

# A Tool to Assist in the Analysis of Gaze Patterns in Upper Limb Prosthetic Use

**Peter J. Kyberd** School of Engineering, University of Derby, Kedleston Road, Derby, DE22 1GB, UK.

**Alexandru Popa** Institute of Biomedical Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada.

**Théo Cojean** Université de Lyon, Université Gustave Eiffel, Université Claude Bernard Lyon 1, Lyon, France.

**Address for correspondence:** Peter Kyberd, UK. email: p.kyberd@derby.ac.uk

## 1 Abstract

Gaze tracking, where the point of regard of a subject is mapped onto the image of the scene the subject sees, can be employed to study the visual attention of the users of prosthetic hands. It can show whether the user pays greater attention to the actions of their prosthetic hand as they use it to perform manipulation tasks, compared with the general population. Conventional analysis of the video data requires a human operator to identify the key areas of interest in every frame of the video data. Computer vision techniques can assist with this process, but a fully automatic systems requires large training sets. Prosthetic investigations tend to be limited in numbers. However, if the assessment task is well controlled, it is possible to make a much simpler system that uses initial input from an operator to identify the areas of interest and then the computer tracks the objects throughout the task. The tool described here, employs colour separation and edge detection on images of the visual field to identify the objects to be tracked. To simplify the computer's task further, this test uses the Southampton Hand Assessment Procedure (SHAP), to define the activity spatially and temporarily, reducing the search space for the computer. The work reported here is the development a software tool capable of identifying and tracking the Points of Regard and Areas of Interest, throughout an activity with minimum human operator input. Gaze was successfully tracked for fourteen unimpaired subjects, which was compared with the gaze of four users of myoelectric hands. The SHAP cutting task is described and the differences in attention observed with a greater number of shorter fixations by the prosthesis users compared to unimpaired subjects. There was less looking ahead to the next phase of the task by the prosthesis users.

## 2 Introduction

The operation of an upper limb prosthesis requires the users to control their device using limited feedback paths [1]. Skill in control is attained as the result of sustained practice and concentration on the part of the user. One of the primary feedback paths used is vision. The hands remain in the view of the operator most of the time and the high quality information supplied about the hand's position and grasp is crucial to successful operation. In contrast, natural hand control is achieved using many different modalities [2] and the hand is generally in only in the person's peripheral vision [3].

Gaze tracking is an established technology, in its portable form it uses cameras mounted on the head to record eye movement and the visual field. A computer maps the focus of the gaze onto the scene in front of the subject. The output of the system is a video of visual field

with the foveal point superimposed upon it. This provides information about the direction of the gaze. The visual attention can be inferred from this, by observing where the subject looks and for how long. The technology has allowed exploration of understanding of the role of gaze in reaching, manipulation [3,4], sports [5], and skills acquisition [4]. Other eye tracking systems can be used to control computer cursors, but as they rest on the desk they have less use in this application where the subject needs to move about [6]. More recently, this exploration has moved into the investigation of prosthetic grasping [7,8]. For natural grasping the gaze tends to anticipate the hand, looking to where the hand will be, and moving on before the hand has grasped the object, to where the hand will go next. In prosthetic manipulation it has been shown that the users tend to follow the hand closely not looking ahead to the next target until the current one is acquired [7,8].

Employing a gaze tracker is the first step towards understanding how prostheses users look at a task involving their prosthesis. Potentially will allow improvement in the design of the control of prosthetic hands while reducing the cognitive burden. Analysis of the recorded scene requires human intervention. Generally, it is a human operator who goes through the video data frame by frame. They must identify and mark the key items and events in the visual field. This is clearly time consuming and dependant on the skills of the operator. Computer vision techniques are available to interpret the scene, but machine learning needs to be able to segment the data into significant actions and aggregate real, noisy data into statistically similar groups. Thus it needs copious amounts of data to learn from [9]. In contrast, similar to the development of low cost motion tracking systems, (such as the Kinect for home gaming), it is possible to create a simplified system that will speed up the image analysis without the expense of highly sophisticated computers and software [10,11].

This paper describes a system designed speed up the process of processing eye tracking data for the analysis of prosthetic hand function. The program uses a standardised set up, including an activity of daily living tasks. A computer is able to identify the key areas of interest and track them in the visual field throughout the activity, with minimal human intervention. It is believed this is the first time this assisted analysis has been created for this application.

## **2.1 Background**

The study of gaze behaviour and visual attention for upper-limb activities of daily living (ADLs) has been conducted since the technology became compact enough to make field studies practical [12]. More recently, it has garnered the attention of the prosthetics field, with an initiative from three groups [7,8,13]. This was later expanded to the wider community [14,15,16]. All of these studies have required hand labelling of the video data by a skilled operator. The overall goal of this program of study was to assess visual attention of a prosthesis user when faced with a new task, and to gauge it's influence of the usability of upper limb prostheses by observing the eye gaze patterns of users of prosthetic hands. The aim of the project detailed herein was a first step; to develop a method for assessing visual attention with the minimal human intervention.

In the commercial systems, the video data from cameras mounted on the head is combined by a computer to map the Point of Regard (PoR) onto the image field. One camera records the motion of the eye and uses the position of the pupil to derive the PoR. This is mapped on to the image taken by a second head mounted camera of the scene. Before use the system is calibrated.

Generally, to analyse the videos generated, the experimenter has to step through the video frame by frame, identifying items of interest and inputting them into a computer for numerical analysis. This is laborious and time consuming, it relies on human interpretation of the Areas of Interest (Aols) in the scene and is dependent on the level of skill and attention of the observers [17]. Computer systems and image processing software are now able to analyse complex video images through forms of machine learning [18,19]. However, to achieve an adequate level of proficiency to be useful [9], the systems require many examples of the test data to train the system. Studies with prosthesis users rarely generate enough data to train an advanced machine learning model sufficiently. The number of repetitions each subject would have to perform is unrealistic. This project created a system that did not require the computer system to be that competent or need many examples to train the software. Instead, the software was designed to extract key items of interest from the video of the visual field, with the operator marking the first instance of a Aol and then the computer being able to track it in subsequent frames. The automatic identification was based on three details: The colour of the object, its shape and that an object cannot have moved far between successive frames. This approach has not been used previously for this sort of application.

The project was not focused on generating new ideas and technology in the computer vision field, but to use existing tools to extract information. Thus an interface was created that used tools within Matlab (Mathworks, Natick, 1 Apple Hill Drive Natick, Massachusetts, USA) to extract sufficient information to assist in the identification of a range of points of interest in the visual field. Data processing and analysis used a combination of pre-existing Matlab functions and custom written routines.

## 2.2 Assessment framework

Assessment of complex activities is a compromise between extreme abstraction, (which is easy to measure but may not determine realistic results) and verisimilitude (which is highly variable and is hard to draw general conclusions about a population) [20]. For this application, an assessment form was chosen that controlled the task and the environment to assist the computer in tracking the task while ensuring it had clinical validity.

While some experiments have employed simulated prosthesis users as subjects [7,14,21] and abstract tasks, for this study an assessment tool was chosen to allow for the measurements to be made while the routine users of prosthetic hands performed simulated Activities of Daily Living (ADLs). The Southampton Hand Assessment Procedure (SHAP) was chosen because it has qualities that make it particularly appropriate for this application [22]. SHAP is a tool validated for assessment of users of prosthetic hands. It measures the functional capabilities of the user and hand *in combination* [23]. All functional tests are a compromise between the practical (what can be achieved in the time, space and budget) and the accurate use of real world operations. SHAP uses a series of standardised tests based on tasks that were already proven to be repeatable and validated [24]. These tasks are timed by the operator as this too increases the reliability of the measurement [24]. A score is produced which is based on the Mahalanobis distance of the times of the subject from a standardised set measurements. The overall score is out of 100, with scores above 95 being the range for the general population. There are eight abstract tasks, (picking and placing abstract shapes) and fourteen simulated ADLs. The layout of the tasks are controlled by a form board. SHAP has been used in a wide range of conditions including prosthetic limbs [25,26,14] and grasp kinematics [27,28], (in conjunction with motion analysis systems).

A SHAP task is set up on the form board. The start and end positions for all elements are marked on the board, the board also fixes the timer in the centre. Participants are instructed to perform the tasks with the hand under test (the non-dominant or the prosthesis in these experiments). The subject presses the button on the timer, performs the test and then presses the button again to turn off the timer and complete the task, Figure Set\_up. This time is entered in a database which calculates the overall score and the score for each grip form [22].

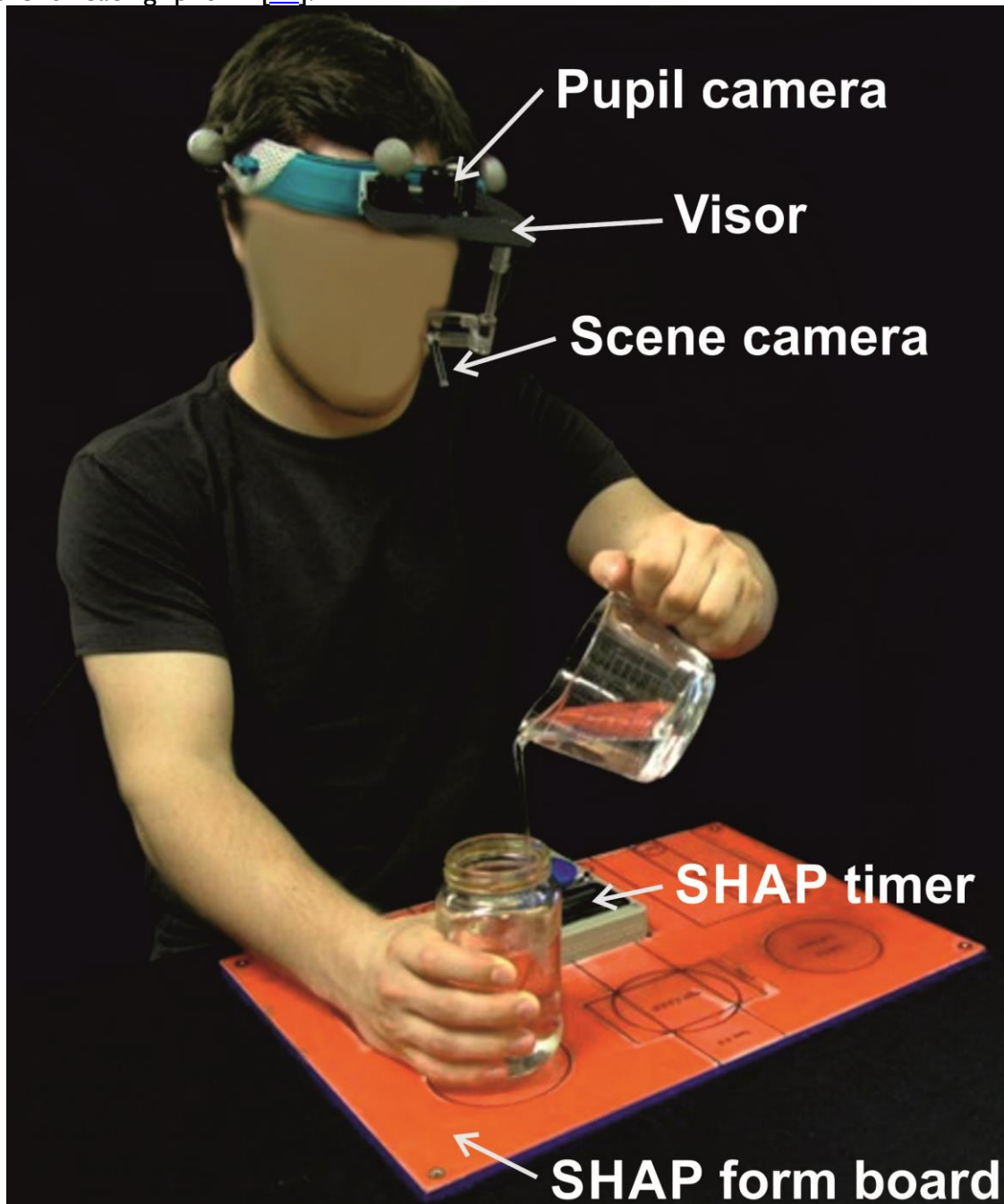


Figure 1: Experimental set up: Shown is the IScan visor that holds the pupil and scene cameras. The pupil camera reads the position of the pupil via a mirror, this data is used to reconstruct the PoR in the scene recorded by the second camera. The SHAP task shown here is the pouring task with the form board and the timer in front of the subject.

For this application SHAP has the additional advantage of providing a framework that allows the computer system to make assumptions about the underlying structure of the tests: The form board holds the parts of the task in a fixed starting point. Each surface of the form board is a strong standardised colour (red for ADLs and blue for abstract objects). The timer button is in the centre and is blue surrounded by the grey of it's case. This makes the prediction of where points of interest lie in the first frame and their extraction from a noisy video field easier. Thus any computer based search can start with these features assumed e.g. colour, contrast and relative position in the visual field and search for known shapes and colours in the most likely place.

## 2.2.1 Task

For this paper, the SHAP task chosen for consideration is the Cutting task. The task is laid out with a cylinder of plasticine placed in the middle of the form board, beyond the timer. The knife is placed on the side of the hand under test. A test begins with the user starting the timer with the hand under test. The subject is allowed to pick it up with their contralateral hand and place it in the prosthesis (if they are prosthesis users). They then grasp the plasticine with their contralateral and cut the strip before replacing the knife beside the timer and turning the timer off with the test hand.

The gaze information was analysed in relation to a series of areas of interest (Aols). Every activity has different Aols depending on the form of the ADL that is captured in the scene. For example; for the cutting task the Points of Regard (PoRs) are the knife, the plasticine, and the timer button. Investigation of the visual attention was then achieved by evaluating the amount of time the PoR was fixated on the specific Aols throughout a given activity.

## 3 Method

Subjects were asked to perform simulated ADLs, using SHAP. During the tasks, the visual attention was recorded using an head mounted eye tracker, ISCAN, (ISCAN, Inc., 21 Cabot Road, Woburn, MA 01801, USA). The IScan system collects video data of the scene together with the coordinates of the pupil of one eye. The IScan identifies and suppresses saccades (quick movements of the eye to a new PoR) and maps the PoRs onto the video of the image field. This is stored in a standard video format. It is this data of the visual field and the point of regard that was used as the input to the analysis software described here.

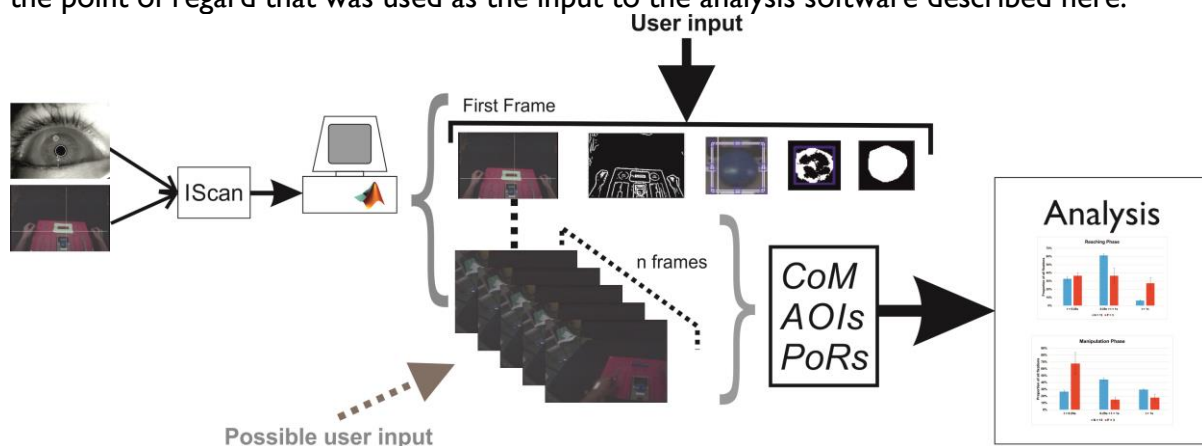


Figure 2: Schematic of the process. The IScan system uses two cameras to provide point of view data. This video is then analysed for information on the point of view, and areas of interest in each frame, having been trained on the areas of interest in the first frame the system steps through frame by frame. The data produced is then analysed for insight into gaze behaviour.

A schematic of the entire system is Figure 2. The guiding principle of the software design was that it should be simple to use and not need extensive development of specialist machine learning software. The initial identification of different areas would be made by the operator on the first image of the sequence and the computer would be able to extract sufficient information about the image to track through the rest of the video. Should the computer lose track of an area of interest, the operator could help the program re-acquire the area, through simple interactions with the computer.

The custom software presented the first frame to the operator, who used a mouse to identify the key areas of interest in the particular task. The software then identified the colour and shape of the object and subsequently stepped through the video frame by frame using this data on the colour and shape to identify the Aols in each new frame. As the Aols cannot move very far between successive frames the computer need only search in small area close to the position the object in the succeeding frame. This reduces both the search time and the chances of another item being misinterpreted as one of the Pols. If the computer had not found the object within the search area it then requested the operator to point to its location. Once the Aol was found then its Centre of Mass (CoM), was recorded and the computer went on identifying all other Aols in this frame. Once complete the computer moved on to the next frame, and continued frame by frame until the end of the video was reached.

The IScan was worn on the head and cables run to a computer. The video stream information was captured at a frame rate of 29.97 frames per second. The area surrounding the table was curtained off to remove unwanted additional images in the peripheral vision. The view generated by the system is shown in Figure 3. This image is of the form board and the timer. The cross-hairs for the point of regard are visible. This will be referred to as the 'primary image' from which the analysis software bases its initial decisions and the results will be shown throughout the rest of the paper.

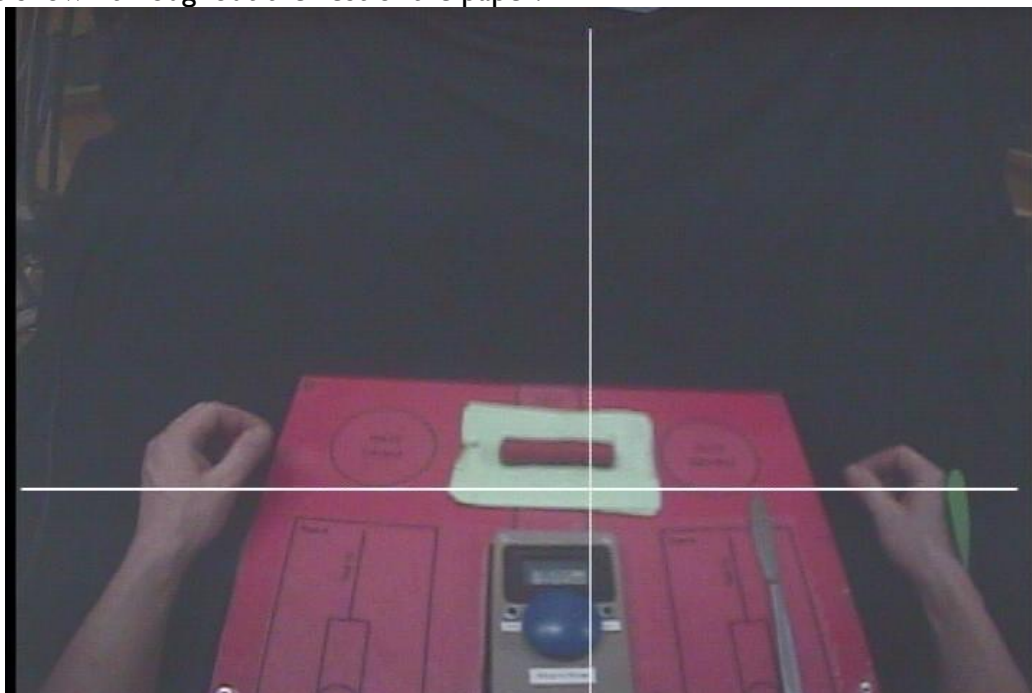


Figure 3: An example of the image created by the IScan system. The form board, subject's hands, and the cross shows, where the subjects gaze is focused. The form board is set up for the cutting

task, the object to be cut is in the top centre, the knife is to the right and centre bottom is the timer with the on/off switch in blue.

### 3.1 Image Processing

The video is a sequence of images or frames, the analysis was made frame by frame starting with the first image. The system identified and tracked the desired objects within each image. The code was not constructed to work without some human intervention. The software coordinated the identification of the key points of interest in the first frame by the operator and the computer tracked the movement through the visual field in subsequent frames using distinctive visual features. The software extracted the boundaries and coordinates of the Aols from the videos captured. It used established image processing techniques [29].

The task was divided into three parts:

1. Train the system on the first frame
2. Iterate through subsequent frames using the information to locate the Pols and record their coordinates
3. Calculate the relative position of the gaze from the Pols and determine a glance or fixation and display the information graphically

Details of the first two stages are reported here and some initial results from the analysis shown.

#### Training on first frame

The system presents the first frame and prompts the operator to identify all the Aols for the particular task and the system then extracts the features: Colour, shape and Centre of Mass and stores them, Figure 4 a).

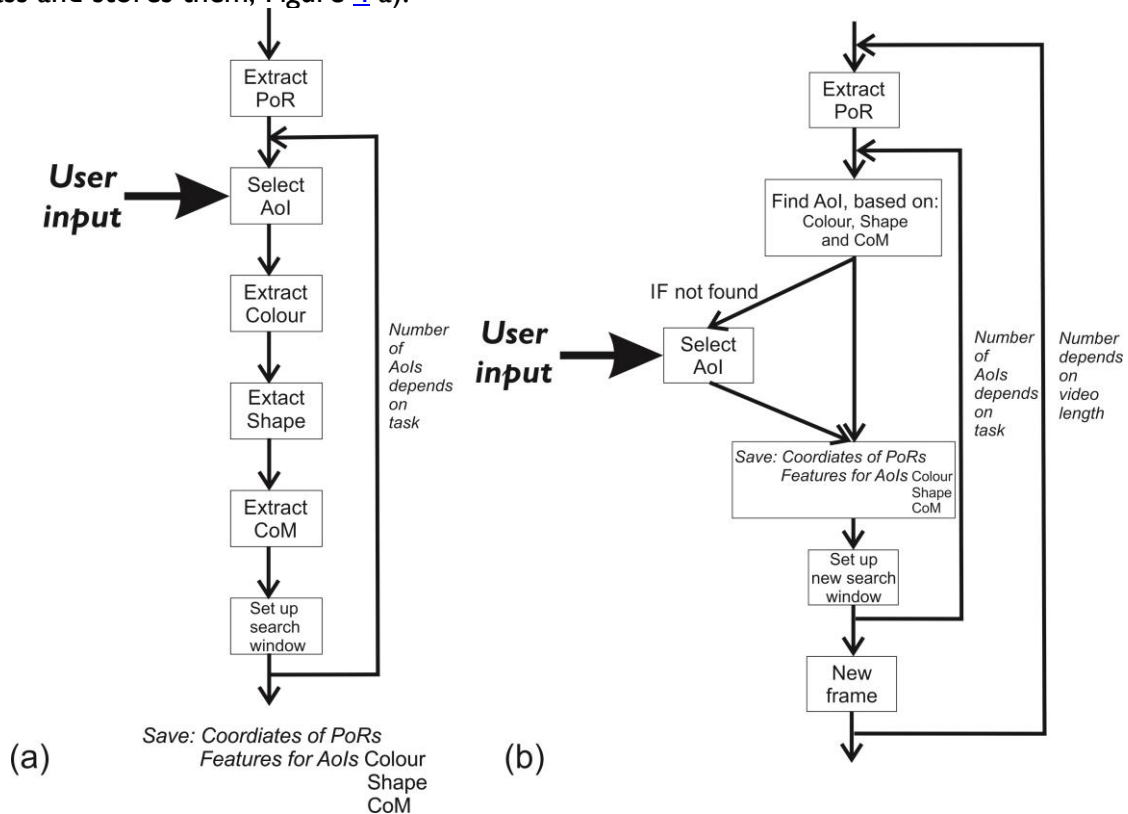


Figure 4: Extraction of Aols and PoRs. a) Training the system on the first frame for the features; colour, shape and movement. b) Extraction of Aols in subsequent frames.

## Identification on subsequent frames

For all frames in the video the basic sequence remains similar, until the end of the video: Extract cross hair position, using the feature information and search windows to detect all Aols. If one cannot be found prompt the user for input, otherwise update the features and create a new search window, Figure 4 b).

### 3.1.1 Object acquisition

The first step to acquire the data is to locate the objects in the scene. This is done by identifying their boundaries and hence recognizing the items as separate entities. For this two spatio-temporal methods were employed [30], one based on colour and the other on edge segmentation.

**Colour based segmentation** This creates segments on the basis of similarities in colour intensity. The intensity is extracted from maps of the fundamental colours (red, green and blue) [31]. The precise value of the intensity depends on the particular illumination level and a single object can possess different colour attributes at different times. To circumvent this, a measure of 'nuance' was implemented. The colour descriptors were derived from the normalized RGB channels, and were referred to as 'redness', 'blueness' and 'greenness'. The code accepts a range of colour values around the core values of red, green and blue. For an 8 bit map the Red colour is 255 and the other two colours are zero. However, to have 'redness', the red colour need only 70% of the maximum value and maximum of 20% of either or both of Blue and Green. To ensure there was a full range of colours to analyse the ranges were normalised first. The the mean intensity and standard deviations for the RGB were calculated and the nuanced images produced. Two nuanced colour maps are seen in of Figure 5, together with the original image.

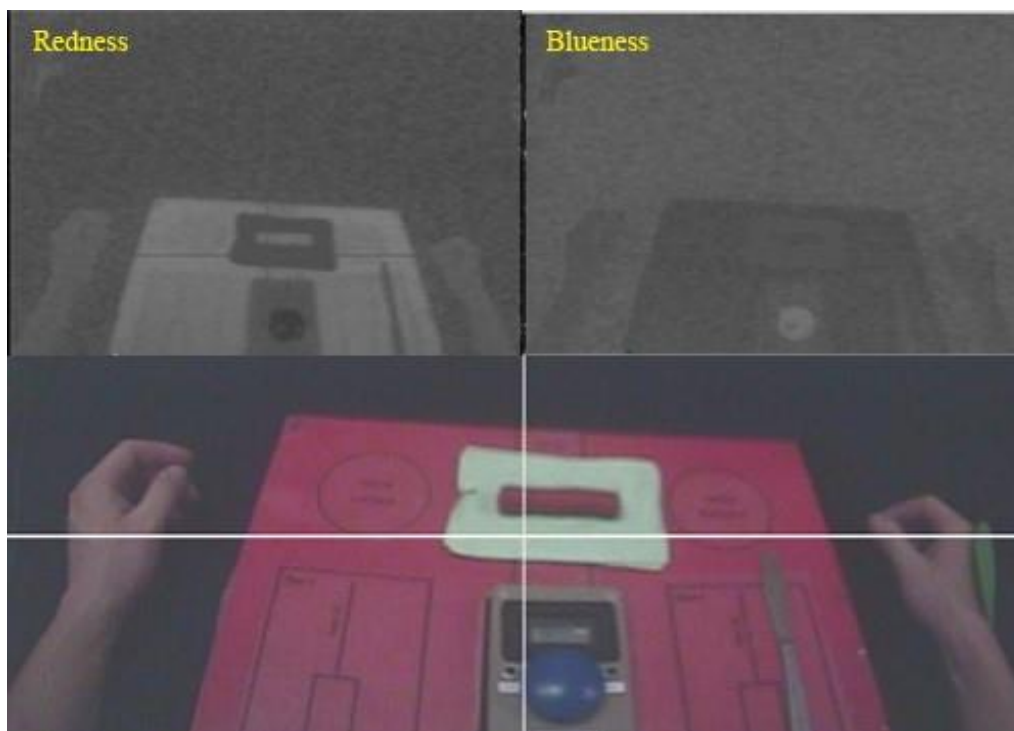


Figure 5: Red and Blue 'nuance' colour images, compared with the primary image. Colour of interest is shown as white in the nuance colour images.



**Edge detection** As colour segmentation cannot discriminate between similarly coloured objects the next stage was to use edge detection techniques. Edge detection is based on detecting sudden changes in intensity [29]. Thus derivative of the intensity in the image locates the majority of edges. The interface applied Canny edge detection methods from the Matlab signal processing toolbox, This applies a Gaussian filter on the data and then calculates the derivative of the image [32]. Figure 5 shows the edges on the primary image.

An example of the shape extraction for the button is indicated in Figure 6. The user inputs a selection of the button by pointing the mouse at any part of the button. The computer constructs a bounding box around the area that is roughly the same colour. At the same time the convex shape based on the edge detection algorithm determines the size of the bounding box. The colour mask then is applied to the shape. The top of the button has a flare from the lights of the room so that to the camera it is not all the same colour, but by combining the colour characteristics and the the edge detection results the final shape approximation is created. In subsequent frames the bounding box can be applied and the button reacquired.

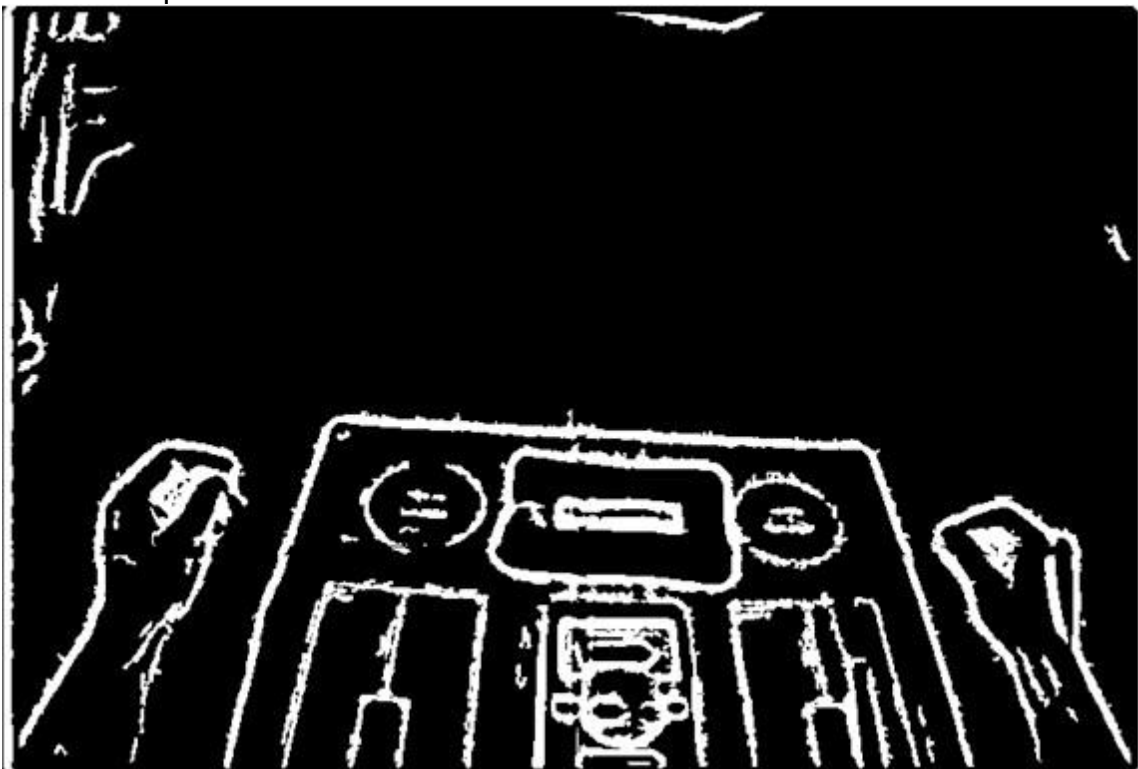


Figure 6: Result of applying edge detection on the primary image.

**Point of Regard tracking** Once the bounding box is defined around the object it is possible to define when the subject is looking at that particular area of interest. This occurs when the cross-hairs are inside or close to the bounding box. From this the gaze behaviour can be built up from the analysis of the individual frames. 'Close' to the bounding box is defined as if the co-ordinates of the PoR to be within 10% of the size of the object, in this way the size of the target object is taken into account.

### 3.1.2 Search window

An object is tracked by being identified in every new frame. This search time to identify an object can be reduced by restricting the search area to that which an object can be reasonably be expected to move in the time between two consecutive frames. For each new

frame, the classifier aimed to identify the objects based on a maximum reasonable distance between the observed feature vectors.

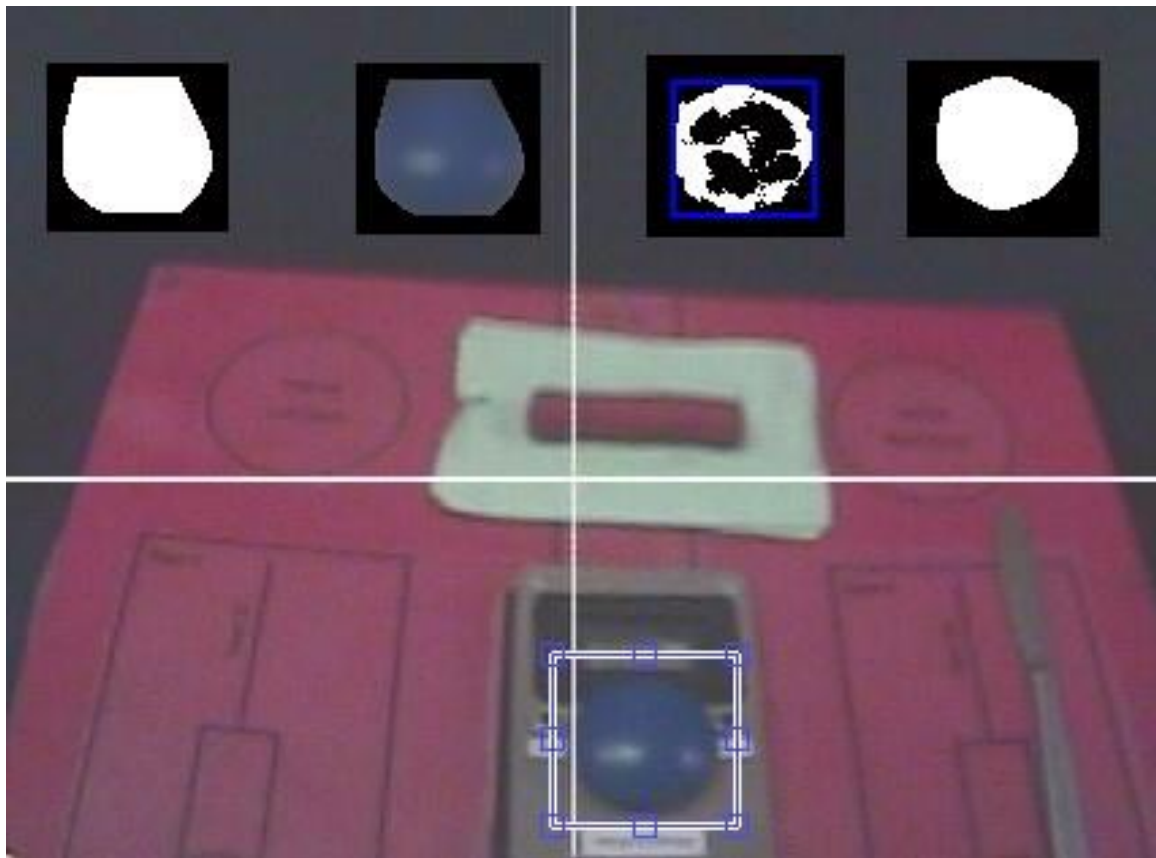


Figure 7: The process of assisted segmentation. The user indicates an item of interest (the time button) and based on the edge detection and colour matching results the computer can track the position of the button in the frame.

The sequential frames are 33ms apart. An object in front of the subject, between them and the form board is up to 300mm from the camera. The form board is 430mm by 310mm. A maximum velocity was set to be  $0.2 \text{ ms}^{-1}$  (four times the peak reaching speed recorded by Bouwsema [33]). If it is moving perpendicular to the visual field this corresponds to 33mm between frames. The furthest edge of the form board occupies half of the frame and for the output of the IScan this corresponds to 720 by 540 pixels (see Figure 3). Thus the range limit is one tenth of the width of the visual field (72 pixels). The search window for subsequent frames is set to be larger by 72 pixels all round. If the object is not found within this window the interface referred back to the operator for a new entry.

### 3.1.3 Data extracted

For analysis of the action the characteristics that were then extracted from the video data were the Euclidean distances between the PoR, and the CoM and bounding boxes of the Aols. The duration of individual fixations were measured, by counting the number of frames the PoR is within an Aol. These data are sufficient for concluding whether a subject is fixating an object or is glancing at it. Based on previous work [34,35] a fixation was defined as a time on the target greater than 200ms (seven frames). Once the data is obtained it is possible to gain insights in behaviour.

### 3.1.4 Subjects

Fourteen subjects without limb differences were recruited from the staff and students of the University of New Brunswick (UNB), nine males and five females, nine were right handed (mean age 29, range 18 to 48) and four users of prosthetic limbs volunteered through the limb clinic at UNB (ages not available). Research ethics for the experiment number, was obtained from UNBs Research Ethics Board (REB2010-099). All users had absences below the elbow and used standard single axis myoelectric limbs with friction wrists and were allowed to adjust the orientation of their prosthesis before each activity. A single prosthesis user performed the test twice, first with their prosthesis and second time with their contralateral hand, allowing additional comparisons.

After the ethics processes were completed, each subject donned the IScan headset and prior to the tests the system was calibrated. Subjects were then asked to perform the six SHAP tasks (Pouring water, Coins, Cutting, Abstract Lateral, Page turning, operating a Zip), for details [22,23]. Subjects then removed the eye tracker headset and performed a SHAP test according to the standard protocol.

## 4 Results

For this paper the emphasis is on the Cutting task. Results of the overall SHAP tasks and other details are included in a separate publication [36]. Due to limitations in the system, gaze data was not available for some instances (see 5). In a test of 100 gaze video sequences the semi-automated tracker only lost the various targets when the IScan device lost track of the gaze. There were 18 users with 6 tasks and 8 sequences where the fixation data was missing.

### 4.0.5 General observations of the task

Non-impaired subject tends to look at the button as they are reaching for it, but quickly changed their focus on to the knife before their hand had gotten to the button. Similarly, their gaze moved ahead to the plasticine while picking up the knife. Subjects may have looked briefly at the blade to ensure it is in the right position, especially if they have not handled the knife before. In contrast the impression from observing the video suggests that prosthesis users looked at the button until it was pressed and attended to the knife until they have it firmly in their grasp. They then fixated on the plasticine until they have finished cutting it and so on.

### 4.0.6 Video information

To assist in the understanding of the usual behaviour afforded by this tool, data for a single unimpaired subject is studied. The processed data is represented in the 2D plane of the image, this allows different ways to represent the data, examples are given here. Instead of showing the data relative to the static field of the observer, once acquired, it is possible to transform that data relative to different points of regard. The software can assess the relative position of the Point of Regard (PoR) with respect to the Centre of Mass (CoM) of each Object of Interest (OI) in the scene. Figure 7 shows an example of the position of a single OI, (the blue timer button, positioned here top left). For each frame, the COM is plotted at the origin ('stabilized' CoM). The PoRs are show as adjusted relative to the button. Their offset to the origin, (relative to the CoM), is plotted on the vertical and horizontal axes. Each cross represents the PoR in a specific frame in the video. The blue rectangles surrounding the origin represent the bounding boxes of the button, (i.e. the minimal area rectangles that completely confine the OI). There are multiple bounding boxes

for each frame, hence a heavy blue line. The crosses are the PoR and they are shown in blue outside the bounding box and red inside, suggesting that for those frames, the subject was fixating on that OI. Using this technique one can observe that the subject does not attend to the button except when they are reaching for it.

Figure 8 shows the time series of the same sequence. The horizontal axis is the frame number, hence the time. The vertical axis is the distance from the blue button CoMs. Red circles mark the frames when the PoR was inside the bounding box of the reference object (fixating on the button). This is the analysis of 150 frames or approximately five seconds of recorded video. The subject looks at the button only once. However, in Figure 9 which shows the Euclidean distances between the PoR and the CoM of another tracked object within the scene (the knife). It can be seen that the PoR is never within the bounding box of the OI, so the subject never looks directly at the knife, but see it in their peripheral vision only.

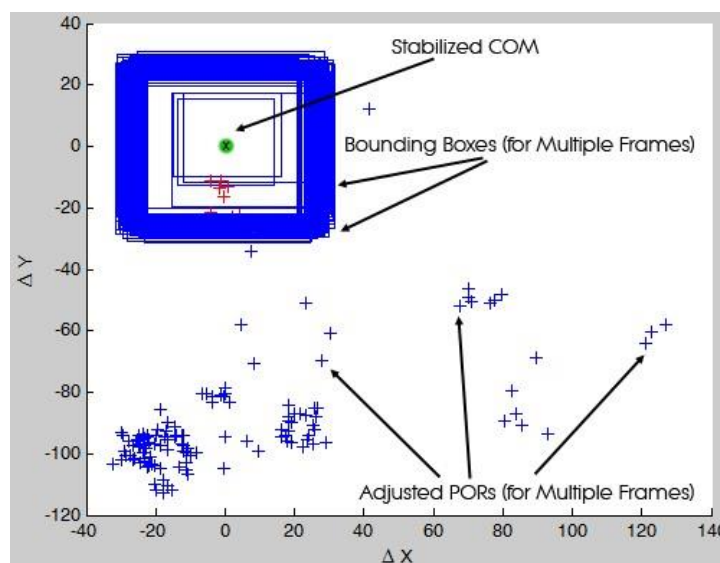


Figure 8: Relative PoRs to CoM coordinates for the cutting task for an unimpaired subject. The timer button is the stabilised as the centre of mass and is at the origin. Note: Upper left. The multiple PoRs are then show relative to it. The red crosses are when the attention is on the button.

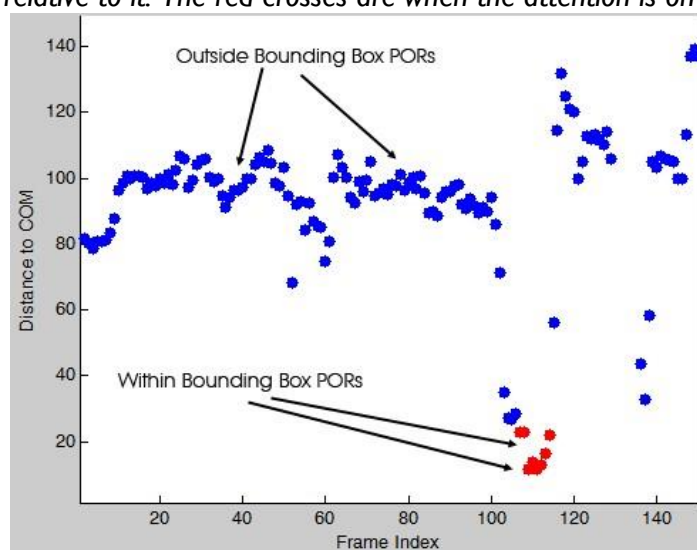


Figure 9 The motion of the point of regard (PoR) relative to the blue button (CoM) for an unimpaired subject. The Euclidean distance is on the Y-axis and the frame number (hence time) is on the X-axis. This unimpaired subject looks at the target before they reach towards it.

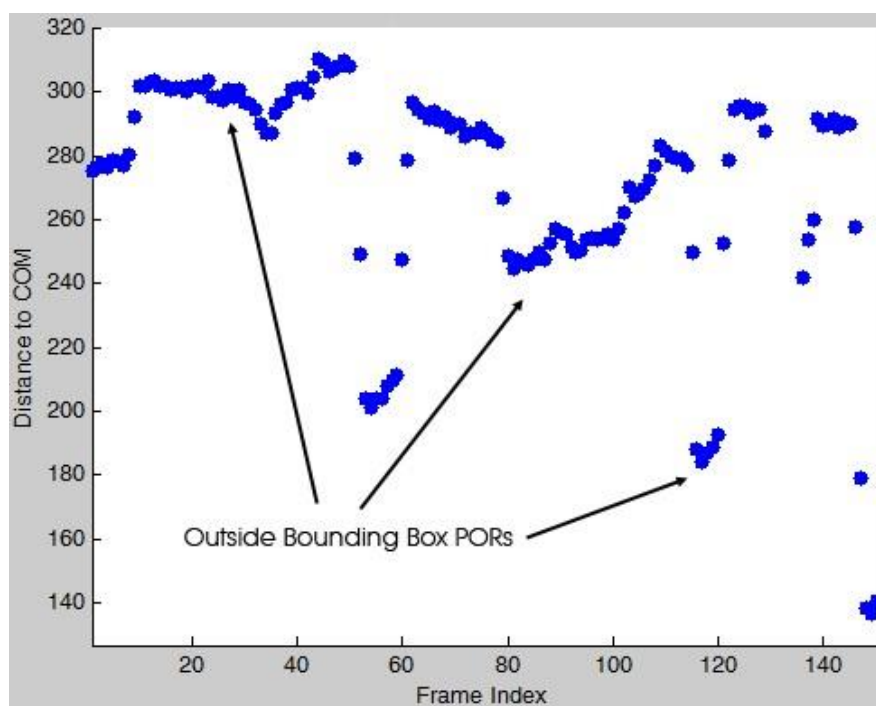


Figure 10 The motion of the point of regard (PoR) relative to the knife (CoM) for an unimpaired subject. The Euclidean distance is on the Y-axis and the frame number (hence time) is on the X-axis. The unimpaired subject looks never needs to look directly at the target. Merely keeping it in their peripheral vision.

#### 4.0.7 Analysis of the statistical data

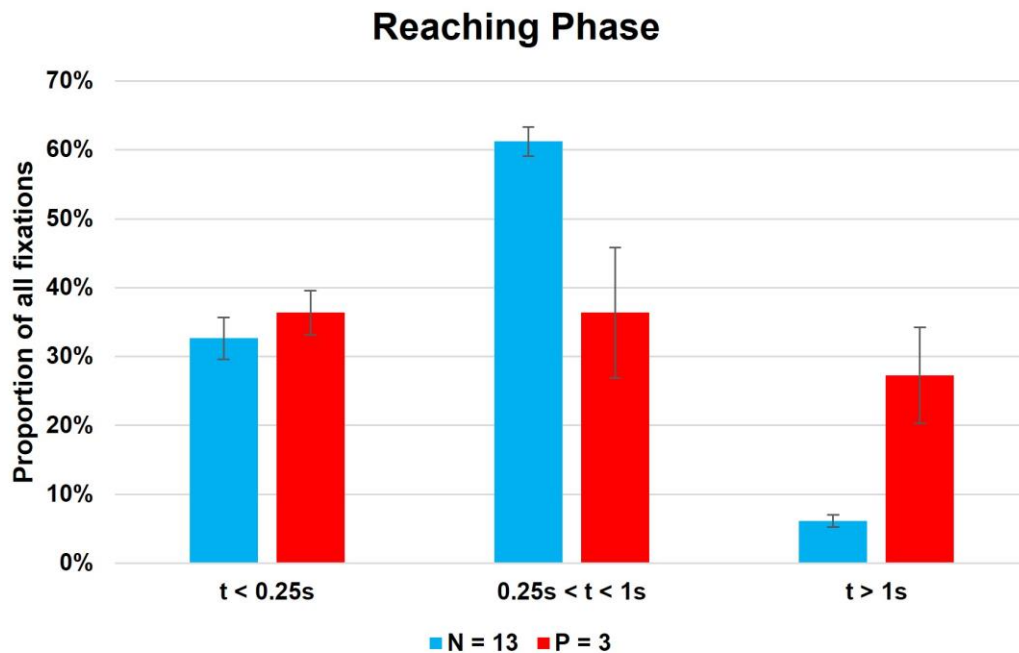
To demonstrate the other results that can be obtained from this analysis an aggregation of the data for cutting task is shown Figure 10. It depicts the proportions of the average number of gazes for both groups. The data is divided up into the reaching and manipulation phases. The data is further divided up into three temporal phases; below a quarter of a second, from that to one second and gazes longer than a second ('glances', 'looks' and 'stares').

The results show that when they employed their prosthesis the users showed a greater range of glances across the visual field. The proportion of the number of 'glances' were similar in both groups, but the users fixated on the target for longer. Note, the users take longer overall to perform the tasks. Unimpaired average trial duration is  $4.8 \pm 1.5s$ , with  $6 \pm 3$  fixations, while the prosthesis users are  $20 \pm 12s$  with  $14 \pm 10$  fixations. During the manipulation phase the number of glances was far higher in the prosthesis population while the unimpaired subjects stared directly at the plasticine.

## 5 Discussion

The system proved to be reliable enough to automate much of the process. There was a greater number of times the IScan system was unable to track the gaze of the subjects, than the custom software could track the OIs. This limitation can be reduced by ensuring the IScan system is set up correctly at the beginning. There are numerous factors that dictate the success of the system; position of the tracker on the head of the subject, level of illumination, the shape of the subjects orbits and if the visor carrying the scanner has been inadvertently moved by the subject. The experimenter needs to be vigilant in detecting and reducing these problems.

(a)



(b)

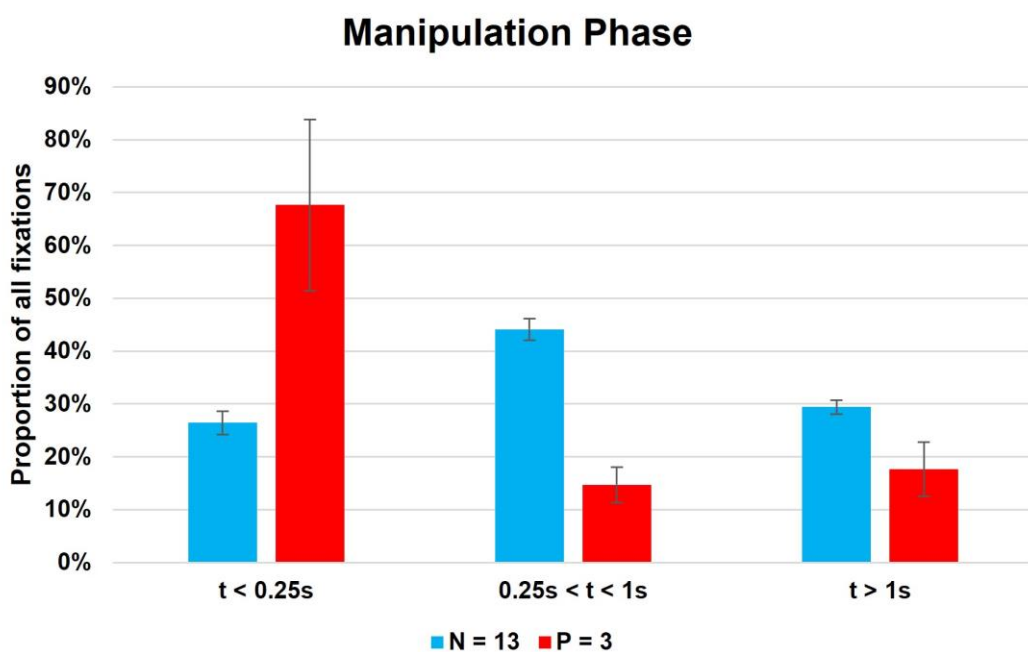


Figure 11: Number of fixations for the cutting task divided into three phases: Below 0.25s, between 0.25 and 1s and greater than one second. (a) Data for reaching phase, and (b) the manipulation phase. Error bars are standard deviation across subjects.

## 5.1 Gaze tracking

This simplified approach to the automation of the analysis of the visual field is novel in this application to prosthetics gaze tracking. The time to undertake the work of basic analysis was much reduced and the data produced was available for many different forms of subsequent analysis. It has been shown that it is possible to evaluate the gaze behaviour by identifying the frames at which the PoR is fixated at specific OIs or Aols with less user input

and thus needing reduced reliability on the skills of the operator compared with a standard process in [17].

Different insights can be gained from the different ways of creating and displaying the data. Time series data can allow reading of the single activity, collected numerical data allows aggregation. By counting the number of consecutive frames, the duration of each fixation can be extracted, and a measure of the amount of attention paid to any particular point of regard derived. An example is shown in Figure 9. The PoR was fixated at the timer button for ten consecutive frames. This amounts to 0.66% of the total time analysed (150 frames or roughly 5 seconds). Whereas in Figure 10, local minima (seen in frames 50 to 60, 115 to 120 and around frame 150, for the knife), shows that the PoR quickly approaches the CoM of that object of interest. This suggests either a glance towards the OI or a focus towards another OI that is in the vicinity. Study of the video frames shows there is nothing of interest at these points, thus the subject is simply glancing *towards* the target.

Data aggregation allows a nuanced analysis of the population data. One example from this task is the amount of time the operator looks at the button compared with the time regarding the knife. Comparing Figures 9 and 10 with the number and length of different fixations (Figure 11), shows how much attention the subjects need to attend to the sub tasks.

The key interest for this investigation is the comparison of the actions of the prosthesis users compared with the majority of the population. In both groups, reaching has a similar percentage of glances at objects (times less than 0.25s). The prosthesis users then 'stare' at the objects (such as the button) for far longer than the unimpaired (percentage of fixations greater than 1s, compared with mid length 'looks' between 0.25s and 1s). During the manipulation phase this is very different with the users spending more than 60% of the time glancing between the prosthesis and the task itself, while once again the unimpaired subjects spent the majority of the time looking at the plasticine. The inference from this it that the user of a prosthesis needs to monitor the prosthesis and the objects equally, moving their eyes away from the task to the hand and back, checking that the prosthesis is still retaining the object. This results in many more short fixations relative to the unimpaired subjects, who can rely on touch feedback to tell them if the object is slipping from the grasp.

## 5.2 Head orientation

Humans can acquire visual information using different behaviours. They can look directly at an object (foveate), glance at it or to use the peripheral vision to check on the location of important, but secondary details [5]. Additionally, the individual can divert their gaze towards the object, but for more concerted focus they can rotate their head towards the object and fix it in the centre of their visual field. This is likely to be associated with longer fixation times or when the priority given to the task. To extract this information from eye tracker data, one solution would be to require the computer to numerically transform the entire visual field and obtain the details of the rotations of what is seen [37,38]. Using the objects within SHAP makes the analysis far simpler: If the timer's bounding box remains in the lower centre of the frame then the subject is pointing ahead. Only if the button moves away from centre of the frame and the PoR moves towards the centre, is the subject is concentrating on the activity.

This difference can be observed by plotting all the PoRs on a single two dimensional field. If the subject turns their head towards a target the PoRs will be closer to the centre of the image, Figure 12. An illustrative comparison can be made between the behaviours of the single subject that performed the test with their prosthesis and their unimpaired hand. With their contralateral side the subject kept their head pointing forwards and used peripheral vision to track the activity (green crosses). With their prosthesis, the operator turned their head towards the knife on the left side of the form board, putting it closer to the centre of the visual field (red crosses towards the right side). The gaze can be seen to cover more of the visual field as they look about more from the greater spread of red crosses. This, coupled with the difference in fixation durations, paints a picture of someone with a far greater focus on the task.

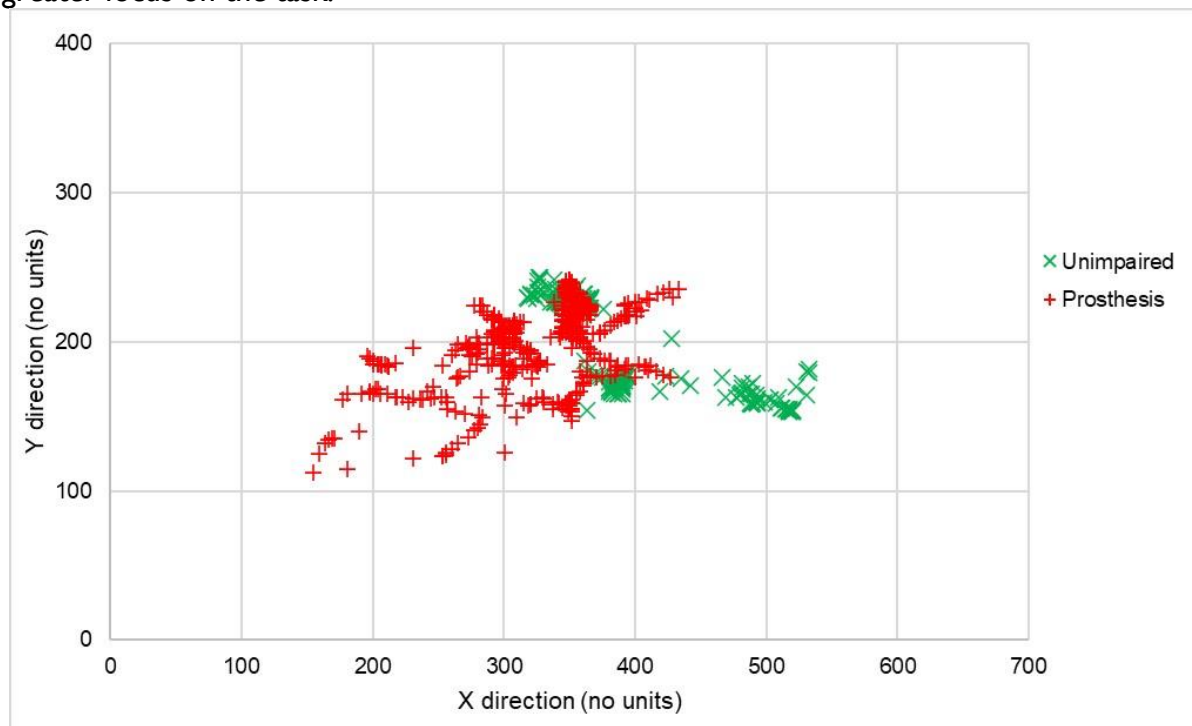


Figure 12: The range of regard across the visual field for the unimpaired arm and the prosthesis. Values are dimensionless numbers from the digitisation of the visual image. Green crosses are for an unimpaired hand, red are for a prosthesis.

Finally, the point of view of the subject provides clues as to how they compensate for the lack of flexibility in some of the joints in their prosthesis. Users of prosthetic limbs tend to compensate for the lack of flexibility in the wrist and forearm by using a combination of humeral abduction and trunk sway instead of flexion/extension and pro-supination [39]. This motion tends to cause the visual field to rotate, and this is clear in a comparison of frames during the pouring task; Figure 13. The first example is an early frame and the second is taken during the pour, showing considerable rotation of the body to the right in compensation.

The choice of task has a strong influence on the applicability of the assessment. As has been indicated, the greater the abstraction of the task the more general the conclusions, but these are more removed from real activities in the field [20]. SHAP was designed to create a balance between the contrasting pressures, clinically valid tasks that are sufficiently well controlled to allow broader conclusions about manipulation [22]. Other assessment tools have been developed that could also employ similar techniques to automate the process of



analysis [40,41,42]. Using SHAP tasks restricts the activity to small scale indoor manipulations, but as the hand is generally used in this circumstance this is not too great a restriction.



Figure 13: Example of the rotation of the visual perspective as a prosthetic user employs a combination of trunk sway and humeral abduction to compensate for a lack of pro-supination in their prosthesis. The subject has lent to the right as they pour the carton. The image rotates to the left. First image is the initial set up and the second during the pour.

An additional advantage of this method is it allows a measure of the closeness of the gaze to an object to be quantified. With conventional video analysis, the points of interest are marked by hand. If the size of an object of interest is considered it has to be arbitrarily assigned by the operator too. This software derives a size from the visible outline of the object and a centre of mass is based on the image itself. Any a fixation or a glance in the peripheral vision can be measured and recorded repeatedly and objectively.

### 5.3 Refinements

While the system proved satisfactory there are a number of possible refinements to automate the system further:

It is possible to detect trunk sway and measure its timing or extent from the data. Again the choice of a SHAP task makes it far simpler as the displacement of the button from the lower centre of the visual field combined with the angle of the form board, is sufficient to detect the motion without needing to derive the precise angle of flexion.

The system does not rely on any particular form of visual information, such as how the PoR is defined or shown on the images which is a particular characteristic of each gaze tracking system., This tool has been shown to work with other gaze tracking hardware (Dikablis Professional Wireless Eye Tracker, (<https://ergoneers.com/en/eye-tracker/dikablis-hdk-eye-tracker/>) [43]).

A refinement to the code would be to use the context of the known activity (such as cutting) to automatically search for the knife, button and plasticine in the early image based on their known colour where they should be in any SHAP sequence, although this level of automation will make the tool less general for use with other tasks. Additionally, it is possible to use the motion data of prior frames to restrict the search box further, predicting position based on the preceding motion of the last few frames, rather than an arbitrary constant maximum speed. Again this level of sophistication was not found to be needed in this application.

This interface has been used to automate the process of studying gaze during the learning process. When performing an unfamiliar task subjects tend to fixate for shorter periods of time compared with routine operations, thus it is more important to have objective measures of gaze as provided by this program.

## 6 Conclusion

Using a standardised format for a task, it is possible to create a simple interface that tracks the points of interest in a video sequence that records the motions and the point of regard within the visual field. Once the system has been given some points of interest in the first frame of the video sequence by the operator the system can track their motion through the activity of interest, determining how the subject looks at the key points on a visual field during the activity and suggests prosthesis users devote more focus and attention to a task. This system has been used to analyse the data from the tasks and investigate the behaviours when a task is new to the operators.

## References

- [1] P. Kyberd. *Making Hands: A History of Prosthetic Arms*. Academic Press, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK, 2021.
- [2] G. Westling and R.S. Johansson. Factors influencing the force control during precision grip. *Experimental brain research*, 53:277-284, 1984.
- [3] M. Land, N. Mennie, and J. Rusted. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311-1328, 1999.
- [4] B. Law, M.S. Atkins, Arthur E Kirkpatrick, and Alan J Lomax. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 symposium on eye tracking research & applications*, pages 41-48, 2004.
- [5] C. Vater, A. M. Williams, and E-J. Hossner. What do we see out of the corner of our eye? the role of visual pivots and gaze anchors in sport. *International Review of Sport and Exercise Psychology*, 13(1):81-103, 2020.
- [6] K. Y Cheng, M. Rehani, and J. S Hebert. A scoping review of eye tracking metrics used to assess visuomotor behaviours of upper limb prosthesis users. *Journal of NeuroEngineering and Rehabilitation*, 20(1):1-22, 2023.
- [7] H. Bouwsema, P.J. Kyberd, W. Hill, C.K. van der Sluis, and R.M. Bongers. Determining skill level in myoelectric prosthesis use with multiple outcome measures. *Journal of Rehabilitation Research and Development*, 49(9):1331-1348, 2012.
- [8] M. Sobuh, L.P.J. Kenney, A. Galpin, S. Thies, J. McLaughlin, J. Kulkarni, and P. Kyberd. Visuomotor behaviours when using a myoelectric prosthesis. *Journal of Neuro Engineering and Rehabilitation*, 11(17):72-83, 2014.
- [9] K. L Jackson, Z. Duric, S. M Engdahl, A. C Santago II, Siddhartha Sikdar, and Lynn H Gerber. Approaches for segmenting the reaching and targeting motion primitives in functional upper extremity reaching tasks. *IEEE Journal of Translational Engineering in Health and Medicine*, 2023.
- [10] J. Han, Ling Shao, D. Xu, and J. Shotton. Enhanced computer vision with Microsoft Kinect sensor: A review. *IEEE transactions on cybernetics*, 43(5):1318-1334, 2013.
- [11] J. Smisek, M. Jancosek, and T. Pajdla. 3D with Kinect. *Consumer depth cameras for computer vision: Research topics and applications*, pages 3-25, 2013.
- [12] M. F Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision research*, 41(25-26):3559-3565, 2001.
- [13] F.A. Popa and P.J. Kyberd. Identification of patterns in upper limb prosthetic usage by analysis of visual attention to areas of interest. In "MEC'11 Symposium - Raising the Standard", *Proceedings of the 2011 MyoElectric Controls/Powered Prosthetics Symposium*, pages 226-229, Fredericton, New Brunswick, Canada, August 14-19 2011. Institute of Biomedical Engineering.
- [14] JVV Parr, S, J Vine, NR Harrison, and G. Wood. Examining the spatiotemporal disruption to gaze when using a myoelectric prosthetic hand. *Journal of motor behavior*, 50(4):416-425, 2018.
- [15] V. Gregori, M. Cognolato, G. Saetta, M. Atzori, Megane Pro Consortium, and Arjan Gijsberts. On the visuomotor behavior of amputees and able-bodied people during grasping. *Frontiers in bioengineering and biotechnology*, 7:316, 2019.
- [16] K. Y Cheng, C. S Chapman, and J. S Hebert. Spatiotemporal coupling of hand and eye movements when using a myoelectric prosthetic hand. In *2022 International Conference on Rehabilitation Robotics (ICORR)*, pages 1-6. IEEE, 2022.

- [17] M Sobuh, L Kenney, A Galpin, S Thies, P Kyberd, and R Raffi. Coding scheme for characterising gaze behaviour of prosthetic use. In *Myoelectric Controls Symposium - Raising the Standard*, pages 106-109, 2011.
- [18] N. Pokhrel, S. Snäll, O. I Heimo, U. Sarwar, A. Airola, and T. Säntti. Accelerating image processing using reduced precision calculation convolution engines. *Journal of Signal Processing Systems*, pages 1-12, 2023.
- [19] S Gnanapriya and K Rahimunnisa. Helping hand: A GMM-based real-time assistive device for disabled using hand gestures. In *Machine Learning Techniques for Smart City Applications: Trends and Solutions*, pages 23-33. Springer, 2022.
- [20] P.J Kyberd. Outcome measures. In O.C Aszmann and D. Farina, editors, *Bionic Limb Reconstruction*. Springer Nature, Gewerbestrasse 11, 6330 Cham, Switzerland, 2021.
- [21] S.B.A Thies, LPJ Kenney, M Sobuh, AJ Galpin, P Kyberd, R Stine, and MJ Major. Skill assessment in upper limb myoelectric prosthesis users: Validation of a clinically feasible method for characterising upper limb temporal and amplitude variability during the performance of functional tasks. *Medical Engineering and Physics*, 47:37-143.
- [22] C.M. Light, P.H. Chappell, and P.J. Kyberd. Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function. *Archive of Physical Medicine and Rehabilitation*, 83(6):776-783, 2002.
- [23] P. J. Kyberd, A. Murgia, M. Gasson, T. Tjerks, C. Metcalf, P. H. Chappell, K. Warwick, S. E. M. Lawson, and T. Barnhill. Case studies to demonstrate the range of application of the Southampton Hand Assessment Procedure. *British journal of Occupational Therapy*, 72(5):212-218, May 2009.
- [24] C.M. Light. *An Intelligent Hand Prosthesis and Evaluation of Pathological and Prosthetic Hand Function*. PhD thesis, Electrical Engineering Department, University of Southampton, 2000.
- [25] L. Resnik, M. Borgia, J. M Cancio, J. Delikat, and P. Ni. Psychometric evaluation of the Southampton Hand Assessment Procedure (SHAP) in a sample of upper limb prosthesis users. *Journal of Hand Therapy*, 2021.
- [26] T. Tanaka, K. Taguchi, J. Shikano, I. Motomatsu, N. Ootaki, M. Nakagawa, T. Hamagushi, and M. Abo. Validity of the Japanese version of the Southampton Hand Assessment Procedure in stroke patients. *The Japanese Journal of Rehabilitation Medicine*, pages 18009-18009, 2019.
- [27] A. Murgia, P.J. Kyberd, P. Chappell, and C.M. Light. The use of gait analysis techniques in the measure of hand function. *Journal of Hand Surgery*, 30B(Supplement 1):83-84, June 2005.
- [28] M. Sobuh. *Visuomotor Behaviours During Functional Task Performance with a Myoelectric Prosthesis*. University of Salford (United Kingdom), 2012.
- [29] M. Nixon and A. Aguado. *Feature extraction and image processing for computer vision*. Academic press, 4th edition, 2019.
- [30] D. Zhang and G. Lu. Segmentation of moving objects in image sequence: A review. *Circuits, Systems and Signal Processing*, 20:143-183, 2001.
- [31] Q. Zaidi and M. Bostic. Colour strategies for object identification. *Vision research*, 48(26):2673-2681, 2008.
- [32] MATLAB. version 9.13.0 (R2022b). The MathWorks Incorporated, Natick, Massachusetts, USA, 2022.

- [33] H. Bouwsema, C.K. van der Sluis, and R.M. Bongers. Movement characteristics of upper extremity prostheses during basic goal-directed tasks. *Clinical Biomechanics*, 25(6):523-529, 2010.
- [34] T. A Salthouse and C. L Ellis. Determinants of eye-fixation duration. *The American journal of psychology*, pages 207-234, 1980.
- [35] B. R Manor and E. Gordon. Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of neuroscience methods*, 128(1-2):85-93, 2003.
- [36] P. Kyberd, A. Popa, and T. Cojean. The role of visual attention in controlling a prosthesis in a novel task. *In Preparation*.
- [37] L. Ritchie and B.T. Sharpe. Gaze behaviour of a cellist: A case study. In *International Symposium on Performance Science*, 2023.
- [38] L. Ritchie and B.T. Sharpe. Gaze behaviour of a cellist: From sight-reading to performance. *PsyArXiv*, 2023.
- [39] A. Zinck. The investigation of compensatory movements in prosthesis users and the design of a novel wrist. University of New Brunswick, Masters, Department of Mechanical Engineering, University of New Brunswick, 2008.
- [40] A. Chadwell, L. Kenney, S. Thies, A. Galpin, and J. Head. The reality of myoelectric prostheses: Understanding what makes these devices difficult for some users to control. *Frontiers of Neurorobotics*, 10(7), 2016.
- [41] J. S Hebert, Q. A Boser, A. M Valevicius, H. Tanikawa, E. B Lavoie, A. H Vette, Patrick M Pilarski, and Craig S Chapman. Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. *JAMA network open*, 2(9), 2019.
- [42] M. M. White, W. Zhang, A. T Winslow, M. Zahabi, F. Zhang, H. Huang, and D. B Kaber. Usability comparison of conventional direct control versus pattern recognition control of transradial prostheses. *IEEE Transactions on Human-Machine Systems*, 47(6):1146-1157, 2017.
- [43] A. Chadwell. *The Reality of Myoelectric Prostheses: How do EMG skill, unpredictability of prosthesis response, and delays impact on user functionality and everyday prosthesis use?* PhD thesis, school of Health Sciences, University of Salford, 2018.