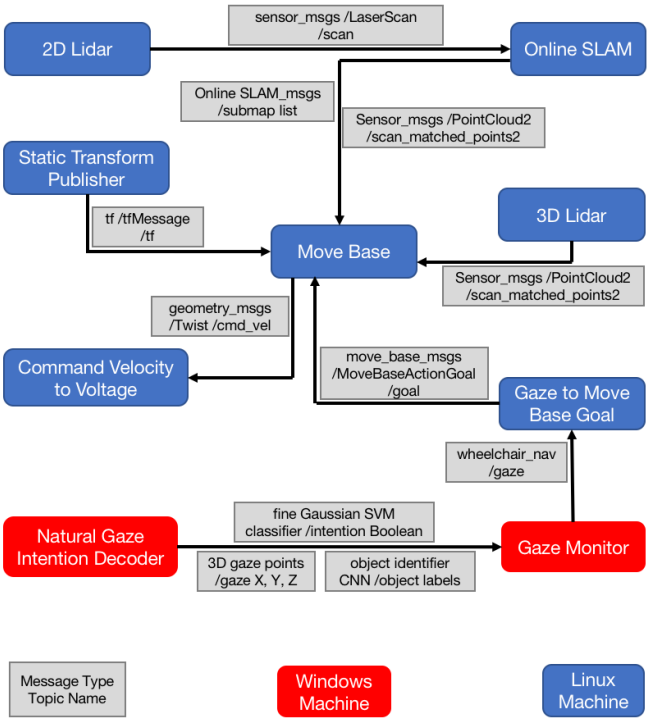


**Figure 2. Diagram showing the A.Eye Drive system architecture. The elements in red were hosted on a windows machine, and the elements in blue were hosted on a Linux machine. Grey elements are ROS topics.**

The autonomous navigation architecture was then incorporated with the gaze-based destination commands. The `natural_gaze_intention_decoder` node publishes `predictor_msgs` and `object_identifier_msgs`, i.e. whether the user intends to interact with the object of interest within the field of view. The `gaze_monitor` node subscribes to these messages, and the gaze-based commands are published in ROS message (`wheelchair_nav/gaze`) received from Windows client. The `gaze_to_move_base_goal` node subscribes to this message and publishes the goal pose as a `move_base_msg` to the `move_base` node. Once a path to the goal has been computed, the required velocity commands are sent to the `cmd_vel_to_wheelchair_drive` node which is the driver for the wheel motors. All of these processes were achieved

online. Equations 1 and 2 were used to determine the wheel speeds from the wheelchairs' linear and angular velocities represented. This kinematic model was found to be suitable for this 6-wheeled wheelchair (2 main, 4 castor wheels).

$$V_{right} = \left(\frac{\omega}{4}\right) + V_x \quad (1)$$

$$V_{left} = 2 . V_x - V_{right} \quad (2)$$

where $V_{right}$ is the right wheel velocity, $V_{left}$ is the left wheel velocity, $V_x$ is the wheelchairs' linear velocity, and $\omega$ is the wheelchairs' rotational velocity. These velocities were prevented from assuming high values that might put the user at risk; however, an emergency stop was added as a precaution.

## III. Results and Discussion

### A. 3D Gaze-based Destination Estimation

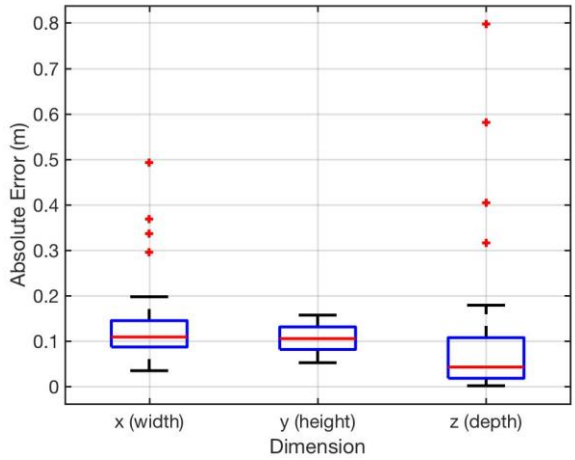


**Figure 3. Absolute errors in three dimensions. Depth dimension contains the highest variation due to missed targets resulting in overshoot.**

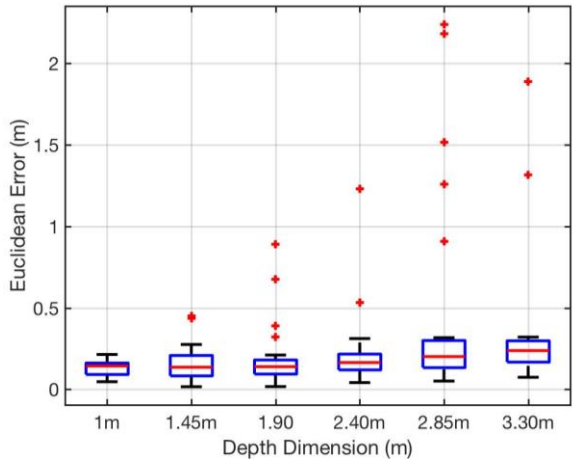


**Figure 4. The Euclidean error and error variation increase with the depth dimension. Depth distances are measured assuming the RGB-D camera position as z = 0.**

For validating the alignment of the eye-tracker scene and the RGB-D point cloud map, the calibration accuracy was measured in units of absolute pixel error between reference coordinates on the depth image and the transformed 2D gaze screen coordinates. In result of the experimentation with 5 subjects, the average absolute pixel error was 1.38px ± 0.46px and 2.69 px ± 0.49px in x and y dimensions respectively. Fig. 3. shows the absolute errors obtained in three dimensions from the grid experiment performed on five subjects. The overall Euclidean error was 25.0 cm ± 20.2 cm. The errors in 3D destination estimation were 14.2 cm ± 10.3 cm in x dimension, 10.5cm ± 2.94 cm in y, and 11.3 cm ± 18.3cm in z dimension. As the largest variation is observed in the depth dimension, the Euclidean error was investigated across the depth dimension and is plotted in Fig. 4. to show the increase in the Euclidean error as the target point distance increases. As seen from the boxplot, the magnitude and the frequency of outliers increase with the depth dimension. This is due to the target object occupying a smaller number of pixels in the image, and therefore, when missed, results in significant overshoot values. This also shows the sensitivity of the alignment procedure to the depth dimension. Further work will explore 3D calibration procedures for better alignment of sensor field-of-view. Overall, the average Euclidean error was within the box dimensions (60cm x 30cm) used as the target object, proving that the destination estimation needs improvement for more accurate destination decoding, however still practical in the 3.5m (width) x 1m (width) x 1m (height) workspace.

### B. Autonomous Wheelchair Performance Evaluation

To find the optimal specifications for safe use of the wheelchair, we investigated three parameters: planner frequency, position tolerance and orientation tolerance. These parameters were first evaluated in three different tasks by measuring the time it takes the wheelchair to reach the destination. For the first task, the wheelchair was instructed to travel 4m ahead in the x-direction. In task 1 (n=5), the wheelchair was unobstructed by any obstacles. For task 2 and 3 (n=5), the wheelchair was made to move 5m ahead in the x-direction. In task 2, a static obstacle of height 2.5m and width 0.5m was placed 2.5m ahead of the wheelchair; and in task 3, the static obstacle remained in the same position and another person, acting as a dynamic obstacle, of height 2m and width 0.5m was instructed to walk by the wheelchair and stand near the static obstacle. It was found that planner frequency of 5Hz [out of 20, 10, 5 Hz], position tolerance of 0.13m (13 cm) [out of 100, 50, 25, 13, 6 cm] and orientation tolerance of 0.06 radians (3.4 degrees) [out of 1, 0.5, 0.25, 0.125, 0.0612, 0.0306 radians] gave the best performance. As these parameter values corresponding to the lowest travel times. Based on these results, the parameters for the system architecture were optimised, and task 4 (n=6) was performed. In task 4, the goal for the autonomous wheelchair was defined as to move from A to B, within different dynamic environment scenarios, i.e. the number of static and dynamic obstacles and their position were changed for each run to simulate different routes but within the same room for the same start and end goal coordinates. Task 4 results - Time taken to move from A to B varied for different routes. However, the autonomous wheelchair was able to detect both static and dynamic obstacles with perfect accuracy with using optimised parameters. Next, a questionnaire about the comfort level was provided to 3 volunteers (Two quadriplegic and one paraplegic) who showed interest in evaluating our system further to a demonstration. State of the passenger during autonomous navigation was recorded. Based on the observations, the torque during turning was reduced; voltage output to the wheel motors was adjusted based on the user's weight; a moving average window filter was added to the

`cmd_vel` to the `wheelchair_drive` node to smoothen wheelchair transmission.

### *D. A.Eye Drive Evaluation*

The 3D gaze-based destination defining module was then integrated with the wheelchair to test the semi-autonomous functionality. The wheelchair was positioned between two static obstacles of height 2.5m and approximately 2m ahead of the chair. The wheelchair participant's (subject) intention to get to the chair that the second participant occupied (on the far left) was decoded. This was done to mimic the potential use-case when a wheelchair user would like to pull up beside a table to engage in an activity at the table, e.g. a luncheon or game. The task was carried out (n=3), and it took 29±28s for the `gaze_monitor` to detect the intention and publish it. This variance was due to intention detection node running at a much higher rate than the rate the `gaze_monitor` could publish. This meant that many intentions were missed and hence the system has to be as responsive as the system presented in [1]. When an intention has successfully detected the chair was able to navigate to the destination as illustrated in Fig. 5. The navigation goal set was locked until the wheelchair arrived at the destination, ignoring all other intentions during the navigation. For this task, the final orientation of the wheelchair was always set to its initial orientation. This meant that the wheelchair would stop, facing the table, rather than beside it as desired. The best average lap time yielded for task 1 was (23s ± 1s. For the range of goal distances that were tested, the wheelchair typically chose `cmd_vel` values between 0.28m/s and 0.37m/s. At the lower end of this spectrum, 0.28m/s, we would expect the lap time for a 4m course to be approximately 14s. The wheelchair was capable of localising itself efficiently for velocities within the range mentioned above. Future work should involve improving system architecture to achieve high throughput intention decoding.

**Figure 5. Picture showing the layout for gaze incorporated navigation. The wheelchair occupant's intention to get to the chair on the far end (left) is decoded. The wheelchair reaches to the chair, which is the destination defined by the user while avoiding static obstacles along the way (right).**

## IV. Conclusion

Such smart technology shows much promise to give people with severe disabilities the best quality of life possible, and the opportunity to maximize their human potential as it takes into account of the human factors, the importance of natural interfaces in assistive robotics [18] and has the ability to be mounted on any powered wheelchair [19]. Finally, the technology will be evaluated systematically in end-users.

## References


[1] L.A. Raymond, M. Piccini, M. Subramanian, P. Orlov, G. Zito, and A. A. Faisal, "Natural Gaze Data Driven Wheelchair," *bioRxiv*, p.252684, 2018.

[2] J. J. Belić, amd A. A. Faisal, "Decoding of human hand actions to handle missing limbs in neuroprosthetics," *Frontiers in computational neuroscience*, vol. 9, pp. 27, 2015.

[3] M. Xiloyannis, C. Gavriel, A. A. Thomik, A. A. Faisal, "Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol 25, pp. 1785-801, 2017.

[4] P. Orlov, A. Shafti, C. Auepanwiriyakul, N. Songur, A. A. Faisal, "A gaze-contingent intention decoding engine for human augmentation," *in Proc. ACM Symposium on Eye-Tracking Research & Applications*, pp. 91 2018.

[5] W. W. Abbott, A. Zucconi, A. A. Faisal, "Large-field study of ultra low-cost, non-invasive task level bmi," *Proc. 5th IEEE/EMBS Conf. Neural Engineering (NER)*, pp. 97-100, 2013.

[6] W. W. Abbott and A. A. Faisal, "Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain–machine interfaces," *Journal of neural engineering*, vol. 9, pp. 046016, 2012.

[7] M. Tall et al., "Gaze-controlled driving," in *Proc. ACM CHI Extended Abstracts on Human Factors in Computing Systems*, pp. 4387–4392, 2009.

[8] E. Wästlund, K. Sponseller, and O. Pettersson, "What you see is where you go: testing a gaze-driven power wheelchair for individuals with severe multiple disabilities," in *Proc. ACM Symposium on Eye-Tracking Research & Applications*, pp. 133– 136, 2010.

[9] G. Gautam, G. Sumanth, K. C. Karthikeyan, S. Sundar and D. Venkataraman. Eye movement based electronic wheel chair for physically challenged persons. *International Journal of Scientific & Technology Research*, *vol.* 3, no. 2, 2014.

[10] R. Xu, R. Hartshorn, R. Huard, J. Irwin, K. Johnson, G. Nelson, J. Campbell, S. A. Ay, and M. E. Taylor, "Towards a Semi- Autonomous Wheelchair for Users with ALS," in *Proc. IJCAI Workshop on Autonomous Mobile Service Robots*, New York City, USA, 2016.

[11] S. I. Ktena, W. Abbott and A. A. Faisal, "A virtual reality platform for safe evaluation and training of natural gaze-based wheelchair driving," in *Proc. 7th IEEE/EMBS Conf. Neural Engineering (NER),* pp. 236–239, 2015.

[12] P. M. Tostado, W. W. Abbott, A. A. Faisal, "3D gaze cursor: Continuous calibration and end-point grasp control of robotic actuators," *in Robotics and Automation (ICRA), IEEE International Conference,* pp. 3295-3300, 2016.

[13] R. O. Maimon-Mor, J. Fernandez-Quesada, G. A. Zito, C. Konnaris, S. Dziemian, A. A. Faisal, "Towards free 3D end-point control for robotic-assisted human reaching using binocular eye tracking," *in Rehabilitation Robotics (ICORR), IEEE International Conference,* pp. 1049-1054, 2017.

[14] B. Noronha, S. Dziemian, G. A. Zito, C. Konnaris, A. A. Faisal, ""Wink to grasp"—comparing eye, voice & EMG gesture control of grasp with soft-robotic gloves," *in Rehabilitation Robotics (ICORR), IEEE International Conference,* pp. 1043-1048, 2017.

[15] M. Li, N. Songur, P. Orlov, S. Leutenegger, A. A. Faisal, "Towards an Embodied Semantic Fovea: Semantic 3D scene reconstruction from ego-centric eye-tracker videos," *arXiv preprint*, pp.1807.10561, 2018.

[16] S. Dziemian, W. W. Abbott, and A. A. Faisal, "Gaze- based teleprosthetic enables intuitive continuous control of complex robot arm use: writing & drawing," *In Proc. of the 6th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob). IEEE*. pp. 5, 2016

[17] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM," *in Robotics and Automation (ICRA), IEEE International Conference,* pp. 1271-1278, 2016.

[18] T. R. Makin, F. de Vignemont and A. A. Faisal, "Neurocognitive barriers to the embodiment of technology," *Nat. Biomed. Eng,* vol. 1, pp. 0014, 2017.

[19] J. Leaman and H. M. La, "A comprehensive review of smart wheelchairs: past, present, and future," *IEEE Transactions on Human-Machine Systems*, vol. *47*, pp.486-499, 2017.