This paper appears in Journal of Medical Robotics Research, 2020.
https://doi.org/10.1142/S2424905X19500053

Comparison of Attentive and Explicit Eye Gaze Interfaces for Controlling Haptic Guidance of a Robotic Controller

Javier L. Castellanos-Cruz[a], María F. Gómez-Medina[a], Mahdi Tavakoli[b], Patrick M. Pilarski[c], and Kim Adams[a]

[a] *Faculty of Rehabilitation Medicine, University of Alberta, 116 St 85 Ave, Edmonton, AB, T6G 2R3, Canada*
E-mail: javierle@ualberta.ca

[b] *Electrical and Computer Engineering Department, University of Alberta, 116 St 85 Ave, Edmonton, AB, T6G 2R3, Canada*

[b] *Department of Medicine, University of Alberta, 116 St 85 Ave, Edmonton, AB, T6G 2R3, Canada*

Children with physical impairments may have challenges in play due to limitations in reaching and handling objects. Telerobotic systems that provide guidance towards toys may help provide access to play, but intuitive methods to control the guidance is required. As a first step towards this, adults without physical impairment tested two eye gaze interfaces. One was an attentive user interface that predicts the target toy that users want to reach using a neural network, trained to recognize the movements performed on the user-side robot and the user's point of gaze. The other interface was an explicit eye input interface that detects the toy that a user fixates on for at least 500ms. This study compared the performance and advantages of each interface in a whack-a-mole game. The purpose was to test the feasibility of activating haptic guidance towards toys with an attentive interface and to assure the safety of the system before children use it. The prediction accuracy of the attentive interface was 86.4% on average, compared to 100% with the explicit interface, thus, seven participants preferred using the explicit interface over the attentive interface. However, using the attentive user interface was significantly faster, and it was less tiring on the eyes. Ways to improve the accuracy of the attentive eye gaze interface are suggested.

Keywords: Attentive user interface; Eye gaze interface; Explicit eye input interface; Haptic guidance; Neural network.

1. Introduction

Children with physical impairments face significant challenges to play because of their physical limitations, for example in reaching and handling objects [1]. Children with physical impairments may have less access to opportunities to play, thus leading to negative impacts on their social, emotional, or psychological development [2]. Assistive technologies such as assistive robots can be a tool to help children with physical impairments to play, giving children the opportunity to manipulate objects and interact with their surroundings. For example, robotic arms and car-like robots can be teleoperated by switches, allowing children to interact with objects in the environment at a distance (e.g., from their wheelchairs) [3], [4]. Playing using robots can contribute to a child's independence, cognitive, social and linguistic skills [5]. Current robots for play do not include physical feedback to inform the user about the properties of the objects that the robot is touching [6]. Haptic interfaces can exert forces to provide haptic feedback (the sense of touch) to the user [7]. Haptic feedback can benefit the user's understanding and exploration of the environment, making it possible to perceive object properties such as shape and size [8].

Haptic interfaces can also exert forces with the purpose of helping and guiding the user to complete and increase their performance in manual tasks [9]. With guidance children with physical impairments might be able to use a joystick-like interface to control a robot, rather than just switches, which are limiting in function. Haptic-enabled robots could provide this guidance. A pen-like haptic interface was used to help children with cerebral palsy to write Chinese characters [10]. The haptic guidance consisted of forces that attracted and kept the children's hand on the trajectory of the character. Haptic guidance can also be implemented as forbidden region virtual fixtures (FRVF) which constrain the robot to stay inside or outside a predefined region. For example, FRVF can be used for supporting colouring where forces are applied on the haptic user interface whenever the user tries to colour outside the predefine shape template [11]. Also, FRVF can define the space of a trajectory for transport of objects from one location to another, for example, to support sorting [12]. However, these types of haptic guidance are commonly implemented for a specific task, and they do not necessarily correspond to the user's movement intentions [9].

Eye gaze systems could be used to orient the haptic guidance towards the target toys that children intend to reach. The state-of-the-art of eye gaze-controlled robots mainly involves explicit eye input interfaces. This type of interface consists of voluntarily and consciously controlling one's eye movements, or gaze direction [13]. Explicit interfaces relate the user's eye movements and point of gaze (POG), i.e. the location where the user's gaze is focused, to robot commands. For example, voluntary blinking and closing the eyes has been used as the 'stop' command for mobile robots such as electric powered wheelchairs [14]. Gaze gestures or patterns are another form of explicit eye inputs. For example, electric wheelchairs can be commanded to move forward if the user looks upwards and downwards [15].

A user's eye gaze can also be used as a pointer. A mobile robot was controlled by selecting buttons on a computer screen that commanded the robot to stop and move forward, left, right and backwards [16]. The user had to fixate on the buttons for a period, known as dwell time, to activate the button's action. Dwell times are a form of solving, to some extent, the Midas Touch problem. This problem relates to discriminating between intended and unintended activation of commands on the screen whenever the POG goes over them [17]. Fixating during the dwell time is a way of confirming that the user has the intention of activating the command. Electric wheelchairs have been controlled by fixating at a computer screen that displays arrow-buttons for direction commands [18]. The eye gaze can be combined with other types of devices to control robots. For instance, users preferred to control the robot's motion with the joystick and use their eye gaze to control the rotation of the robot's mounted camera [19]. In the case of a drone, it was easier and more reliable to control the rotation and speed with the gaze, and translation and altitude with buttons [20]. The eye gaze has also been employed to point out or indicate which object the user wants to reach with a robot, for instance, using a robotic arm and an exoskeleton [21], [22]. Using an explicit interface requires making predefined eye gaze patterns, or fixating for a dwell time at a location to trigger an action, which can be uncomfortable and tiring to the eyes [13].

Encarnação et al. [23] tested an explicit eye input interface with three children with cerebral palsy, two of them were three years old and the other child was six years old. Children controlled the movement of a Lego robot by looking at a computer screen that displayed buttons to move forward, backwards, turn left and right. Children had trouble operating the robot because they had to fixate on the screen to make a selection and then look back at the robot to observe its action. The three-year olds were not able to complete the activities due to the complexity to control the robot.

Unlike explicit interfaces, attentive user interfaces do not require the user to change his/her eye behaviour. Instead of taking explicit commands, attentive interfaces interpret and process the user's natural eye movements, and respond according to them [13]. An example of an attentive interface is the recognition of activities (e.g., evaluating statistical graphs or completing a quiz) by interpreting the user's eye gaze while the user is interacting with a computer [24]. A laparoscopic system for minimally invasive surgery interpreted the surgeon's eye-gaze patterns and determine the viewing target where the laparoscope should focus [25]. Predicting targets during the interaction with computer applications was also possible by analyzing the user's eye gaze [26], [27].

The aim of this study was to develop an attentive interface to interpret the user's eye movements and select the target toy to reach with the robot. This study is the continuation of a previous preliminary study [28], where a multilayer perceptron neural network was trained to predict the target that adults were supposed to move a robot towards. The current paper describes the development and testing of a telerobotic haptic system, which integrates a neural network into an attentive user interface to activate haptic guidance, with the purpose of guiding the robot interface towards target toys. The goal is for children with physical impairments, who are seated in their wheelchair, to use a joystick-like interface positioned on their wheelchair to control a robot situated in the play environment. The robot will give guidance through the interface towards the toys of their choice, determined from their eye gaze. At this stage, the robotic system was tested by adults without impairments and not with children or adults with physical impairments, in order to test the feasibility of applying haptic guidance with an attentive interface, to test its performance, identify possible improvements, and to ensure the safety of the system before children use it. The objectives of this study were: 1) To create and test the performance of an attentive eye gaze interface for activating the haptic guidance towards the correct target toy during a game. 2) To compare the performance of the attentive interface with an explicit eye gaze interface that employed dwell times.

2. Methods

The study consisted of two sessions. The first session was carried out to collect data for developing the attentive interface. The second session was carried out to address the objectives of this study. Each session took between 30 and 60 minutes and were scheduled one month apart.

2.1. *Participants*

Ten adults without physical or cognitive impairments participated in this study. Two of them were females. The participants' ages ranged from 18 to 36 years old (Mean=25.4, SD=5.34, mean and standard deviation). Four participants had normal and six had corrected-to-normal vision. Ethical approval was obtained from the Health Research Ethics

Board at the University of Alberta.

2.2. *Materials*

The telerobotic haptic system included two 3-DoF PHANToM Premium 1.5A haptic robots (3D Systems, Inc., Rock Hill, SC, USA). One of the robots was placed in the environment as the environment-side robot (slave robot) where it interacted with the target objects in the game. The other haptic robot was placed at the user side as the user-side robot (master robot), for participants to control. The robots were controlled in position-error-based bilateral teleoperation mode, thus the environment-side robot mimicked the movements performed by the user on the user-side robot. In a second computer, the Tobii EyeX (Tobii Technology, Stockholm, Sweden) stationary eye tracking system was interfaced to acquire the POG signal. The eye tracker's operating distance is between 54 and 80 cm. The POG x and y coordinates for the left and right eyes of the user were recorded at a sampling frequency of 40Hz. Additionally, the robots' PC and the eye tracker's PC communicated via Ethernet and used the user datagram protocol (UDP).

Fig. 1 shows a participant controlling the robotic system, the main components and the setup of the system and the activity. More details about the system are described in [28].

The chosen activity for testing the robotic system was a Whack-A-Mole Arcade Game by Fischer-Price. The game was customized to light up and turn off the lights of the five moles, and detect the pressing of the moles, using a Arduino Leonardo microcontroller.

A stand was set up as illustrated in Fig. 1 so that participants had to sit behind it and look through the hole. This way, the participants' eyes were within the operating distance of the eye tracker (i.e., 45 to 80 cm), and avoided losing the calibration of the eye tracker with respect to the location of the moles due to head movements. The distance between the eye tracker and the hole was approximately 65cm, and from the rear moles to the hole, it was about 90cm.

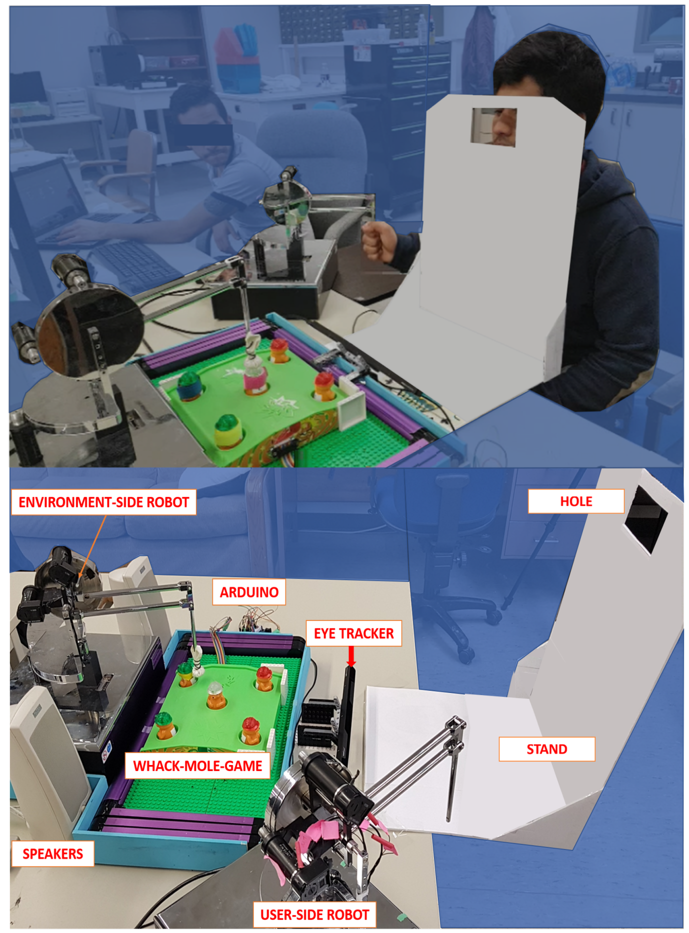

Fig. 1. A participant controlling the robotic system, and illustration of the setup of the system and the activity. Participants sat behind the stand and looked through the hole, to maintain the eyes within the operating distance of the eye tracker. They manipulated the user-side robot with their non-dominant hand (left in the figure) to make the environment-side robot "whack" the moles when they lit up.

2.3. *Haptic guidance*

FRVF with the shape of a cone was chosen as the guidance method for this study. The cone allowed the user to move the robot end-effector closer to the target and prevented the user from moving away from it. The design takes into account the findings of a previous study [12], where an adult with cerebral palsy tested FRVF in a sorting task. The shape of the FRVF was a cylinder that went from the place where she had to pick up the objects to a bin where she was supposed to drop them. The findings of that study indicated that the cylindrical shape of the FRVF opposed the natural and preferred motion of the user, which was an arc-like movement. The cone shape gives users more freedom to move the way they want.

The design of the FRVF is illustrated in Fig. 2A. The 3D cone was aligned along the line that crosses the target mole's and the robot's end-effector coordinates, the origin of the cone was at 1cm behind the robot's end-effector, and the cone angle was 30 degrees. The 1 cm distance was set so that the user did not feel like he/she was always up against the cone. The orientation and the origin of the cone were updated as the user moved closer to the target. If the user attempted to go outside the cone, forces were applied perpendicularly to the cone surface. The force was proportional to the distance between the robot end-effector's position and the cone surface, $F = K * |distance|^2$ (N/m). K was the constant that regulated the haptic guidance force. The K constant was set to 50 (N/m) for this study for both eye gaze interfaces. As pressing the moles was somewhat difficult, a potential field was implemented to attract the robot end-effector towards a virtual vertical line over top of the mole, as illustrated in Fig. 2B. The potential field helped the user to keep the robot end-effector over top of the mole so that the user just had to push down. This form of haptic guidance activated only when the environment-side robot end-effector was within 1.5cm from the virtual line extending vertically through the mole (the mole's x and y coordinates). Once the potential field was activated, the cone-shaped guidance was deactivated. The cone-shaped guidance was activated again, and the potential field deactivated when the target changed. The force of the potential field was proportional to the distance between the robot end-effector's position and the virtual line. For the attentive interface, the K constant was 10 (N/m). In contrast, the constant was set to 50 (N/m) for the explicit interface. The reason for the different constants is discussed later.

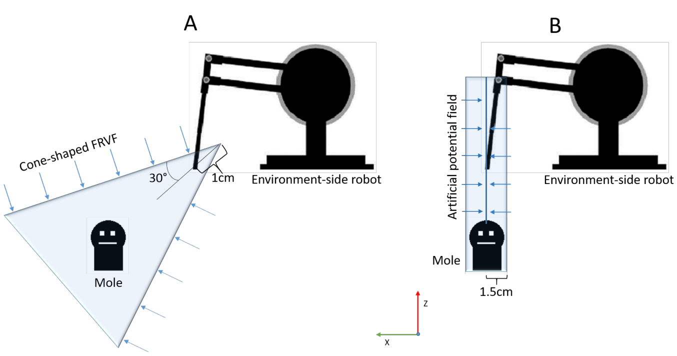

Fig. 2. Illustration of the haptic guidance projected from 3D to a 2D plane: A) 3D cone-shaped FRVF with origin 1cm behind the environment-side robot's end-effector and projected towards the mole. The origin's location was updated when the end-effector moved closer to the mole. The cone-shaped guidance helped a user move towards a mole and prevented them from moving away from it; B) 3D artificial potential field helped to keep the end-effector over the virtual line on top of the mole, thus making it easier to whack it. The artificial potential field was activated, and the cone-shaped guidance deactivated, when the end-effector was 1.5cm away from the virtual line.

2.4. *Eye gaze interfaces*

2.4.1. *Attentive user interface*

This interface consisted of implementing an artificial neural network for predicting the target mole that the user wanted to reach. The neural network analyzed the user's eye gaze and the movements performed at the environment-side robot. The output of the neural network activated the haptic guidance to guide the user towards the predicted mole.

The performance of training three multilayer perceptron (MLP) neural networks with combinations of input variables was compared to decide which structure to use in session 2: 1) the POG's x, and y coordinates, 2) the POG's x and y coordinates, and the environment-side robot's x, y, z positions, and 3) the POG's x and y coordinates, the environment-side robot's x, y, z position's, velocity and direction. The input layer had 80, 200, and 360 nodes, respectively. For the hidden layer, the number of nodes was explored from 0 to the size of the input layer. The neural networks were implemented as classifiers, therefore all of them had five nodes in the output layer, one for each mole in the game. A 5-fold cross-validation was performed, thus the neural networks were trained on 80% of the dataset and tested on the remaining 20%. The training was done using the scaled conjugate gradient method, having $\sigma = 5.0e - 5$ and $\lambda = 5.0e - 7$.

The accuracies of the neural networks with the three variable combinations were $84.91 \pm 5.02\%$, $90.06 \pm 2.44\%$, and $90.13 \pm 2.27\%$ (mean and standard deviation), respectively. Input variable combination number three was chosen due to the accuracy and because the neural network did not required nodes in the hidden layer, which makes it simpler and computationally less expensive. Fig. 3 illustrates this neural network. The neural network was trained one more time using the complete dataset and was integrated it in the Simulink code for activating the haptic guidance in session 2.

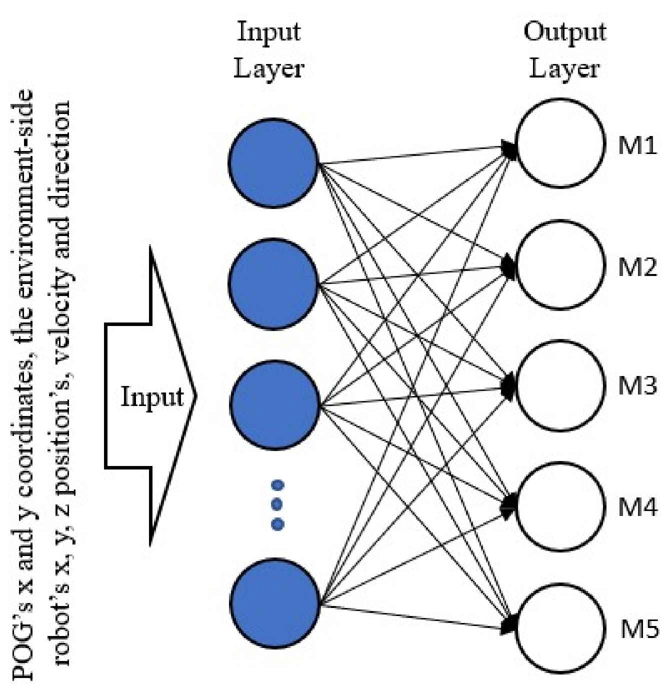

Fig. 3. Illustration of the neural network that was chosen to implement in session 2.

2.4.2. *Explicit eye input interface*

This interface was designed using dwell times, and worked in a two-step process to activate the haptic guidance towards the moles. First, the user was required to fixate at the middle light of the eye tracker (herein called change-selection-spot) for at least 500ms. The computer produced a beep sound when the user fixated successfully at the change-selection-spot. Second, the user had to fixate at the desired mole for 500ms. The computer spoke out loud the mole ID number (i.e., 1-5) to let the user know which mole he/she had selected. The user could repeat the two-step process in case he/she wanted to go to a different mole.

2.5. ***Procedure***

2.5.1. *Session 1*

The first session was carried out with the purpose of collecting the data needed to train the MLP neural network of the attentive interface. Participants were asked to use the robot interface with their non-dominant hand to whack 60 moles without haptic guidance. Three moles were lit up at a time and after whacking them the next three moles were lit up after 1 second. Each mole corresponded to one episode, which consisted of the time interval that the user took to whack the mole. The moles were lit up randomly. To simulate involuntary movements on the users, the y and z axes of the teleoperation were mirrored, i.e. when the user-side robot moved to the left or upwards the environment-side robot moved in the opposite direction, right and downwards, respectively. The purpose of the mirroring was to encourage the participants to move the robot interface in wrong directions such that the haptic guidance that would be applied in the second session would be engaged. Participants had the chance to get familiarized with the system by whacking each mole twice.

2.5.2. *Session 2*

The second session consisted of three parts with the purpose of evaluating the system. There was a five-minute break between each part of the session. In the three parts, the participants did the activity by looking through the hole of a stand, as illustrated in Fig. 1. Participants did the activity by looking through the hole to ensure that the participants' eyes were within the operating distance of the eye tracker, and to avoid losing the calibration of the eye tracker. The distance between the eye tracker and the hole was approximately 65cm, and from the rear moles to the hole, it was about 90cm. Before starting the activity, the eye tracker was calibrated with respect to the five moles. Then, participants had the chance to get familiarized with the system before starting each part of the session by whacking each mole twice.

In the first part, participants did the activity using the system in normal teleoperation ("normal teleoperation" condition), i.e. the robot axes were not mirrored as in session 1 and haptic guidance was not applied. They whacked a total of 60 moles.

The second part was focused on evaluating the attentive eye gaze interface and its neural network. The evaluation was carried out using an experimental crossover design. For this part, participants whacked 120 moles. As in session 1, the teleoperation was mirrored. To control for learning effects, the starting condition was randomized and counterbalanced, i.e. half of the participants started with the haptic guidance ("with guidance" condition) and the other half without haptic guidance ("without guidance" condition). Additionally, the condition was changed multiple times in order to have more reliable responses from the participants. The 120 moles were divided into 10 sets of 12 moles, in which the condition, "with guidance" or "without guidance", was randomly assigned, with a maximum of two consecutive sets with the same condition. A short break was given between sets to ask the participants whether it was easier than the previous set, but they were not told if guidance was on or off. Participants did five sets in the "with guidance" condition and five in the "without guidance" condition. At the end of the trial, participants were asked if their eyes felt tired and their responses were recorded by the researcher into the research notes.

The third part had the purpose of testing the explicit eye input interface. Participants were asked to whack 60 moles having the axis mirrored as in session 1, with haptic guidance. At the end of the trial, the participants were asked if their eyes felt tired. Also, they were asked to comment on which interface they preferred to use, the explicit interface or the attentive interface, and the reason why. Participant's responses were also recorded into the research notes.

2.6. ***Data collection and analysis***

In session 1, the user's POG and the environment-side robot's x, y, z positions were collected. The user's POG x and y coordinates for the left and right eyes were averaged. The velocity and direction of the environment-side robot were derived from its x, y, z positions. As each participant whacked 60 moles, a dataset of 600 episodes was created. Seven were excluded because the eye gaze data was lost. Additionally, 49 episodes were also excluded because they were shorter than 1s, which is the window size that was chosen for creating the training dataset for the neural network. Those episodes that were shorter than 1s occurred because the user was already close to the mole that was lit up. Thus, 544 episodes were considered for training the neural network of the attentive user interface. From the dataset, a training set of 57128 examples was constructed by having the moving window of one second (40 samples). In a previous study [28], the performance of MLPs trained with windows of 0.25, 0.5, 0.75 and 1 second were compared. Training the MLP with input data framed in a window of 1s had the best performance, thus, it was used in this study. In the present study, the training dataset contained nine input variables: x, y coordinates of eye gaze; the environment-side robot's position (x, y, z coordinates in space), velocity, and direction (α -angle respect to the x-axis, β -angle respect to the y-axis, and γ -angle respect to the z-axis).

In part 1 in session 2, only the environment-side robot's position was recorded since the eye gaze interfaces were not employed for part 1. Two episodes were excluded because of a malfunction with the whack-a-mole game. From, each episode the average time the user took to whack each mole, the distance travelled, and the jerkiness of the movements, were obtained. Jerkiness was measured by using the Dimensionless Jerk formula and taking the negative logarithm to improve the sensitivity (LDLJ) [29], [30]:

$$LDLJ = -ln\left(\frac{(t_2 - t_1)^3}{v_peak^2} * \int_{t1}^{t2} |\frac{d^2v(t)}{dt^2}|^2 dt\right) \quad (1)$$

where V is the velocity at which the environment-side robot was moving. LDLJ is a valid measure for measuring jerkiness of movements [30]. The lower the value of LDLJ the jerkier the movements.

From part 2 in session 2, the eye gaze and the environment-side robot's position were recorded. Episodes for the "with guidance" or "without guidance" conditions were analyzed separately. There was one episode excluded during the "without guidance" condition because of a malfunction with the whack-a-mole game. During the "with guidance" condition there were six excluded for this reason. Additionally, there were 32 episodes excluded because the user's eye gaze was lost due to head movements.

For the "with guidance" condition, the average time the user took to whack each mole, the distance travelled, and the jerkiness of the movements were measured. To analyze the neural network performance, the accuracy of the neural network to predict which mole the user was going to reach throughout each episode was measured. The analysis was performed after 300ms at the start of each episode, as humans do not react instantly to light. Kiselev, Espy, and Sheffield [31] reported that the reaction to light for adults was 270ms on average with a standard deviation of 31ms. For the "without guidance" condition, the same measures were obtained except for accuracy.

In part 3 in session 2, the average time the user took to whack each mole, the distance travelled, and the jerkiness of the movements were measured. Additionally, the number of mis-selections when the users selected other moles before selecting the intended one during each episode, was recorded. Two episodes were excluded due to a malfunction with the whack-a-mole game.

The results of time, distance, and jerkiness were compared between the robot conditions ("normal teleoperation", "with guidance", and "without guidance"), and between the eye gaze interfaces using linear mixed-effects models [32]. Linear mixed-effects models are an extension of linear regression models, but they account for differences or effects of each participant within each condition. As a linear model, if the slope between two datasets is zero then there is no statistical difference. For instance, the model for time was:

$$Time = \beta_0 + \beta_1 * condition + \beta_2 * (1|participant) + \epsilon \quad (2)$$

where condition refers to the conditions of interest for comparing, for example, "without guidance" and "with guidance". β_0 is the intercept of the linear model, β_1 is the coefficient of the slope for the fixed effects (conditions), and β_2 is the coefficient for the random effects accounted for each participant, and ϵ is the error. The hypothesis tested was $Ho : \beta_1 = 0$. If the hypothesis is rejected with a confidence level of 0.05, then there is a significant difference between the two conditions.

3. Results

Table 1 lists the means and standard deviations for time, distance, and jerkiness, for the trials with normal teleoperation (from session 2 part 1 where there was no mirroring and no guidance) and when the y and z axis were mirrored (from session 2 part 2 with the attentive interface in the "without guidance" condition). Additionally, it lists the results of the linear mixed-effects model obtained from comparing each measure between the two conditions.

Table 1. Linear mixed-effects models for comparing the results of normal and mirrored teleoperation without guidance (mean and standard deviation)

	Normal teleoperation	Mirrored teleoperation	p-value
Time (s)	1.25 ± 0.75	2.26 ± 1.72	0.000
Distance (m)	0.232 ± 0.11	0.32 ± 0.22	0.000
Jerkiness	-20.34 ± 2.39	-22.44 ± 2.75	0.000

Table 2 lists the means and standard deviations of time, distance, and jerkiness, when the participants did the activity "without guidance" and "with guidance" with the attentive interface during session 2 part 2. It also lists the results of the linear mixed-effects model for comparing the results of the two conditions. All participants responded that the activity was easier to do without the haptic guidance activated by the attentive interface. All participants also commented that they felt the haptic guidance was sometimes against their movements, making the activity somewhat more difficult.

Table 2. Linear mixed-effects models for comparing the results of "without guidance" and "with guidance" of the attentive interface (mean and standard deviation).

	Without guidance	With guidance	p-value
Time (s)	2.26 ± 1.72	2.46 ± 1.69	0.039
Distance (m)	0.32 ± 0.22	0.34 ± 0.23	0.183
Jerkiness	-22.44 ± 2.75	-22.81 ± 2.59	0.010

Table 3 lists the performance of the participants when using the attentive and explicit interfaces. It shows the results of the linear mixed-effects model for the comparisons of time, distance, and jerkiness. The neural network of the attentive interface had an accuracy of 86.43 ± 15.58 % (mean and standard deviation). The explicit interface had an accuracy of 100 ± 0 % and the number of mis-selections the participants did was 11, performed by five out of the ten participants. None of the participants felt that their eyes were tired after using the attentive interface. In contrast, six out of the 10 participants felt that their eyes were tired after using the explicit interface. Seven out of the 10 participants preferred the explicit interface over the attentive interface. Regarding the participants' comments, five of them thought that it was bothersome or tedious to fixate and transition their gaze between the change-selection-spot and the moles. Three of out those five participants felt that their eyes were tired after using the explicit interface. The three participants that preferred the attentive interface commented that the attentive interface was faster. The other seven participants preferred the explicit interface because the haptic guidance was not against their movements.

4. Discussion

Mirroring of the axes of the teleoperation did encourage users to move the robot in wrong directions. According to Table 1, participants spent significantly more time, travelled longer distances, and had jerkier movements with mirrored axes. These results imply that mirroring of the axes indeed induced confusion or involuntary movements in the participants, making them move the robot in wrong directions, increasing the time and the distance travelled to whack each mole. More relevant, the participants' movements were jerkier and could be, to some extent, a simulation of the movements performed by a person with physical impairments who experiences involuntary movements.

Table 3. Linear mixed-effects models and results for the comparison of the attentive and explicit interfaces with guidance (mean and standard deviation).

	Attentive	Explicit	p-value
Time (s)	2.46 ± 1.69	4.04 ± 1.84	0.000
Distance (m)	0.34 ± 0.23	0.29 ± 0.15	0.000
Jerkiness	-22.81 ± 2.59	-21.61 ± 2.24	0.000

According to Table 2, when the attentive interface applied the haptic guidance, participants spent more time and had more jerky movements than when haptic guidance was not applied. The main reason for the longer time was that the accuracy of the predictions made by the neural network was not 100%. Therefore, there were periods of time where the guidance was oriented towards the wrong mole, making the participants stop along the way until the neural network made the right prediction. The guidance never forced the participants to whack wrong moles, this is why the distance travelled to whack the moles was not significantly different between having the guidance on or off. A possible reason for not achieving 100% accuracy is that the movements to go from one mole to another could be somehow similar, leading to difficulties for discriminating the movement patterns for each class. For example, going from the bottom-left mole to the upper-right mole has, to some extent, a similar movement trajectory (e.g. the direction) as going to the mole in the middle.

There were many differences found between the attentive and explicit interfaces, as seen in Table 3. Regarding time, the participants spent significantly more time to whack each mole with the explicit interface than when using the attentive interface. Theoretically, the control of the explicit interface takes one second (two dwell times) longer than the attentive interface but it was about 1.5 seconds on average, because of the time the participants took to hear the mole ID spoken by the computer. The dwell-time could be decreased as participants gain more experience with the system and the explicit interface.

There was a significant difference between the distance travelled to whack each mole when using the attentive and the explicit interface. Participants travelled longer distances to whack each mole when using the attentive interface. One reason is that the force of the potential field (Fig. 2B) was higher for the explicit interface, therefore, providing more help to whack the moles. The movements to reach and whack the moles were significantly jerkier using the attentive interface than the explicit interface, because of the inaccuracy of the neural network. This was why participants commented that the haptic guidance was sometimes against their movement, and a reason why seven out of ten preferred the explicit interface over the attentive interface.

The accuracy of using the explicit interface was 100% because the participants could repeat their selection if they had selected the wrong target, but at the cost of spending more time to whack each mole. The 11 mis-selections by five of the ten participants represents less than the 2% of the total number of episodes. The low number of mis-selections implies that the adults did not experience difficulties using the explicit interface, contributing to seven out of ten preferring this interface. In addition, the interface allowed the participants to point out where they wanted to go. Thus, the haptic guidance did not oppose their movements. In the case of children, we hypothesize the number of misselections will be higher. Furthermore, as five adult participants thought it was bothersome or tedious to transition their gaze, it is possible that children may not use the explicit interface. Research is lacking on how children perform with explicit interfaces, especially children under five years old who lack some of the cognitive skills required to control robots [33].

An advantage of the attentive interface is that the users did not have to change their eye behavior, which makes the interface more intuitive and faster to user than the explicit interface. Also, user's eyes did not get as tired as with the explicit interface. None of the participants experienced tiredness in their eyes after completing the task with the attentive interface. In contrast, six participants out of ten felt their eyes were tired from having to fixate for 0.5s at the moles and the change-selection-spot. This tireness is also reported in other studies with explicit interfaces [13].

Testing with adults without impairments showed that it was feasible and safe to apply haptic guidance based on input from an attentive eye gaze interface, and the interface had advantages such as faster time and less eye-strain than the explicit interface. Also, it avoids problems with explicit interfaces for robot control identified in Encarnação et al. [23] about children having to switch their attention between the screen and the robot. However, its accuracy performance was low enough that it caused the adults to prefer the explicit interface instead. It is possible that children will be less tolerant than adults of the steps required for the explicit eye gaze interface. However, children may also be less tolerant than adults of prediction errors of the attentive interface. Improvements to increase accuracy and guidance will have to be made before trying the attentive interface system with children with physical impairments.

Different input variables would improve the prediction accuracy. The haptic guidance affected the movements performed by the user, i.e., position, velocity and direction, which were input variables of the neural network. Thus, the haptic guidance influenced both the input and output of the neural network. Therefore, different and additional input variables should be considered, e.g., the force exerted by the user on the haptic interface, as well as different machine learning algorithms such as Support Vector Machines.

Making the K constant of the potential field higher would help to make the haptic guidance more effective. The K constant of the potential field (Fig. 2B) needed to be lower for the attentive interface than the explicit interface (10 N/m compared to 50 N/m). Having a value of 50 for the K constant was helpful to whack the moles, but also prevented the users from moving towards other moles. The potential field affected the movements of the participants, and at the same time affected the position, velocity and direction of the environment-side robot, which were inputs of the neural network. Despite having the POG coordinates as inputs too, the neural network was unable to change its output unless it observed similar movement patterns as in the training. The K constant had to be decreased to a value (10N/m) so the neural network could respond appropriately, allowing the user to move towards other moles but still be supported to whack the moles. In contrast, issues were not found with the explicit interface because the participants explicitly told the system where they wanted to go. This allowed the system to have a higher K constant to provide more support for whacking the moles. Regarding the cone-shaped guidance, the K constant was the same for both eye gaze interfaces, because this type of guidance allowed the user to move with more freedom towards the moles, allowing the neural network to be responsive.

There were some limitations of this study. Only adults without physical impairments tested the system, and not children with physical impairments, who are the target population. However, the results of this study were helpful to understand improvements needed to develop an attentive user interface for children. The results of the comparisons between the two eye gaze interfaces may be biased because all participants tested the attentive interface first and then the explicit interface, with only five minutes for resting. A longer resting time would have been required to washout any learning effects. Also, counterbalancing the order in which participants tested the interfaces would have helped to control for learning effects. The eye tracker used in this study was stationary and required the user to be within its operating distance to obtain a reliable measure of the POG. In this study, operating distance was maintained by requiring the users to look through a hole, which is not a natural method of interaction. However, this eye tracker could be a good choice for children with mobility impairments who use a wheelchair. The eye tracker could be attached to the frame or lap tray of the wheelchair, which will ensure that their eyes are always within the operating distance. Wearable eye trackers could be another option to explore, since they do not have the limitation of the operating distance of the eye tracker. Nevertheless, it is possible that users would dislike wearing it.

5. Conclusion

The attentive interface offered some advantages over the explicit interface in this haptic guidance task, including lower times to complete the activity and the user's eyes did not get as tired as with the explicit interface. Using the attentive user interface could be more intuitive to use than the explicit interface, as it is intended to predict where the user wants to go without the user having to point out with his/her gaze explicitly. However, the accuracy for these predictions were not 100%, and for this reason, all the participants felt like the guidance activated by the output of the neural network was sometimes against their movement. This was one of the reasons why seven out of the ten participants preferred the explicit interface.

If prediction accuracy and the haptic guidance could be improved, attentive interfaces should be easier and more intuitive to use than explicit interfaces, which can contribute to a more natural interaction between the user, the robot, and the environment. Next stages of the project will be directed to improve the attentive interface and then test it with children with physical impairments.

Acknowledgments

This research was funded by Collaborative Health Research Projects (CHRP), which is a joint initiative of the National Sciences and Engineering Research Council (NSERC) and Canadian Institutes of Health Research (CIHR). Grant numbers: 462227-14 and 134744.

References

[1] A. C. Eliasson, L. Krumlinde-Sundholm, B. Rösblad, E. Beckung, M. Arner, A. M. Öhrvall and P. Rosenbaum, The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability, *Dev. Med. Child Neurol.* **48**(7) (2006), 549–554.

[2] J. C. Deitz and Y. Swinth, Accessing Play Through Assistive Technology, in *Play in Occupational Therapy*, (Mosby Elsevier, St. Louis, USA, 2008), pp. 395–412.

[3] G. Kronreif, B. Prazak, M. Kornfeld, A. Hochgatterer and M. Fürst, Robot assistant 'PlayROB' - User trials and results, in *16th IEEE International Workshop on Robot and Human Interactive Communication*, (Jeju, Korea, 2007), pp. 113–117.

[4] A. M. Ríos-Rincón, K. Adams, J. Magill-Evans and A. Cook, Playfulness in children with limited motor abilities when using a robot, *Phys. Occup. Ther. Pediatr.* **36**(3) (2016) 232–246.

[5] K. Adams, L. Alvarez and A. Rios, Robotic Systems for Augmentative Manipulation to Promote Cognitive Development, Play, and Education: Principles and Practice, in *Robotic Assistive Technologies. Principles and practice* (CRC press, Boca Raton, 2017), pp. 219–260.

[6] A. Miguel Cruz, A. M. Ríos Rincón, W. R. Rodríguez Dueñas, D. A. Quiroga Torres and A. F. Bohórquez-Heredia, What does the literature say about using robots on children with disabilities?, *Disabil. Rehabil. Assist. Technol.* **12**(5) (2017) 1–12.

[7] V. Hayward, O. R. Astley, M. Cruz-Hernandez, D. Grant and G. Robles-De-La-Torre, Haptic interfaces and devices, *Sens. Rev.* **24**(1) (2004) 16–29.

[8] S. Demain, C. D. Metcalf, G. V Merrett, D. Zheng and S. Cunningham, A narrative review on haptic devices: relating the physiology and psychophysical properties of the hand to devices for rehabilitation in central nervous system disorders, *Disabil. Rehabil. Assist. Technol.* **8**(3) (2013) 181–189.

[9] N. Jafari, K. Adams and M. Tavakoli, Haptics to improve task performance in people with disabilities: A review of previous studies and a guide to future research with children with disabilities, *J. Rehabil. Assist. Technol. Eng.* **3** (2016) 1–13.

[10] K. S. Choi and K. H. Lo, A hand rehabilitation system with force feedback for children with cerebral palsy: two case studies, *Disabil. Rehabil.* **33**(17–18) (2011) 1704–1714.

[11] N. Jafari, K. Adams, M. Tavakoli and S. Wiebe, Development of an assistive robotic system with virtual assistance to enhance play for children with disabilities: A preliminary study, *J. Med. Biol. Eng.*, (2017) 1–13.

[12] P. Majaranta and A. Bulling, Eye tracking and eye-based human-computer interaction, in *Advances in Physiological Computing* (Springer-verlag, London, 2014), pp. 39–65.

[13] M. Nakanishi and Y. Mitsukura, Wheelchair control system by using electrooculogram signal processing, in *The Korea-Japan Joint Workshop on Frontiers of Computer Vision*, (Incheon, South Korea, 2013), pp. 137–142.

[14] A. Banerjee, S. Datta, P. Das, A. Konar, D. N. Tibarewala, and R. Janarthanan, Electrooculogram based online control signal generation for wheelchair, in *2012 International Symposium on Electronic System Design (ISED)*, (Kolkata, India, 2012), pp. 251–255.

[15] A. Păsărică, G. G. Andruseac, I. Adochiei, C. Rotariu, H. Costin, and F. Adochiei, Remote control of an autonomous robotic platform based on eye tracking, *Adv. Electr. Comput. Eng.* **16**(4) (2016) 95–100.

[16] R. J. K. Jacob, The use of eye movements in human-computer interaction techniques: what you look at is what you get, *ACM Trans. Inf. Syst.* **9**(3) (1991) 152–169.

[17] E. Wästlund, K. Sponseller, O. Pettersson, and A. Bared, Evaluating gaze-driven power wheelchair with navigation support for persons with disabilities, *J. Rehabil. Res. Dev.* **52**(7) (2015) 815–826.

[18] G. S. Watson, Y. E. Papelis, and K. C. Hicks, Simulation-based environment for the eye-tracking control of tele-operated mobile robots, in *Modeling and Simulation of Complexity in Intelligent, Adaptive and Autonomous Systems (MSCIAAS) and Space Simulation for Planetary Space Exploration (SPACE)*, (Pasadena, California, 2016).

[19] J. P. Hansen, A. Alapetite, I. S. MacKenzie, and E. Møllenbach, The use of gaze to control drones, in *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA'14*, (Florida, USA,2014), pp. 27–34.

[20] K. Arai and K. Yajima, Robot arm utilized having meal support system based on computer input by human eyes only, *J. Hum. Comput. Interact.* **2**(1) (2011) 1–9.

[21] A. Frisoli, C. Loconsole, D. Leonardis, F. Bann, M. Barsotti, C. Chisari and M. Bergamasco, A new

gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks, *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **42**(6) (2012) 1169–1179.

[22] P. Encarnação, T. Leite, C. Nunes, M. Nunes da Ponte, K. Adams, A. Cook, A. Caiado, J. Pereira, G. Piedade and M. Ribeiro, Using assistive robots to promote inclusive education, *Disabil. Rehabil. Assist. Technol.* **12**(4) (2017) 352–372.

[23] F. Courtemanche, E. Aïmeur, A. Dufresne, M. Najjar, and F. Mpondo, Activity recognition using eye-gaze movements and traditional interactions, *Interact. Comput.* **23**(3) (2011) 202–213.

[24] S. Li, X. Zhang, F. J. Kim, R. Donalisio da Silva, D. Gustafson, and W. R. Molina, Attention-aware robotic laparoscope based on fuzzy interpretation of eye-Gaze patterns, *J. Med. Device.*, **9**(4) (2015) 1-10.

[25] D. Novak, X. Omlin, R. Leins-Hess, and R. Riener, Predicting targets of human reaching motions using different sensing technologies, *IEEE Trans. Biomed. Eng.* **60**(9) (2013) 2645–2654.

[26] P. Biswas and P. M. Langdon, Multi-modal target prediction, in *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access. Lecture Notes in Computer Science*, eds. C. Stephanidis and M. Antona (2014), pp. 313–324.

[27] J. L. Castellanos-Cruz, M. F. Gómez-Medina, M. Tavakoli, P. M. Pilarski, and K. D. Adams, Preliminary Testing of a Telerobotic Haptic System and a Neural Network to Predict Targets during a Playful Activity, in *7th International Conference on Biomedical Robotics and Biomechatronics*, (Enschede, Netherlands, 2018), pp. 1280–1285.

[28] I. Sakamaki, K. Adams, M. F. Gomez Medina, J. L. Castellanos Cruz, N. Jafari, M. Tavakoli and H. Janz, Preliminary testing by adults of a haptics-assisted robot platform designed for children with physical impairments to access play, *Assist. Technol.* (2017) 1-9.

[29] N. Hogan and D. Sternad, Sensitivity of smoothness measures to movement duration, amplitude, and arrests, *J. Mot. Behav.* **41**(6) (2009) 529–534.

[30] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, On the analysis of movement smoothness, *J. Neuroeng. Rehabil.* **12**(1) (2015) 1–11.

[31] S. Kiselev, K. A. Espy, and T. Sheffield, Age-related differences in reaction time task performance in young children, *J. Exp. Child Psychol.* **102**(2) (2009) 150–166.

[32] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware, Linear Mixed Effects Models, in *Applied Longitudinal Analysis*, (John Wiley & Sons, Inc., Hoboken, NJ, 2011).

[33] L. Poletz, P. Encarnação, K. Adams, and A. Cook, Robot skills and cognitive performance of preschool children, *Technol. Disabil.* **22**(3) (2010) 117–126.

Javier L. Castellanos Cruz received his BEng in Biomedical Engineering in 2016 from Universidad del Rosario and Escuela Colombiana de Ingeniería Julio Garavito, Bogotá, Colombia. In January 2019, he completed all the requirements for the degree of MSc in Rehabilitation Science from the University of Alberta, Canada. His research interests include the development of robotics for helping children with physical disabilities to play, machine learning for developing intuitive user interfaces and intelligent robots, the development of mobile applications for controlling robots, and eye gaze interfaces.

María Fernanda Gómez Medina received her BEng degree in Biomedical Engineering in 2016 from Universidad del Rosario and Escuela Colombiana de Ingeniería Julio Garavito in Bogota, Colombia. She has completed all the requirements for the degree of MSc in Rehabilitation Science from the University of Alberta in Edmonton, Canada. Her research focuses on areas of assistive technology for people with disabilities. She has been involved in projects related to using machine learning algorithms for the development of technology to support play in children with physical disabilities

Mahdi Tavakoli is a Professor in the Department of Electrical and Computer Engineering, University of Alberta, Canada. He received his BSc and MSc degrees in Electrical Engineering from Ferdowsi University and K.N. Toosi University, Iran, in 1996 and 1999, respectively. He received his PhD degree in Electrical and Computer Engineering from the University of Western Ontario, Canada, in 2005. In 2006, he was a post-doctoral researcher at Canadian Surgical Technologies and Advanced Robotics (CSTAR), Canada. In 2007-2008, he was an NSERC Post-Doctoral Fellow at Harvard University, USA. Dr. Tavakoli's research interests broadly involve the areas of robotics and systems control. Specifically, his research focuses on haptics and teleoperation control, medical robotics, and image-guided surgery. Dr. Tavakoli is the lead author of Haptics for Teleoperated Surgical Robotic Systems (World Sci-

entific, 2008). He is an Associate Editor for IEEE/ASME Transactions on Mechatronics, Journal of Medical Robotics Research, Control Engineering Practice, and Mechatronics.

Patrick M. Pilarski is a Canada Research Chair in Machine Intelligence for Rehabilitation at the University of Alberta, an Associate Professor in the Division of Physical Medicine and Rehabilitation, Department of Medicine, and a Fellow of the Alberta Machine Intelligence Institute (Amii). He received the B.ASc. in Electrical Engineering from the University of British Columbia in 2004, the Ph.D. in Electrical and Computer Engineering from the University of Alberta in 2009, and completed his postdoctoral training in computing science with Dr. Richard S. Sutton at the University of Alberta. Dr. Pilarski's research interests include reinforcement learning, real-time machine learning, human-machine interaction, rehabilitation technology, and assistive robotics. He is the author or co-author of more than 70 peer-reviewed articles, a Senior Member of the IEEE, and is currently supported by provincial, national, and international research grants.

Kim Adams is an Assistant Professor in the faculty of Rehabilitation Medicine at the University of Alberta, Edmonton, Canada. She received her PhD degree in Rehabilitation Science from the University of Alberta, Canada, in 2011. She is a past-RESNA (Rehabilitation Engineering and Assistive Technology Society of North America) board member and past-chair of RESNA's special interest group on internationally appropriate technology. She has 20 years of clinical experience in the field of assistive technology. She has 29 peer reviewed journal articles and 63 conference research articles. Her research interests are assistive technology design, development and evaluation, use of assistive robots for children with physical disabilities to engage in play and learning activities, and augmentative and alternative communication.