Using Visual Attention to Evaluate Collaborative Control Architectures for Human Robot Interaction

Tom Carlson[1] and **Yiannis Demiris**[2]

Abstract. Collaborative control architectures assist human users in performing tasks, without undermining their capabilities or curtailing the natural development of their skills. In this study, we evaluate our collaborative control architecture by investigating the visual attention patterns of robotic wheelchair users. Our initial hypothesis stated that the user would require less visual attention for driving, whilst they are being assisted by the collaborative system, thus allowing them to concentrate on higher level cognitive tasks, such as planning. However, our analysis of eye gaze patterns—as recorded by a head mounted eye tracking system—supports the opposite conclusion: that patterns of saccadic activation increase and become more chaotic under the assisted mode. Our findings highlight the necessity for techniques that assist the user in forming an appropriate mental model of the collaborative control architecture.

1 INTRODUCTION

Smart wheelchairs are being developed to augment a mobility impaired person's capabilities, enabling them to safely perform precise manoeuvres. In order to achieve this, we must share the control appropriately between the user and the robotic chair, such that the user still feels in control [10]. The human driver knows what they want to achieve and is good at interpreting complex, cluttered environments, however a robot can be much more precise in executing low level commands. Therefore, we have proposed an effective collaborative control methodology, which infers the user's intentions from their joystick input, along with the affordances of the surrounding environment [3, 1]. Based on these predictions, the wheelchair can alter the motor control signals to provide assistance, where necessary.

Our collaborative controller successfully increased the safety of trajectories driven in narrow spaces, whilst simultaneously reducing the need for excessive corrective joystick movements [2]. However, after processing feedback from these earlier trials, we are now investigating the possibility of using additional physiological input from the user, to help the wheelchair behave as naturally as possible during interaction. We propose to utilise the user's eye gaze, to estimate their attention, which could simultaneously enhance our prediction of intention algorithm and indicate when the wheelchair does not behave as the user expects.

Whilst the user is being assisted by the collaborative system, we hypothesize that the user would require less visual attention to effectively manoeuvre the wheelchair. This would allow them to simultaneously perform other higher level cognitive tasks, such as environmental exploration, or planning future manoeuvres. We also expect the driver to fixate on objects of interest, which may help to strengthen our intent-prediction system.

We do not treat eye gaze as an active input device, in which the user tries to control the wheelchair by moving their head and/or eyes, as was demonstrated in [9]. Instead, we aim to use it as a passive device, to non-intrusively increase the user state vector (the knowledge we possess about the user at each time step).

In this exploratory study, we observe the characteristics of the user's eye movements, whilst performing typical manoeuvres, such as driving around offices and passing through narrow doorways. The observations are made over one independent variable, which can take one of two states: provide adaptive assistance, or provide no assistance.

Figure 1. A participant driving our robotic wheelchair. The software on the tablet PC combines the stimulus from the joystick with the localisation data derived from the camera, to collaborate with the user in controlling the wheelchair motion. Simultaneously, the head-mounted eye-tracking system records the driver's gaze.

2 BACKGROUND

In this section, we will introduce the robotic wheelchair platform that we have developed (Figure 1). After briefly describing the hardware

[1] Imperial College London, UK, email: tom.carlson02@imperial.ac.uk, web: http://www.iis.ee.ic.ac.uk/~tcarlson
[2] Imperial College London, UK, email: y.demiris@imperial.ac.uk, web: http://www.iis.ee.ic.ac.uk/yiannis

components, we focus more on the underlying collaborative control architecture. We then go on to explain the eye tracking system, describing the hardware setup along with the fundamental operation of the tracking algorithm.

2.1 The Robotic Wheelchair Platform

Our system is built upon a mid-wheel drive EPIOC (electrically powered indoor/outdoor chair), that would typically be prescribed to a severely mobility impaired patient. We have interfaced a tablet PC with the wheelchair's joystick and motor control unit. The joystick signals are intercepted and altered, where necessary, before being forwarded to the wheelchair's motor control unit.

The collaborative controller (Figure 2) is realised in software on the tablet PC. The adaptive assistance is provided by the shared controller module, which uses information from the *safe mini-trajectory* generator, along with the intention predictor, to decide exactly how to adapt the joystick signals [3]. The system is currently underpinned by the computer vision-based localisation system that we have developed to work in mapped, indoor environments (with minimal modification of the environment) [3].

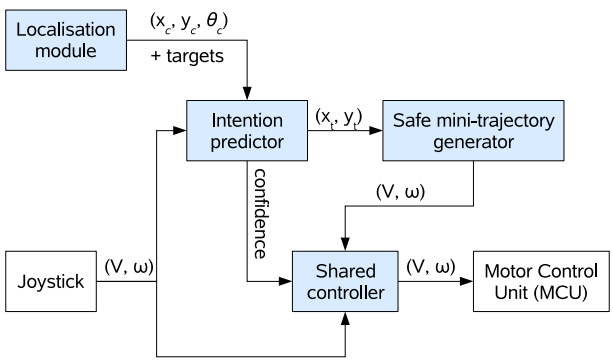


Figure 2. The collaborative system shares the control appropriately between the user and on-board computer [3]. (x_c, y_c, θ_c) and (x_t, y_t) describe the wheelchair's current and target positions respectively. (V, ω) represent the target translational and rotational velocity tuple to be sent to the motor control unit.

2.1.1 *Intention prediction*

In our system, we base our intention prediction on the multiple hypothesis approach, as described in [4]. Our prediction models are task based, so we define targets of interest, such as doorways and desks, which the user may wish to drive through or approach. We constructed a confidence function (Equation 1), which only increases when moving towards a target. This function is the product of two parts: the first (Equation 2) is computed using the Euclidean distance from the current wheelchair position $\mathbf{w}$ to the target $\mathbf{w_t}$, the second (Equation 4) is based upon the heading of the chair θ, compared with the angle to the target ϕ (Equation 3). The scaling factor of k in Equation 4 determines the sensitivity towards the angular error and in our case was experimentally set to 2.

$$C = C_d C_\theta \tag{1}$$

$$C_d = \exp\left\{-||\mathbf{w_t} - \mathbf{w}||\right\} \tag{2}$$

$$\phi = \arg\left(\mathbf{w_t} - \mathbf{w}\right) \tag{3}$$

$$C_\theta = \exp\left\{\frac{k(\pi - |\theta - \phi|)}{\pi} - k\right\} \tag{4}$$

Essentially the *safe mini-trajectory* generator computes a path to reach the predicted target safely, once the confidence threshold has been reached for that target. The wheelchair is then guided gently towards the first waypoint of the *safe* path. However, we allow the user to gradually deviate from this path, if they consistently oppose this attraction. The confidence value will then fall accordingly; eventually allowing them to regain full control if appropriate. Conversely, we will prevent them from deviating from the safe path if it is going to result in a collision (e.g. they are in a doorway and might hit the doorframe). However, the speed of the manoeuvre is always controlled by the user, in a manner similar to that of Zeng et al. [12]; it is proportional to the component of the joystick input vector, which lies in the *safe* direction. This continues until the corresponding confidence value has dropped below an experimentally set threshold, C_{thresh}, which occurs once the chair has safely reached its target destination. We also allow the user to reverse backwards along the safe path at any time, until the confidence value drops below C_{thresh}, at which point they regain full control. This strategy strives to make the user feel more in control than using a rigid assistance method, which forces them to stay on a computer-controlled path at all times, for example, when the *CALL smart wheelchair* uses its "optical track follower" [10].

2.2 The Eye Tracker

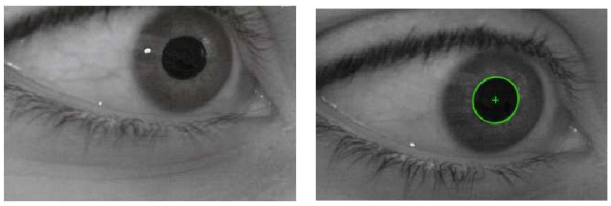
In the BioART lab, we have constructed a portable monocular eye-tracker, which is based on the openEyes system developed at Iowa State University [7]. It allows us to indicate the user's point of gaze (POG) on a projection of their field of view (the scene image) [11]. We decided to use a cycle helmet as our substrate, to comfortably and securely support the hardware, whilst allowing quick adjustments to be made for new users (Figure 5).

The scene image is produced by a firewire camera with a fish-eye lens (111° field of view), mounted on the headpiece above the tracked eye, which reduces parallax error. Concurrently, the subject's right eye is illuminated by an infra-red LED, which is observed by a second firewire camera. To reduce the sensitivity to lighting conditions, this second camera is fitted with a Kodak Wratten 87C infrared filter, which blocks the visible spectrum.

2.2.1 *The PCCR algorithm*

The eye-tracking algorithm uses the pupil centre and corneal reflection method (PCCR) [5]. We based our implementation on the openEyes project [7], which utilises the Starburst algorithm [8]. Our implementation is in C++ rather than C, which has allowed us to make a portable class-based solution, abstracting away the interface dependence on the HighGUI/GTK widgets. This means we can easily use the real-time eye tracking class in our existing multithreaded QT framework, or as a standalone application.

Since we are using filtered, reflected infrared light, the corneal reflection should be the brightest part of the image, as shown in Figure 3(a). An adaptive binary thresholding method is used to detect this. Once the reflection has been found, its position is recorded and it is subsequently removed from the image, so as not to interfere with the pupil detection. This is done by replacing it with the average pixel intensity of its neighbours.



(a) Corneal reflection. (b) Finding the pupil.

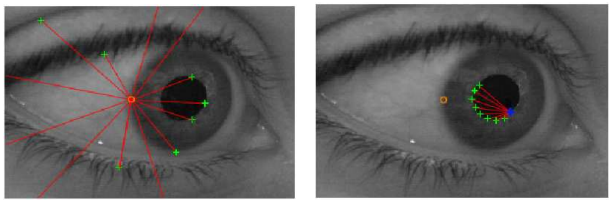
Figure 3. The gaze direction is computed from the vector between the centre of the pupil and the corneal reflection. To do this, we use a modified version of Starburst algorithm [8].

We have extended this method to prevent erroneous detections that sometimes occur in bright lighting conditions, or if the user is wearing glasses. Once the pupil has been detected, we check that the corneal reflection lies within 150 pixels of the pupil centre. If this is not the case, we discard the reflection result and restart the search, looking for the next largest area of high intensity pixels.

2.2.2 *Determining the pupil centre*

Robustly finding the pupil is more of a challenge, we apply the feature-point detection method which is also part of the Starburst algorithm [8]. We start by setting the estimate of the centre of the pupil $\mathbf{p_{current}}$ to be in the centre of the image, at each iteration, this estimate is updated to be the mean of the generated feature set. In successive frames, $\mathbf{p_{current}}$ is taken to be the final value from the previous image.

The first part of the algorithm generates a number of rays radiating from $\mathbf{p_{current}}$, to the edges of the image. Each ray is then traversed, calculating the derivative intensity between neighbouring pixels, Δ. For the first point $\mathbf{f_i}$ on each ray that this intensity is greater than a certain threshold, δ, it is deemed to be a potential boundary between the pupil and iris. Consequently it is added to the candidate feature set (Figure 4(a)). Once all the rays have been traversed, a similar process is begun for each of the points in the generated feature set, only this time, the rays are constrained to fall within an arbitrary angle of $\mathbf{p_{current}}$ (Figure 4(b)). The whole algorithm is repeated with $\mathbf{p_{current}}$ set to the geometric centre of the feature set, unless they are within 10 pixels of each other, in which case it terminates [8].



(a) Phase 1. (b) Phase 2 (1st iteration).)

Figure 4. Detecting the pupil using the Starburst algorithm [8]. In phase 1, potential features (green crosses) are found along n rays radiating from $\mathbf{p_{current}}$ (red blob in centre). In phase 2, new feature points are located along n rays from each of the features $\mathbf{f_i}$ in our current set, back towards $\mathbf{p_{current}}$.

The final stage is to fit an ellipse to the candidate feature set, which is performed using the random sample consensus (RANSAC) paradigm in a similar manner to [6]. This allows the true centre of the pupil to be estimated.

2.2.3 *Calibration*

Now that we have a system for estimating the *eye difference vector* from the centre of the pupil to the corneal reflection, we can determine the POG on the scene image. This is achieved by using the homographic mapping function, H, which is a 3×3 matrix with eight degrees of freedom. H is computed by performing a standard calibration, whereby the user sequentially fixates on 3×3 known grid points in the scene image [8].

Figure 5. Our portable eye-tracker consisits of two firewire cameras; one of which observes the subject's right eye, whilst the other records the scene (the person's field of view). After performing a manual calibration, the software will mark the person's focal point on the scene image.

It is important to note that this eye tracking system does not give pixel-level resolution on the scene image, as might be required for the manipulation of a cursor on a screen. Instead, it is sufficiently accurate to detect when the subject's POG dwells on regions of interest, such as doorways, desks, people etc., however it remains extremely sensitive to any calibration error.

2.2.4 *Improved calibration for inexperienced users*

In our experiments, many of the subjects have never experienced the use of an eye tracking system before, so we must ensure the process is as simple and non-intrusive as possible. Although the calibration procedure is relatively quick and simple, it is crucial to get it right first time. It has proven to be disruptive when running trials with several participants that each have to make successive re-calibrations. We observed that many of these errors occurred due to the subject blinking, or moving the eye just as the calibration point was set.

Therefore, we improved the calibration stage in order to automatically detect and eliminate errors in producing the homographic mapping matrix. We now compute the stability of the *eye difference vector* over a few frames before and after the calibration point has been set. If the vector is deemed unstable (± 5 pixels on each end point), the point is rejected and the subject must perform that fixation again, before moving onto the next grid point. This intuitive adjustment to

the algorithm greatly reduced the need for re-calibration, which has saved experimental time, reduced erroneous results and helped to instill greater confidence in our test subjects.

3 METHODOLOGY

This paper presents the approach we followed for conducting a pilot study. It is important to test the methodology before carrying out the full scale trials, which will take a considerable length of time. We followed some generally accepted principles for our experimental design, to ensure our results are as meaningful as possible and eliminate any bias. A script was used to ensure that all participants were told the same information. Additionally, none of them were aware of the goals of the experiment.

In contrast with previous work [2, 3], the participants were not informed prior to the trial whether the system would be assisting them to perform the manoeuvre or not. This allowed us to gain some interesting results relating to people's mental models of the underlying control system, which will be further explored in the Discussion section.

Each participant was asked to drive from the start point in one office, through a narrow doorway to an adjoining office, exit the office via another doorway and stop in the corridor, as shown in Figure 6. This route was driven four times by each participant. Half of the trials were driven with no assistance, whereas the other half were driven with the collaborative controller active, thereby providing assistance at certain points, notably when driving through the doorways. In order to prevent biases, the collaborative controller was active during the odd trials for the odd–numbered participants and during the even trials for the even–numbered participants.

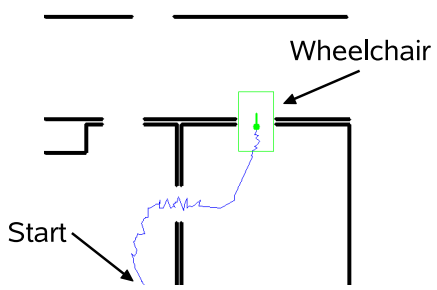


Figure 6. In the trials, participants were required to drive from the start position in one office through to an adjoining office, before arriving at the finish position in the corridor. When using the collaborative controller, it provides assistance, where necessary, in the approach to and whilst driving through the doorways.

In our previous experiments [3, 2], we received many useful comments from the participants, so this time we decided to collect information that we could quantify in a structured questionnaire. After driving each trajectory, the participant was asked to rate the following carefully chosen statements on a five point Likert scale, ranging from "strongly agree" to "strongly disagree":

- The wheelchair was easy to manoeuvre.
- The wheelchair behaved as I expected.
- I had to concentrate hard to drive the wheelchair.
- It felt natural driving the wheelchair.

4 RESULTS

We took two approaches to processing the large quantity of video data that we had collected. The first was to qualitatively gain information from watching the playback of the scene images, which had been automatically annotated with the gaze points. The second was to perform a quantitative analysis, by computing statistics on clusters of the gaze points, for instance: their standard deviation, whilst passing through a doorway.

Using our first approach, we noticed that all the participants tended to perform a visual exploration of the scene as they drove through through a doorway, an example of which is shown in Figure 7. This comprised of rapid saccadic eye movements. However, once they were in a room they tended to fixate on objects or perform smooth pursuit, as shown in Figure 8.

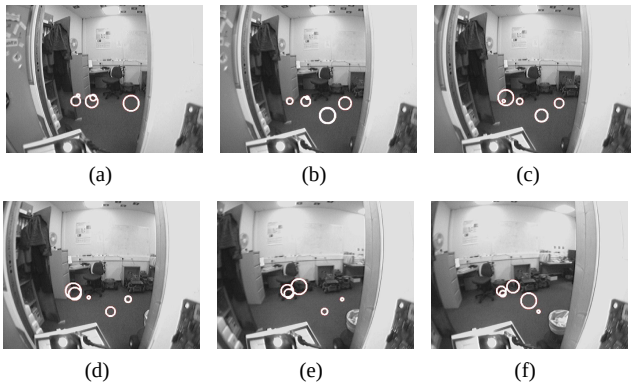


Figure 7. A sequence showing the visual exploration as a user drives through a doorway. The white circles indicate the gaze of the user; the largest one is the current POG, the smaller ones indicate the previous four POGs.

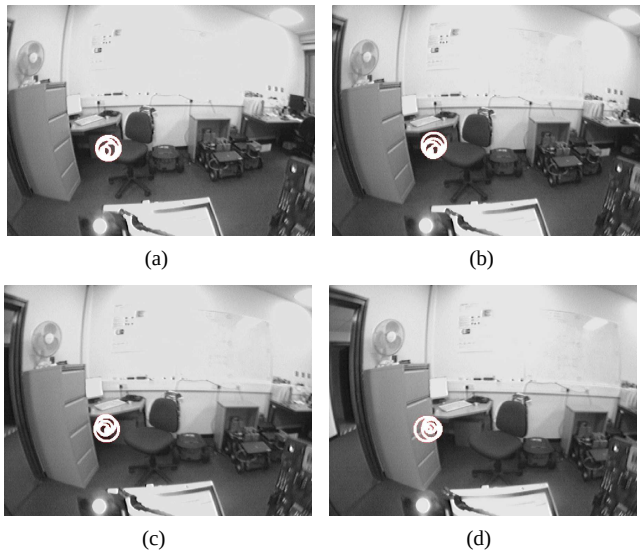


Figure 8. Once a user is in an office, which they have already visually explored, their gaze tends towards the forward position. This is clearly shown by the sequence of processed POGs, which form approximate concentric circles.

Typically, users driving without any assistance exhibited patterns of movement in the points of gaze (POGs) as shown in Figure 9. In this case, the driver was looking predominantly straight ahead, in the direction of the chair's movement. However, when the collaborative control system assisted the user (unbeknown to them) to drive through a narrow doorway, the typical pattern of POGs changed dramatically to resemble that of Figure 10.

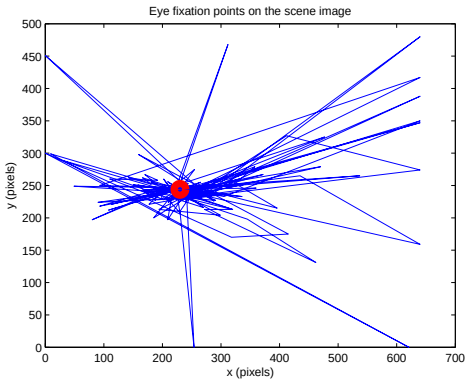


Figure 9. A superposition of the the typical POGs, when driving through a doorway without assistance. The large red blob indicates the median, which corresponds to the user looking straight ahead.

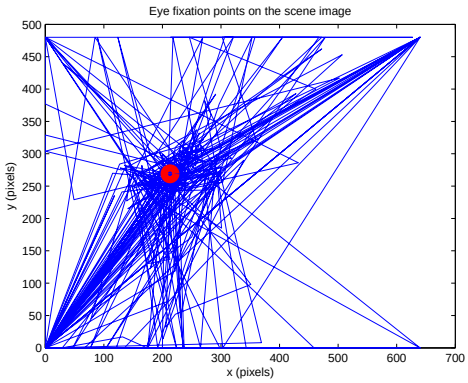
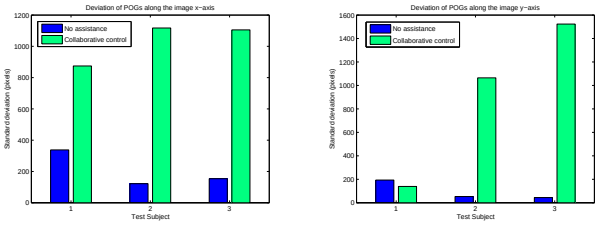


Figure 10. A superposition of the the typical POGs, when driving through a doorway using the collaborative control method. The large red blob indicates the median, which corresponds to the user looking straight ahead.

5 DISCUSSION

When people were driving without any assistance, their POGs predominantly clustered around the median position, with a relatively low deviation, as you can see in Figure 9. The median position corresponds to the times at which the user is looking straight ahead. In Figure 11, the graphs highlight the significant increase in the deviation of the POGs, once the user is being assisted by the collaborative controller. All the subjects exhibited increased deviation in the horizontal plane of the scene image (Figure 11(a)), however the results for the vertical plane of the image (Figure 11(b)) may yield an interesting explantion for the increased saccadic eye movements.



(a) Scene x-axis. (b) Scene y-axis.

Figure 11. The standard deviation of the points of gaze (POGs) for subjects whilst driving through narrow doorways. The case when no assistance was given is compared with the case when the collaborative controller was active.

When using the collaborative controller—compared with not being given any assistance—all of the participants agreed more strongly with the statement: "I had to concentrate hard to drive the wheelchair" and tended to disagree more with "the wheelchair behaved as I expected". However their reasons for this differed. Predominantly people tended to comment about the wheelchair "not behaving correctly", whereas the first participant described feeling that it was their own fault for not understanding how to operate the wheelchair properly. This could explain why the first participant was the only person not to also significantly increase deviation in the vertical plane of the image (Figure 11(b)). They tended to look only at the doorframes (Figure 12), whereas everyone else additionally focussed on the tablet PC.

This suggests a potential lack of a mental model for the wheelchair. We define a person's mental model to be their perceived forward model of a system's behaviour. For example, if we apply a control signal to the current system state, what would be the next state of the system [4]? In our case, how does the user expect the wheelchair to move as a result of a joystick manipulation?

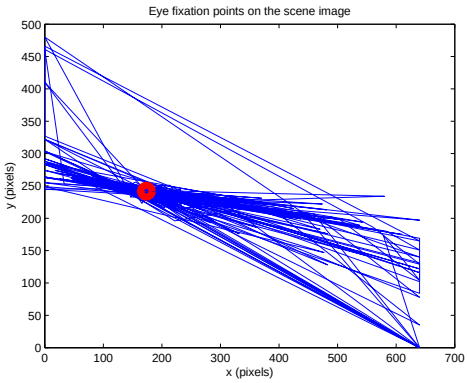
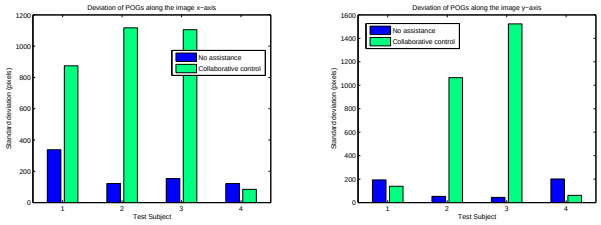


Figure 12. An increased deviation in the horizontal plane, without a significant increase in the vertical plane, suggests the user is more concerned with performing the manoeuvre, than what the control system on the tablet PC is doing.

Inspired by our previous results, we decided to carry out an additional trial, to test our hypothesis regarding the requirement of an adequately formed mental model. The results of which are plotted against our previous findings in Figure 13. We informed an additional fourth test subject about the shared control policy within the collab-

orative control architecture and how it would assist with manoeuvres whilst it was active. This meant they were able to form an appropriate mental model of the wheelchair's expected dynamic behaviour. Consequently they produced no significant difference in either axis of their eye gaze patterns between the case when they were given assistance and the case when they were in full control.

This would suggest that a high degree of saccadic eye movement could indeed be triggered by the lack of an appropriate mental model. However it would require further trials, which we are currently undertaking in the BioART lab, to produce statistically significant results, to support this premise.



(a) Scene x-axis. (b) Scene y-axis.

Figure 13. The additional fourth test subject was aware of the shared control policy within the collaborative control architecture. This meant they had an appropriate mental model of the wheelchair's behaviour, consequently resulting in no significant difference in the eye gaze patterns.

6 CONCLUSIONS

We have constructed a head-mounted, portable eye-tracking system and interfaced it with our existing robotic wheelchair. Through the use of our collaborative control architecture, people have improved the quality of their driving, in terms of the safety of the trajectories followed. Conversely, an analysis of the user's eye movement, combined with a questionnaire and verbal feedback, has indicated potential difficulties that users have in the recognition and understanding of adaptive interfaces.

In our work, we have again demonstrated the importance of using physiological measures in addition to the more traditional system performance metrics (e.g. speed, distance etc.) when evaluating intelligent HRI architectures. The results can be counter-intuitive; for example, it may require more concentration to perform a task when being given adaptive assistance. However, such results are important to discuss and in this case, they follow the premise that it could take longer to create a mental model of an adaptive interface than of one that has a fixed response.

Our initial hypothesis stated that the user would require less visual attention for driving, whilst they are being assisted by the collaborative system. This would allow them to concentrate on higher level cognitive tasks, such as planning or performing a visual search. However, our analysis of eye gaze patterns for untrained users supports the opposite conclusion; that patterns of saccadic activation increase and become more chaotic under the assisted mode. Therefore, our findings reiterate the necessity for techniques that assist the user in forming an appropriate mental model of the collaborative control architecture.

7 FUTURE WORK

Inspired by our unexpected findings and the success of our pilot study, we intend to perform a full-scale study, investigating the user's mental model of the collaborative control system. It would be interesting to do a *between* subjects trial, where some users only operate the chair with the collaborative controller active, whilst others are not given any assistance. Over the course of a number of trials, we would expect both the groups to form different perceptual models of the wheelchair's behaviour. Despite these differences, the degree of saccadic eye movements may converge.

ACKNOWLEDGEMENTS

The authors would like to thank Kaveh Yousefi, Philipp Leskovitz and Valerio Chang for their contributions to the development of the eye-tracker. We are also extremely grateful to all those who participated in our experiments. Finally, we would like to thank the members of the BioART team — Simon Butler, Murilo Fernandes Martins, Bálint Takács, Paschalis Veskos and Yan Wu — for their insightful comments and continued support.

REFERENCES

[1] Tom Carlson and Yiannis Demiris, 'Evidence of shared control improving accuracy in human-wheelchair interaction', in *Proc. The MAIA Project Workshop 2007*, K.U. Leuven, Belgium, (Nov 2007).

[2] Tom Carlson and Yiannis Demiris, 'Collaborative control in human wheelchair interaction reduces the need for dexterity in precise manoeuvres', in *Robotic Helpers: User Interaction, Interfaces and Companions in Assistive and Therapy Robotics, a Workshop at ACM/IEEE HRI 2008*, pp. 59–66, Amsterdam, The Netherlands, (March 2008).

[3] Tom Carlson and Yiannis Demiris, 'Human-wheelchair collaboration through prediction of intention and adaptive assistance', in *Proc. of IEEE International Conference on Robotics and Automation*, pp. 3926–3931, Pasadena, CA, (2008).

[4] Yiannis Demiris, 'Prediction of intent in robotics and multi-agent systems', *Cognitive Processing*, **8**(3), (September 2007).

[5] E. D. Guestrin and M. Eizenman, 'General theory of remote gaze estimation using the pupil center and corneal reflections', *IEEE Transactions on Biomedical Engineering*, **53**(6), 1124–1133, (June 2006).

[6] Dan Witzner Hansen and John Paulin Hansen, 'Robustifying eye interaction', in *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*,, pp. 152–159, (2006).

[7] D. Li, J Babcock, and D. J. Parkhurst, 'openeyes: A low-cost head-mounted eye-tracking solution', in *ACM Eye Tracking Research and Applications Symposium*, (2006).

[8] D. Li, D. Winfield, and D. J. Parkhurst, 'Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches', in *IEEE Vision for Human-Computer Interaction Workshop at CVPR*, pp. 1–8, (2005).

[9] Y. Matsumoto, T. Ino, and T. Ogasawara, 'Development of intelligent wheelchair system with face and gaze based interface', in *Proc. of 10th IEEE Int. Workshop on Robot and Human Communication (ROMAN 2001)*, pp. 262–267, (2001).

[10] P. Nisbet, 'Who's intelligent? Wheelchair, driver or both?', in *Proc. IEEE Intl. Conference on Control Applications*, Glasgow, Scotland, U.K., (September 2002).

[11] Kaveh Yousefi, *Multimodal human-computer interaction : eye tracking*, Master's thesis, Department of Computing, Imperial College London, 2007.

[12] Qiang Zeng, Etienne Burdet, Brice Rebsamen, and Chee Leong Teo, 'Evaluation of the collaborative wheelchair assistant system', in *IEEE Conference on Rehabilitation Robotics*, The Netherlands, (June 2007).