DEVELOPMENT OF AN ATTENTIVE USER INTERFACE FOR CONTROLLING A TELEROBOTIC HAPTIC SYSTEM TO SUPPORT PLAY IN CHILDREN WITH PHYSICAL DISABILITIES

by

Javier Leonardo Castellanos Cruz

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Rehabilitation Sciences

Faculty of Rehabilitation Medicine

University of Alberta

Abstract

BACKGROUND: Children with physical impairments may face difficulties when playing because of limitations in reaching and handling objects. Children with physical disabilities may have limited opportunities to play, and they may experience negative impacts on their social, emotional, or psychological development. Children are able to control robots to play, explore and manipulate the environment. Telerobotic systems could allow children to control the robot at a distance, for example, from their wheelchairs. Haptic interfaces are capable of providing the sense of touch, so that the children can feel the properties (e.g., hardness) of the objects that the robot is interacting with in the environment. Additionally, haptic interfaces can provide guidance to help children with physical impairments to reach different locations with the robots.

There two types of eye gaze interfaces for the control of robots. Explicit eye input interfaces require the user to voluntarily control the eye movements for fixating at objects for a dwell time or for doing eye gestures. Children may have difficulties controlling their eye movements and the robot at the same time. Attentive user interfaces respond according to the user's visual behavior when they interact with the technology. Attentive interfaces can potentially remove the cognitive demand of thinking about controlling the eye movements. An attentive interface could predict the toy that a child wants reach with the robot and apply haptic guidance to help him/her get to that location.

OBJECTIVES: The main objective of this thesis was to develop and test an attentive user interface for activating the guidance of a telerobotic haptic system for supporting children with physical impairments to reach toys in a playful activity. A secondary objective of this thesis was to develop an explicit eye input interface for activating the haptic guidance and compare its performance with the attentive user interface.

METHODS: The robotic system included two haptic robots, one for the user to control the movements of the other robot, which interacted with the objects in the environment. It also included an eye tracking system to track the user's eye gaze. Adults without physical disabilities, typically developing children, and a child and an adult with cerebral palsy were recruited to test the system. Studies with the adults without disabilities contributed towards developing the attentive user interface and testing the robotic system before children used it. Different algorithms for predicting the target object that the user wanted to reach with robotic system were tested and compared. Two different attentive user interfaces were developed to predict and activate the haptic guidance towards the predicted target toy. One of the interfaces predicted the target object by using a neural network (tested with adults) and the other was based on the participant's eye-robot coordination (tested by children and participants with disabilities). Also, two explicit eye input interfaces were developed for activating the haptic guidance (one tested by the adults, and the other by the children and participants with disabilities), and their performance was compared to the attentive interfaces.

RESULTS: The average accuracy of the target predictions made by the attentive interfaces was higher than 85%. Adults had 100% success rate at selecting the correct moles using the explicit interface. Adults did not feel that their eyes were tired after using the attentive interface, as they did with the explicit interface. Three of ten adults without disabilities and an adult with disabilities preferred using the attentive interface because it was faster than the explicit interface. In the case of children, they did not have 100% success rate and required prompting to use to use the explicit interface. When participants used the attentive interface, they spent less time to whack each mole than when they used the explicit interface. When children and the adult with cerebral palsy used the attentive interface, the distance travelled by the robot was less than when they used the explicit interface.

CONCLUSIONS: The eye gaze represented a significant predictor of the target mole that participants wanted to whack using the robot. Results demonstrated that the attentive user interface was faster to use than the explicit eye input interface, especially for children. Attentive user interfaces may reduce the cognitive load of having to control eye movements, thus children can focus on playing using the robot.

Preface

This thesis is an original work by Javier Leonardo Castellanos Cruz. The two research projects, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board:

"Robotic System to Support Children with Physical Disabilities in Reaching Toys", Pro00077923, March 6, 2018.

"Robotic interface with Haptic Guidance and Artificial Intelligence for People with Disabilities", Pro00049751, March 20, 2017.

Chapter 3 of this thesis has been published as:

Castellanos, J., Gomez, M. F., & Adams, K. (2017). Using machine learning based on eye gaze to predict targets: An exploratory study. Paper presented at the IEEE Symposium series on computational intelligence. Honolulu, HL, USA. https://doi.org/10.1109/SSCI.2017.8285207

Chapter 4 of this thesis has been published as:

Castellanos-Cruz, J. L., Gómez-Medina, M. F., Tavakoli, M., Pilarski, P. M., & Adams, K. D. (2018). Preliminary Testing of a Telerobotic Haptic System and a Neural Network to Predict Targets during a Playful Activity. Paper presented at the 7th International Conference on Biomedical Robotics and Biomechatronics (pp. 1280–1285). Enschede, Netherlands. https://doi.org/10.1109/BIOROB.2018.848761

Permissions from IEEE have been obtained to include the studies of chapter 3 and 4 in this thesis. Chapter 5 has been submitted for publication to the IEEE Transactions on Biomedical Engineering journal, and it is under review.

The student was responsible for the formulating the study designs, ethics application, data collection, data analysis, and writing of the manuscripts. Maria Gomez was a master's student in rehabilitation science and a research assistant for Dr. Kim Adams. Ms. Gomez was involved in the data collection and data analysis. Dr. Kim Adams was the supervisory author and was involved in the formulation of the study designs, data analysis, writing and editing the manuscripts. Dr. Patrick Pilarski and Dr. Mahdi Tavakoli were involved in the data analysis and editing of the manuscripts.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Kim Adams, for all the support, motivation, and guidance to succeed and complete the program, and for helping me grow as a researcher. I am thankful for the opportunity to be part of her research group at the Assistive Technology lab and to be able to learn from all her past and ongoing research projects. I also would like to thank Dr. Mahdi Tavakoli and Dr. Patrick Pilarski for their ideas and guidance on conducting the studies of my thesis. I also want to thank Maria Gomez for her disposition to work together and learn from each of our research projects. I would to express my gratitude and appreciation to all my teachers, the sandbox group, lab mates, and friends for making my time at the UofA memorable.

Finally, I would like to thank and acknowledge the Faculty of Rehabilitation Medicine and Collaborative Health Research Project (CHRP), a joint initiative of the National Sciences and Engineering Research Council (NSERC) and Canadian Institutes of Health Research (CIHR), for the financial support for my education and the research project.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Play is a fundamental right that must be granted and promoted for every child (United Nations, 1989). According to the International Classification of Functioning, Disability and Health in the Children and Youth version (ICF_CY), play is one of the most important activities during childhood and a key component for social participation (World Health Organization, 2007). Play is children’s primary occupation, which greatly impacts the development of sensory, motor, cognitive, communication and social skills (Tanta & Knox, 2014). Play is associated with children's pleasure, creativity, self-expression, and understanding, discovery and mastery of their world (Ferland, 2003). Play is a way for children to expand their knowledge about the self, physical and social world, and allow them to discover their capabilities by trying out objects, making decisions, understanding cause-and-effect relationships, and seeing the consequences of their actions (Garvey, 1990; Missiuna & Pollock, 1991). Play is critical for children since it is the medium to develop the skills needed to assume student, family, and social roles throughout their lives (Henry, 2008).

Children with physical impairments, such as cerebral palsy, experience limitations in reaching and handling objects for manual tasks in everyday life or for playing (Eliasson et al., 2006). As a result, children with physical disabilities have less independent, less responsive, and more compliant play behavior than typically developing children, allowing the direction of play themes to be determined by others (Okimoto, Bundy, & Hanzlik, 2000). Therefore, children who have limited experience with play may experience negative impacts on social, emotional, or psychological development (Deitz & Swinth, 2008; Missiuna & Pollock, 1991).

Giving children with disabilities the same opportunities to play as typically developing children could lead to the improvement of cognitive and social development (Robins et al., 2012). Robots can give children with disabilities opportunities to engage in play, and simultaneously promote the exploration and manipulation of the environment, enhancing learning and social participation (Cook, Encarnação, & Adams, 2010). Besides being fun, robots can provide the opportunity for children to interact with their peers and hold active roles in activities, becoming equal partners with their peers without disabilities (Marti & Iacono, 2011).

Play is exploratory, meaning activities involve movement and manipulation in relation to the environment (Stagnitti & Unsworth, 2000). Manipulative exploration of objects is as important as visual exploration, since the physical interaction can provide additional information about the objects that cannot be sensed visually, such as object properties like texture, weight, rigidity, or temperature (Fenson & Schell, 1985). Recent literature reviews of robots for the purpose of supporting play in children with physical disabilities did not report any robots that provided such physical information as feedback (Miguel Cruz, Ríos Rincón, Rodríguez Dueñas, Quiroga Torres, & Bohórquez-Heredia, 2017; van den Heuvel, Lexis, Gelderblom, Jansens, & de Witte, 2015).

Haptic robots hold the potential to provide physical information during interaction with objects, contributing to the exploration and perception of objects by providing touch feedback interpreted as force (Jafari, Adams, & Tavakoli, 2016). Haptic interfaces provide the means to interact with the environment by generating a communication channel between a machine and a human, making possible the sensing of a natural or virtual environment through touch (Hayward, Astley, Cruz-Hernandez, Grant, & Robles-De-La-Torre, 2004). Haptic feedback can contribute to the user's understanding and exploration of the environment (Demain, Metcalf, Merrett, Zheng, & Cunningham, 2013).

Telerobotic systems, which consist of a user-side robot in communication with an environment-side robot, can allow the user to sense and manipulate objects from a distance (Cui, Tosunoglu, Roberts, Moore, & Repperger, 2003), for example, user-side robot on a wheelchair and environment-side robot on the floor. Telerobotic haptic interfaces have the capability of implementing forces to amplify the input movements of individuals with muscle weakness or poor coordination, as well as filtering and restricting the user inputs of individuals with spasms or tremors (Keates, Langdon, Clarkson, & Robinson, 2000). In addition, by providing force guidance these interfaces have been able to increase human performance in manual tasks such as handwriting, reaching locations, or following pre-defined trajectories, in applications for computer access, powered wheelchairs, mobile robot control and rehabilitation of patients (e.g. post-stroke) (Jafari et al., 2016). Increased performance was defined as lower completion times or number of mistakes; however, the guidance was provided to meet the goals of structured tasks rather than reflecting the user's intentions of movement.

Guidance could be improved by providing the haptic systems with information about user's intentions. For example, the eyes, known as a window into the human brain, can provide information about intentions, emotional and mental states, as well as where our attention is focused (Ruhland et al., 2015). Eye tracking is the process of tracking eye movements to reference the location where the person's gaze is focused, also referred as point of gaze (POG). Eye tracking has been extensively used in medical and psychological research to study human behavior, as well as in human-computer interaction (HCI) (Majaranta & Bulling, 2014).

Eye tracking has been used for several assistive technology applications for individuals with physical disabilities. Computer access is one application, where controlling the mouse can be achieved using the individual's POG when looking at the computer screen (Chin, Barreto, Cremades, & Adjouadi, 2008). Individuals with communication disabilities can use augmentative and alternative communication devices such as speech generating systems with their eye gaze, gazing at letter-keys on the computer screen to spell words (Fager, Bardach, Russell, & Higginbotham, 2012). Using eye gaze as a pointer can be faster at pointing than a manual mouse, requires little conscious effort, and can be useful in hands-free tasks (Sibert & Jacob, 2000). Controlling mobile robots such as electric wheelchairs has also been possible using eye-gaze interfaces, where users command or guide the wheelchair to go in the direction they are looking (Barea, Boquete, Mazo, & López, 2002). It can be a reliable site to use when people have limited physical movements (Fejtova, Figueiredo, Novak, Stepankova, & Gomes, 2009).

Using the eye gaze as a pointer is a form of explicit eye input interfaces, which are characterized by relating eye movements to different actions of the technology (Majaranta & Bulling, 2014). This type of eye gaze interface requires the users to control their eye movements, or gaze direction, voluntarily and consciously, e.g., using the eyes to point at objects. One of the biggest challenges in using the eyes as a pointing device is the so-called “Midas Touch” problem. Midas Touch refers to the problem of distinguishing from intended and unintended eye movements for the activation of commands or actions (Jacob, 1991). Therefore, the interface must be able to distinguish between the intentions of selection and when the user is scanning, exploring, or looking around. Two common techniques to solve this problem are through the implementation of a dwell time or gaze gestures (Møllenbach, Hansen, & Lillholm, 2013). The former requires the user to stare at the target for a certain amount of time, whereas the latter uses eye movement patterns, for instance looking left to right. However, these methods may not be 100% reliable (Dybdal, Agustin, & Hansen, 2012). Additional downsides of these techniques are that users miss what is happening away from the target or the environment for a period of time, and staring at a target can be uncomfortable and tiring to the eyes (Majaranta & Bulling, 2014).

Using an explicit eye input interface to control a mobile robot can be difficult for children with physical impairments (Encarnação et al., 2017). This type of interface adds complexity to the robot control because the user must think about controlling their eye movements (fixating during a dwell time or doing eye gestures) and at the same time think of the movements commands that are required to complete a task and attend to both the interface and the robot. Robots that are controlled by explicit eye input interfaces also have the “Midas Touch” problem and users must be careful with their eye movements to avoid undesired movements of the robots.

There is another type of interface, gaze-aware or attentive user interfaces, in which the eye movements do not give explicit commands. Instead the POG is tracked and processed in the background to provide information about the user’s attentional behavior (Majaranta & Bulling, 2014). A simple example is an attentive TV, which could pause a movie when the user is not looking at the TV. Another example is in supporting readers to translate words they do not know while they read in a foreign language on a computer screen. The translation of words appears next to the text when the user gazes at a word for a predefined period, which is displayed without the explicit order of the user (Hyrskykari, Majaranta, Aaltonen, & Kari-Jouko, 2000). Visual attentive user interfaces do not intend to command the technology per se, instead they intend to analyze the visual behavior and respond according to it in a more natural and intuitive way, fostering and improving the interaction with the environment (Barbuceanu, Antonya, Duguleana, & Rusak, 2011; Li et al., 2015).

A. Objectives

The overall objective of this thesis was to develop and test an attentive user interface for activating the guidance of a telerobotic haptic system for supporting children with physical impairments to reach toys in a playful activity. The POG was analyzed in real-time to predict what toy the user wanted to reach, and it informed the haptic robot to provide force guidance towards that location. The development of this system was based on the assumption that the POG is a significant predictor of the toy/location the user wants to reach. A secondary objective of this thesis was to develop an explicit eye input interface for activating the haptic guidance of the telerobotic haptic system, and compare its performance with the attentive user interfaces when adults and children use them.

B. Papers and their contribution

This thesis consisted of four papers that were written to carry out the aforementioned research objectives. Each paper is presented as a separate chapter in this thesis.

Chapter 3. Using Machine Learning Based on Eye Gaze to Predict Targets: An Exploratory Study

This paper was presented at the 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (Castellanos, Gomez, & Adams, 2017). The study was the first step towards developing the attentive user interface. The objective of this paper was to investigate the feasibility of predicting the target object that a user wanted to reach when using a telerobotic system. Five adults without physical disabilities used the telerobotic system to do a task that consisted of knocking over three blocks. Four algorithms, two fixation-based algorithms and two learning algorithms, were implemented to predict the block the user wanted to reach. In this study, the robotic system did not provide haptic feedback nor guidance. Results revealed that it was possible to predict the target object using the user's POG and a variable indicating movement onset. The algorithm that showed the best performance in the task was a multi-layer perceptron neural network.

Chapter 4. Preliminary Testing of a Telerobotic Haptic System and Analysis of Visual Attention during a Playful Activity

This study was presented in the 2018 International Conference on Biomedical Robotics and Biomechatronics (BioRob) (Castellanos-Cruz, Gómez-Medina, Tavakoli, Pilarski, & Adams, 2018). This study built upon the results of chapter 3 and contributed towards the development of the attentive user interface. The purpose of this study was to test the performance of a multi-layer perceptron neural network (the algorithm that performed the best in chapter 3) when it was trained for a different task. Five adults without disabilities used a telerobotic haptic system to play a whack-a-mole game. The neural network was implemented to predict (offline after the task) the mole that the participants wanted to whack during the game. The neural network was trained with the coordinates of the POG and the position of the environment-side robot's end effector. Also, the performance of the neural network was measured when it was trained with windows of 0.25, 0.5, 0.75, and 1 second of input data. A spherical-shaped haptic guidance was tested by guiding the participants towards the moles in a predefined order, thus the output of the neural network did not activate the haptic guidance. Additionally, the eye gaze was analyzed to measure the percentage of time that adults devoted their visual attention on the target while whacking a mole. This was analyzed with the purpose of understanding how the participant coordinated their eye movements with the movements of the robot (i.e. eye-robot coordination). Results revealed that the neural network trained using a window of one second had the best performance. The haptic guidance did not improve the time and distance that participants moved the robot to whack the moles. Results of the eye-robot coordination showed that the participants mostly fixated at the target mole but also on the environment-side robot's end effector while controlling the robot to whack the moles.

Chapter 5. Comparison of an Attentive User Interface and an Explicit Eye Input Interface for Controlling a Telerobotic Haptic System.

This paper was submitted to the IEEE Transactions on Biomedical Engineering journal. This study was carried out with the objective of testing a neural network for predicting the target mole and activating the haptic guidance towards the predicted mole. In this study, the spherical shaped of the guidance used in chapter 4 was replaced by a cone shape. Ten adults without physical impairments were recruited for this study, and they played the whack-a-mole game with the telerobotic haptic system. A neural network was trained offline with the POG coordinates, and with the position, velocity and direction of the environment-side robot's end effector, using a window size of one second (the window size with highest accuracy in chapter 4). Additionally, the study described the implementation of an explicit eye input interface that used dwell times for selecting the target mole and activating the haptic guidance towards it. The results of using both interfaces were compared and the advantages of the attentive and explicit eye input interfaces in relation to the interaction between user, robot, and the environment were discussed. Finally, it was concluded that the neural network should not have position variables as inputs because the position was affected by the guidance.

Chapter 6. Comparison of eye gaze interfaces for controlling haptic robots that support play in children with physical impairments.

This chapter contains two studies that were performed to fulfill the two objectives of this thesis. The studies of this chapter will be submitted for publication after the paper in chapter 5 has been published, as they build upon those results. The first study was carried out with the purpose of designing an attentive interface that only had the eye gaze, and no robot position variables, as input. Nine typically developing children, a child and an adult with CP participated in the first study. Examination of the participants' eye gaze revealed that the attentive user interface could activate the guidance towards the mole that was closest to the user's POG. The algorithm of the attentive interface was simpler than the neural network implemented in chapter 5, computationally less expensive and did not require training. The objective of the second study was test the attentive interface when it activated the haptic guidance, and to compare it with an explicit interface. The explicit interface implemented in chapter 5 was modified to reduce the number of steps required to operate it. The participants were five typically developing children, and a child and an adult with CP. This chapter presents the accuracy of the predictions of the attentive interface, and the success rate of the participants at selecting the correct moles with the explicit interface. Finally, it compares and discusses the advantages of both interfaces, and highlights that the attentive user interface was easier to use than the explicit interface, and it could have a lower cognitive demand for children.

Chapter 2
Literature review

This section presents the literature of robots that have been developed for helping children with physical disabilities to play. Next, the applications of haptic interfaces are introduced and how they can help people with disabilities to control robots. Last, the literature of eye gaze interfaces for controlling robots is introduced and discussed, highlighting the gaps to which this thesis contributes.

A. Robots to support play

There are three major types of technology developed to promote play in children with physical disabilities: robots, virtual reality, and computer systems (van den Heuvel et al., 2015). However, robots have the advantage of being able to move in the children's physical environment, while the other technologies are stationary. This section focuses on studies that conducted trials with robots to support play in children with physical disabilities, fostering the manipulation and interaction with toys or objects. Robots will be described in terms of the user interfaces and the tasks that children were able to do with the robot. Studies that used play-like activities as a means to accomplish therapeutic or educational goals will not be discussed, nor those for mobility purposes.

Robotic arms have been used in an activity of hiding and seeking toys in a tub of macaroni. Pre-programmed movements of the robotic arm were executed by children with physical disabilities by pressing switches (Cook, Howery, Gu, & Meng, 2000; Cook, Bentz, Harbottle, Lynch, & Miller, 2005). Children had three buttons, one for commanding the robot to move to the location where the toy was hidden, another to make the robot go into the macaroni, and the third to make the robot move out of the macaroni. Children did the activity by pressing the switches in the correct sequence: move, go into, and move out of the macaroni. Children enjoyed playing with the robot, which gave them the opportunity to play with certain degree of independence. They were able to do the activity, however, they often pressed the switches in the wrong sequence due to possible involuntary movements or lack of understanding.

The IROMEC (Interactive Robotic Social Mediator as Companions) is a mobile robot intended to support play in children with severe motor impairments, mild cognitive disabilities, or autism. The robot can be configured for ten different play scenarios which can be done in solitary play or collaborative play, such as: “dance with me”, “build a tower”, “bring me the ball”, and others (Robins et al., 2012). The IROMEC can operate autonomously or be remotely controlled, as it has built-in switches, a touch screen and a movement detection interface. The robot can move around the environment using ultrasound and infrared sensors to recognize and distinguish objects/obstacles, and people’s bodies (Marti & Iacono, 2011). Levels of playfulness increased when children with physical impairments played with the robot. However, the robot may be challenging for the children to control due to their physical limitations, therefore, further hardware and software development were recommended to meet the needs of children with disabilities (Klein, Gelderblom, de Witte, & Vanstipelen, 2011).

Building with Lego bricks was possible for children with various impairments by using a 3-Degree of Freedom (DOF) robot, PlayROB (Kronreif, Prazak, Kornfeld, Hochgatterer, & Fürst, 2007; Prazak, Kronreif, Hochgatterer, & Fürst, 2004). Children with physical and/or cognitive disabilities controlled the movement of the robot using a joystick, sip-puff, or a 5-switch set to go left, right, forward, backwards, and up/down. In addition, single-switch scanning mode was possible; i.e. first the robot moves across the area in one direction, shifts direction when the child presses the switch, and stops to place a brick when the child presses the switch a second time. Children enjoyed playing with the robot; however, children had difficulties understanding its operation. Automation of robot movements were suggested as future work to address this issue.

Lego robots have been widely used to promote play because of their low cost and versatility to be assembled into different forms. Lego robots can be built as car-like vehicles or robotic arms, and their movement remotely controlled by switches, for example, to drive forward or open the robotic arm's gripper (Cook, Adams, Volden, Harbottle, & Harbottle, 2011; Schulmeister, Wiberg, Adams, Harbottle, & Cook, 2006). In addition, Lego robots can be adapted to carry tools such as a pen, giving children the opportunity to draw. Also, a sequence of movements (e.g. dancing) can be pre-programmed so that they can be played by pressing a single switch.

When using Lego robots controlled by switches the levels of playfulness in four children with physical disabilities increased (Ríos-Rincón, Adams, Magill-Evans, & Cook, 2016). However, when typically developing children used a similar robot to play, their pretend play levels were generally lower than when they used their hands to play with the toys (Adams et al., 2017). Controlling the robot using switches can be cognitively demanding for typically developing children under 5 years old (Poletz, Encarnação, Adams, & Cook, 2010). Therefore, authors suggested the robot provide assistance to allow children to focus on playing rather than on how to control the robot, for instance using computer vision to grasp toys that the robot approaches (Adams et al., 2017).

Children with physical disabilities could access play when using the aforementioned robots, however, some of them had difficulties to control them, because of the complexity of the interfaces or due to the children's physical or cognitive impairments. Additionally, when children played with the robots, they only had their eyes and ears to know what the robot was doing to their toys. Haptic feedback is a method to help control robots and have a sense of touch in the interaction between the robot and the objects or toys in the environment.

B. Haptic interfaces

A review of the literature of haptic systems for people with disabilities did not reveal any robots for supporting play in children with physical disabilities (Jafari, Adams, & Tavakoli, 2016). However, haptic interfaces have demonstrated great potential for enhancing manipulation and interaction with objects, which could be of great benefit for the purpose of play. Atashzar et al. (2017) designed a haptic system with two haptic robots in teleoperation mode, one operated by the user and the other interacting with the environment by following the movements performed by the user. The system was tested in a sorting task by a user with cerebral palsy. The system was capable of filtering unintended movements and scaling up the user's volitional movements. Scaling refers to amplifying the user movements by a numerical factor, e.g. if the user moves one centimeters, the environment-side robot could move two centimeters. Results demonstrated that the user had better control of the robotic system, by having smoother and more accurate movements to complete the task.

Force feedback guidance can improve functional capabilities in children with physical disabilities. One form of guidance is applying force as haptic feedback to help users reach locations or complete tasks within a given trajectory or space. Choi and Lo (2011) studied force feedback as guidance to improve handwriting in children with cerebral palsy. Children could write Chinese characters on a computer screen using a pen-like robot to follow the templates given as guidelines. If the child's handwriting was off the template, force feedback was provided to pull the child's hand towards and along the trajectory of the character. The further the child's hand drifted from the trajectory guideline, the stronger the force feedback applied. Using this approach, writing time decreased through repetitive practice, suggesting an improvement in handwriting accuracy and functional capabilities.

A similar approach has been used to help children with mobility impairments steer powered wheelchairs. Children without disabilities and a child with cerebral palsy participated in a study that tested the feasibility of a system using a force feedback joystick to steer a powered wheelchair along a pre-defined line-path (Marchal-Crespo, Furumasu, & Reinkensmeyer, 2010). Children controlled the speed and direction of the wheelchair; however, forces were applied if the wheelchair went off the path, thus helping the child follow the course.

Avoiding obstacles is another skill that can be supported in children with mobility disabilities. A force feedback joystick can apply artificial potential fields, i.e. applying forces in the direction of the given path, moving the wheelchair away from the obstacles in order to avoid them (Chen, Ragonesi, Galloway, & Agrawal, 2011). Kang, Logan, Galloway and Agrawal (2014) used the same approach in a chasing game. Typically developing children were asked to use a wheelchair to chase a caregiver in an environment with obstacles. The wheelchair tracked the caregiver's position and detected obstacles while moving. The joystick applied force in the direction of the caregiver, and away from obstacles. The use of force feedback significantly improved the steering ability of children with physical disabilities, more than without the force guidance (Chen et al., 2011; Marchal-Crespo et al., 2010).

Haptic interfaces can also generate another form of haptic guidance called forbidden region virtual fixtures (FRVF), which are accomplished by the generation of forces to keep the user's hand inside or outside a pre-defined space. For instance, FRVF were created in the form of figure shapes such as circles or squares to support coloring (Jafari, Adams, Tavakoli, & Wiebe, 2017). FRVF served as guides for the users to reduce the amount of area colored outside the given templates. Another study explored the feasibility of FRVF to guide users from one location to another. Ten people without disabilities and one person with CP did a sorting task (Sakamaki et al., 2017). The FRVF were created in straight-cylindrical shape from the pick-up location to the one of two target locations where tokens of two different colors and shapes were sorted. The performance measures of time, area covered, and sorting accuracy improved significantly for all participants when the FRVF were activated than when this feature was off. Results suggested that FRVF could be beneficial for people with tremors or spastic movements, however, considerations about the natural trajectory movement of the arm must be considered.

Natural trajectory of movements that a person does when going from one location to another can be learnt by a learning algorithm. Gaussian mixture models (GMM) and Gaussian mixture regression (GMR) can learn and apply haptic guidance around the learnt trajectory (Najafi, Sharifi, Adams, & Tavakoli, 2017). After demonstration of the trajectory by the user or a third-party person, e.g. a clinician, force guidance can be implemented along the learnt trajectory. An assistance-as-need strategy can be implemented so that forces are applied only when the user falls out the demonstrated trajectory.

To summarize, all the aforementioned methods needed prior information about the task or the user. The guidance applied forces on the user's hand to keep it inside a pre-defined region, to reach fixed locations in the environment or follow pre-defined trajectories. In the case of play, where toys are moved around to different locations constantly, pre-defined assistance is not enough to support play in children with physical disabilities. The guidance should allow them to explore their abilities thus they should be able to go beyond what has been pre-defined as a suitable path for them to follow. If haptic systems had additional input variables (e.g. eye gaze) to better understand where users want to go in the environment, they could provide an assistance that matches the user's choice and intentions. Eye gaze is one input that can be used to determine the user's intentions.

C. Eye tracking interfaces

The three most common techniques to track eye movement are: videooculography (VOG), video-based infrared pupil-corneal reflection (IR-PCR), and electrooculography (EOG). VOG refers to the recording of the eye movements using a camera or a head-mounted device to track the pupil centre. VOG had problems of inaccuracy and sensitivity to head movements, but these were addressed by the IR-PCR technique. IRPCR consists of beaming an artificial infrared (IR) light source into the eyes, which produces a glint or "corneal reflection" that leads to easier image processing and more accurate measure of the point of gaze (POG). EOG measures the electric potential in the muscles around the eyeball. Although EOG is very inaccurate to measure the POG, it provides useful information about displacement of the eye, i.e. left, right, up and down (Majaranta & Bulling, 2014). These techniques have been utilized in eye gaze interfaces that track the user's eye movements for controlling robots. These eye-tracking techniques have been utilized to control robots using two types of interfaces: explicit eye input-interfaces and attentive interfaces. IR-PCR has been more common and widely used for robot control.

I. Explicit eye input interfaces for controlling robots

Explicit eye input interfaces are a type of eye gaze interfaces that consider the eye as a direct input, meaning eye movements correspond to an action e.g. moving a computer pointer. Robots can be teleoperated using the user's eye gaze or eye movement. A Lego robot was controlled by using the user's eye gaze as a pointer to select commands on a computer screen (Păsărică et al., 2016). A camera was mounted on the robot, and the live video was sent to the user's computer screen. The computer screen displayed nine buttons that made the robot move forward, left, right, backwards, and stop. Eye fixation over the buttons for a certain period of time was required for the robot to move, in this case for two seconds.

Hansen et al. (2014) designed an eye gaze-based interface to pilot a drone from a computer. The x-y gaze with respect to the screen and two arrow keys were used to control the speed, altitude, rotations and drafting of the drone. Users were able to fly the drone using such inputs, however, the major drawback of this interface was that users were unable to look away from the screen without flying in the wrong or unintended direction. Similarly, the user's eye gaze was also used to teleoperate a mobile robot and its camera (Watson, Papelis, & Hicks, 2016). Users preferred to control the robot's motion using a joystick and use their eye gaze to rotate the camera, rather than using only their eyes to control both. The reason was because they wanted to scan the scene without inducing any movement to the robot.

A wheelchair was controlled using the direction of the user's eye gaze, e.g. if the user looked upwards the wheelchair moved forward (Ben Taher, Ben Amor, & Jallouli, 2015). Although this control method did not require a computer screen, the user had to look away from the route when making an eye gesture. Other studies have displayed a camera view from the wheelchair on a computer screen, which also displayed arrow-buttons that commanded the direction of the wheelchair. Fixation over the buttons made the wheelchair move in the respective direction while the user could still see the environment on the computer screen (Barea et al., 2002; Wästlund, Sponseller, Pettersson, & Bared, 2015).

The need for computer screens can been avoided by using EOG-based interfaces. A wheelchair was operated based on the voltages generated by the eye muscles when a person looks at the left, right, or straight (Udhaya Kumar & Vinod, 2015). Another interfaced detected eye gestures using EOG, e.g. looking straight, right and back to straight to make the wheelchair turn right (Banerjee et al., 2012; Wijesoma et al., 2005). This kind of interface was able to filter out unintended moves that did not meet the pattern criteria, however, it required the users to look away from their path. Blinking has been used to avoid having the user look away from their zone of interest, increasing safety, e.g., double blinking or closing the eyes to stop (Nakanishi & Mitsukura, 2013; Wijesoma et al., 2005).

People can use their eyes to point at objects they want a robot to reach or manipulate. Arai and Yajima (2011) developed a system to support feeding in people with disabilities and the elderly. The system was composed of a head mounted display that live streamed the robot's view and recorded user's POG. The robot had a camera mounted on the tip of the grip. Users fixated at the food item that they wanted, then the robot fed the user the selected food. Although this study did not describe how to solve the "Midas touch" problem (to avoid the selection of unintended food items), it highlighted that controlling the robot based on the user's eye gaze was faster than controlling it with computer keys. Similarly, Frisoli et al. (2012) developed an upper limb exoskeleton to help people with upper limb disabilities reach objects that they pointed with their eyes. The exoskeleton was interfaced with an eye tracker and brain computer interface (BCI). The eye tracker allowed the users to select objects in the environment by fixating at them. After selection of a target object with the eye gaze, the BCI detected the user's intentions of moving or maintaining the exoskeleton at rest. The exoskeleton reached towards the target object while the user thought of moving the arm.

The downside of the explicit eye input-interfaces is related to the “Midas Touch” problem mentioned in the introduction (i.e. distinguishing between intended and unintended eye movements). The use of fixations or gestures has attempted to solve this problem but has not been 100% reliable (Dybdal et al., 2012). These approaches require the users to change their visual behavior (voluntarily fixating at a location or doing eye gestures), and thus, miss for a certain period of time what is happening around them, making the interaction with the robot and the environment slower. Similarly, with the EOG-based interfaces, closing, winking or doing eye gestures makes the users take attention away from their environment. Additionally, complexity may increase due to the cognitive demand required to voluntarily change the visual behavior and control the movements of the robot to complete a task.

II. Explicit eye input interfaces for children

There is very little literature describing the performance of children when they use eye gaze interfaces, making it difficult to determine at what age children should be able to successfully use eye gaze interfaces (Karlsson, Allsop, Dee-Price, & Wallen, 2017). Borgestig, Sandqvist, Parsons, Falkmer, & Hemmingsson (2016) investigated the performance of children with physical impairments to gaze at objects on a computer screen. That study included 10 children between one and 15 years of age. Nine of them had cerebral palsy and the other child had cervical spinal cord injury. The study was longitudinal, and assessed the children at baseline, after 5 months, 9-11 months, and after 15-20 months. The task was to gaze at a single object displayed on a computer screen for longer than 1 second. If the child did not do it within 30 seconds the object disappeared, and the trial was counted as a failure. Children improved in terms of completion time after 5 months, and the success rate increased after 15-20 months. Children as young as 9 months of age were capable of doing the task with about 80% accuracy, and improved to 100% after 11 months

(Hemmingsson, Ahlsten, Wandin, Rytterström, & Borgestig, 2018). However, the performance on that task cannot generalize to more complex tasks that include multiple objects such as spelling.

The performance of selecting target images with the eye gaze was compared between five typically developing children with ages between four and 13 years old, and five children with spastic quadriplegic cerebral palsy between seven and 11 years old (Amantis et al., 2011). The task consisted of selecting the correct target image out of 2, 4, 8, or 16 images that were displayed on a computer screen. The typically developing children spent an average of 14.2 seconds to complete the task, while the children with CP spent 57.8 seconds. The children with CP had difficulties maintaining their gaze on the target due to their body movements. Similarly, the ability to select targets was assessed in seven children with Rett syndrome (a developmental disorder with cognitive and neuromotor impairments) between four and nine years of age (Baptista, Mercadante, Macedo, & Schwartzman, 2006). Children were asked to fixate at one of two physical pictures according to the verbal instructions, or at the picture that was similar to or was exactly like a third picture presented. Analysis of the eye gaze revealed that children fixated at the correct picture, on average, 62.4% of the time.

Some children with physical impairments may face challenges using eye tracking systems when systems are stationary. A 5 year-old child with athetoid CP used an explicit interface for selecting letters on a computer screen with the purpose of writing (Dhas, Samuel, & Manigandan, 2014). However, he had asymmetric tonic neck reflexes that caused involuntary neck movements and affected his ability to fixate at items on the computer screen. Similarly, two students with athetoid CP with the ages of 13 and 15 years old were not able to use eye tracking technology for selecting target objects on a computer screen (Man & Wong, 2007). They constantly moved out the workspace of the stationary eye tracking system.

Children may have difficulties using explicit interfaces in tasks that require multiple steps. A seven year-old child without impairments was not able to use an explicit eye input interface for drawing on a computer (Hornof, Cavender, & Hoselton, 2004). The interface required users to fixate at a location on the canvas for a dwell time of 500ms to set the starting point for drawing. Then, the user had to select a shape (line or circle) by fixating at one of the buttons displayed on the computer screen. Finally, the user had to fixate at a location on the canvas to set the end point of the figure. The other participants, who were between 10 and 36 years of age, were able to use the interface successfully. However, the results of this study cannot be generalized to the whole population or to all the tasks that can be done using eye gaze.

Children may have difficulties to use explicit eye input interfaces for controlling robots. Children with physical impairments used an explicit eye input interface to control a Lego robot for educational purposes (Encarnação et al., 2017). Three children with cerebral palsy, two of whom were three years old and the other was six years old participated in that study. 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 to fixate on the screen to make a selection and then look at the robot to observe its action. The three-year olds were not able to complete the activities, likely due to the complexity of changing the focus of their attention from the screen to the robot.

III. Attentive user interfaces for controlling robots and computer systems

Attentive user interfaces respond according to the user's visual behavior when interacting with the technology without having to fixate during a dwell time or eye gestures. The literature of attentive user interfaces for controlling robots and computer systems is scarce. Li et al. (2015) implemented a visual attention recognition method to control a laparoscope (i.e. a camera inside the body) during a minimally invasive surgery. The system had an attentive user interface that recognized the surgeon's visual attention by interpreting eye gaze movements and eye gaze patterns, then autonomously moved the robotic laparoscope to the site where the surgeon's visual attention was. The attentive interface allowed the surgeon to concentrate on the visual target while reducing the cognitive load of manually controlling the laparoscope. The interface responded faster than an explicit eye input interface that steered the laparoscope to the location where the surgeon fixated with a 2 s dwell time. Additionally, participants preferred the attentive interface over the explicit interface, saying it was simple, easy and intuitive.

Barbuceanu, Antonya, Duguleana, and Ruzak (2011) designed an attentive interface to identify the user's intentions for object selection in a virtual environment. The user's eye gaze was first analyzed during the selection of virtual objects representing items from a kitchen. The interface incorporated a model of the gaze transitions between the objects, reflecting the possible operations to perform with them, e.g., pour water from a bottle into a glass. The interface made the selections of the objects based on the transition probabilities of the model. Establishing such connections between the objects allowed the system to anticipate to the user's selection of the objects. The attentive interface was compared to an explicit eye input interface that allow the user to make the selections of targets by fixating at the objects for a dwell time (not specified). The attentive interface had an accuracy of 88% while the explicit interface was 100%. Despite the lower accuracy, participants found the attentive interface to be more intuitive than the explicit interface.

Eye gaze can be used to predict the target objects that a person wants to reach or select when interacting with computer applications. It was possible to predict the target object that adults wanted to reach on a computer screen when they were controlling the cursor with a glove (Novak, Omlin, Leins-Hess, & Riener, 2013). The left and right movements moved the cursor horizontally, and the movements towards and away from the screen moved the cursor vertically. In a two and three object task, the predictions of the targets had an accuracy of more than 90% after the user started moving the hand. The predictions were made based on the patterns of the user's POG and/or the position of the hand (i.e. glove). Similarly, it was possible to achieve an average accuracy of around 70% for predicting the target object that users wanted to reach on the computer screen when using a mouse (Biswas & Langdon, 2014). The prediction was performed by training a neural network with the eye gaze data of adults. The limitation of these two studies was that the algorithms were trained and tested offline and were not tested in a real-time task.

Attentive interfaces have the potential for removing the cognitive load that explicit eye interfaces impose because of having to voluntarily control eye movements. Attentive interfaces intend to interpret the user's behavior while he/she is interacting with the technology in a given task. Attentive interfaces can be more intuitive to use than explicit interfaces and can lead to a more natural and faster interaction between the user and the technology while performing a task.

D. Overall summary and gaps

Children with physical disabilities were able to play by controlling robots such as robotic arms and Lego robots. Robot can be controlled with different kinds of interfaces such as switches, joysticks, and others. However, children may have difficulties to control robots due to the complexity and cognitive demands of the interfaces. The robotic systems for play allowed children to see and hear what the robot was doing to the objects, but not feel.

Haptic interfaces can provide the sense of touch and can also provide guidance to people with physical disabilities for carrying out tasks such drawing and sorting with a greater performance. However, robotic systems that provided haptic guidance needed prior information about the task or the user movements, and often the haptic guidance was implemented to meet the goals of the task and not the movement intentions of the user. Thus, it is necessary to include variables that inform the guidance about the user's intentions.

Eye gaze can be used to get information about the intentions of the user and direct the haptic guidance. Explicit eye input interfaces for controlling robots had challenges related to the "Midas Touch" problem. Solutions to the "Midas Touch" problem (i.e. dwell times or eye gestures) slowed down the interaction with the robot and the environment and added complexity to the control of the robot. In the case of children, the explicit interfaces may be difficult to use due to the cognitive demand of voluntarily controlling the eye movements. There are no studies that report the performance or success rate of using explicit eye input interfaces in tasks that include multiple objects. Additionally, it is not clear at what age children can use effectively explicit interfaces for controlling technology such as computers or robots.

There were no studies about children using attentive user interfaces. However, some studies compared the use of explicit and attentive interfaces with adults. Results revealed that attentive interfaces can be more intuitive and faster to use than explicit interfaces but may be less accurate. Attentive interfaces may reduce or take off the cognitive load of having to voluntarily control the eye movements with explicit interfaces, which may be difficult for children. Attentive interfaces could predict the target object that a user wants to reach instead of having the user to explicitly select with his/her eyes. It has been possible to predict the target object that a person wants to reach in computer applications, but it has not been done while a person interacts with a robot in a physical environment. Also, the prediction of a target object has not been utilized for supporting people with disabilities to carry out tasks using a robot. For this reason, this thesis aimed at developing a telerobotic haptic system that used an attentive user interface for predicting the target object and activating the haptic guidance towards the predicted object. An attentive interface of this form could allow children to effectively control robots and support play in children with physical disabilities.

Chapter 3
Using Machine Learning Based on Eye Gaze to Predict Targets: An Exploratory Study

Previous studies revealed that it was possible to predict using eye gaze the target object that users wanted to reach on a computer screen when they controlled a cursor (Biswas & Langdon, 2014; Novak et al., 2013). However, no studies were found about the prediction of target objects when a person uses a robot.

The objective of this paper was to investigate the feasibility of predicting the target object that a user wants to reach with a robot in a three-block task. This prediction was based on the Point of Gaze (POG) data of five participants while performing the task using a telerobotic haptic system. Two fixation-based algorithms, longest fixation and last fixation, and two learning algorithms, a Double Q-learning and a Multi-Layer Perceptron neural network, were implemented, tested, and compared.

I. Methods

A. System Description

In this study the overall system consisted of two subsystems. The main subsystem included a low cost Tobii Steelseries Sentry eye tracking system (Tobii Technology, Stockholm, Sweden), and a LogiTech webcam (Logitech International S.A., Romanel-sur-Morges, Switzerland). The Tobii Steelseries Sentry is a binocular eye tracker intended for gaming, with the same hardware specifications as the Tobii EyeX. It provides estimation in real time of the Point of Gaze (POG), with an operating distance of 540-800 mm, and has a nominal sampling rate of 55 Hz when controlled with the Matlab toolkit (Gibaldi, Vanegas, Bex, & Maiello, 2016). Gibaldi et al. (2016) stated that this gaming eye tracker can be used in research applications in which fixation parameters, saccadic and vergence eye movements, and smooth pursuit are needed. A Windows PC and the 2016b version of Matlab (MathWorks Inc., Nadick, MA,US) software were used to acquire the signals from the eye tracker and the webcam. A simple code was written in Matlab to record the video and the POG signal throughout the trials at a sample frequency rate of 17Hz. The sample frequency was lower than the nominal due to the integration of the webcam and the eye tracker.

The second subsystem included two low cost Novint Falcon Haptic robots (Novint Technologies Inc., Rockville, USA). One was placed in the environment and the other in the user side (Master side), for participants to control. The robots were working in a unilateral teleoperation mode, in which the robot in the environment followed the movements of the Master robot using a PID controller. Another Windows PC was used for this system with the 2011b Matlab/Simulink version and Quarc V2.2 (Quanser Inc., Markham, ON, Canada).

B. Experimental Design

Five adults without physical or visual disabilities, three females and two males, participated in the study. Their age range was from 21 to 42 years (M=28.6, SD=8.53). For each participant, a 9 point-calibration of the eye tracker was performed using the Tobii Eye Tracking Core Software v2.9.0 (Tobii Technology, Sweden). *Figure 3-1* shows the experimental setup for the present study. In this study, 18 wooden blocks were used to create the words SUN, TOE and PIE, as well as the numbers 183, 954 and 726. For each trial a set of 3 blocks were positioned on top of a box.

A piece of cardboard was placed in front of the participants before each trial started to occlude their view, and once the instruction to knock down the blocks was given, the cardboard was removed.

Participants were instructed to knock down the six sets of blocks six times each a) in order to spell a word or a number, b) according to their favorite color, number or letter, c) in order from smallest to largest numbers, and d) in the order of the alphabet from A to Z. In addition, participants were asked to move back to the center position of the robot every time after knocking down a block. For example, in the set of blocks shown in *Figure 3-1*, the participant was asked to knock down the blocks in order to spell the word TOE, then he/she knocked down first the block with the letter T and went back to the center position of the robot, then knocked down the block with the letter O and went back to the center position, and knocked down the block with the letter E and went back to the center position.

Figure 3-1 Experimental setup for trials

The eye tracker was placed in front of the participants, and the *x* and *y* coordinates of the POG signal were recorded during the sessions. It was not necessary to calculate the depth of the eye gaze because the blocks were placed parallel to the eye tracker and in fixed positions. The webcam view covered the environment, including the environment-side robot and the blocks. The videos were analyzed to register the time-step when the trials started, the moment at which the end-effector started moving, and the block the participants knocked down (target). For each trial three episodes, one per block, were recorded. The cardboard removal indicated the time at which the trial started, and the time at which the last block was knocked down indicated the time at which the trial ended. A total of five datasets, one per participant, were collected and put together to create a final dataset. Episodes that had eye gaze data missing were excluded.

Figure 3-2 shows the signal for the x-coordinate of the POG of participant 1 when knocking down two blocks (i.e. two episodes). The scanning interval (i.e. the time interval the participant took to decide which block to knock down) for the first episode was labeled from the time in which the cardboard was removed until the time the robot started to move. The moving interval for this same episode was labeled from the time the robot moved until the block was knocked down. For the second and third episodes in each trial, the scanning interval started from the time the previous block was knocked down until the time the robot started to move from the center position of the robot towards the target. The moving interval was labeled from the time the robot moved until the block was knocked down. Additionally, a binary variable that indicated if the participant was moving or not was created. This variable (movement onset variable) took values of zero when the participant was in the scanning interval and one when the participant was in the movement interval.

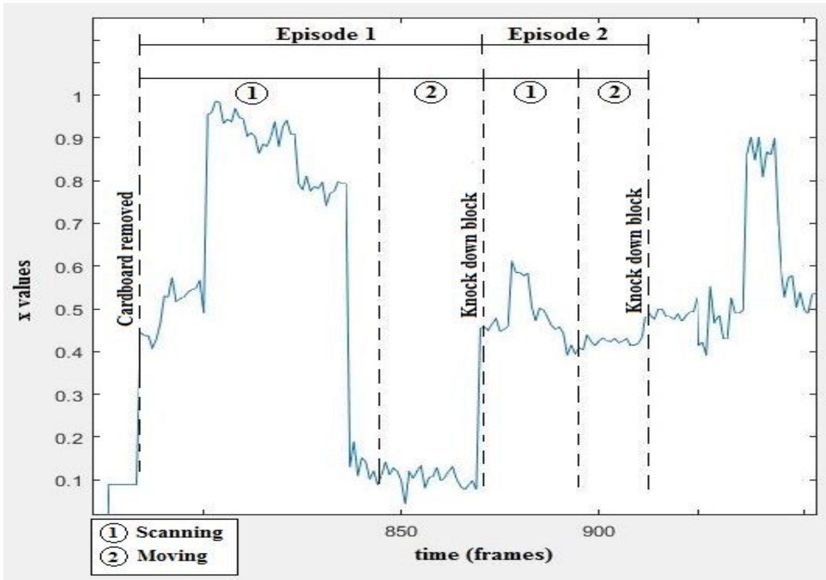

Figure 3-2 POG signal for participant 1 for two episodes

C. Algorithms implementation

*Figure 3-3***¡Error! No se encuentra el origen de la referencia.** illustrates the end goal of this project, in which the algorithm will take as inputs the information from the user's eye gaze, features about the movement of the user-side robot such as position or velocity, and the input from a camera to know the location of the toys (to be able to map the POG to those locations). The final algorithm must be able to identify the toys that a child wants to reach, according to the inputs, and provide the suitable haptic guidance when the user-side robot starts moving.

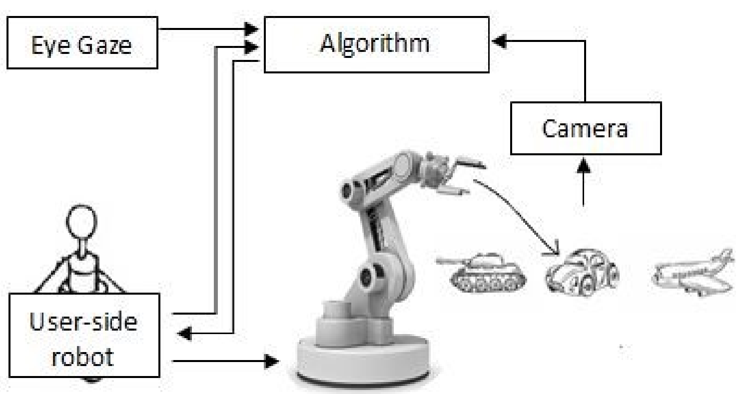

Figure 3-3 End goal for the principal project illustration.

This paper shows the first results towards the end goal project. In this study four algorithms were implemented to predict which block the user intended to knock down before they started moving. It is worth mentioning that an attempt was made to predict the onset of movement based only on the eye gaze data, however, accuracy was close to zero.

Initial results when training the algorithms using only the POG data to predict the target and also the time at which the robot should move, showed an accuracy of about 3%. With that in mind it was decided that the movement onset variable needed to be included as an input variable.

1. Fixation-based Algorithms

For the fixation-based algorithms, unintended eye movements such as blinks and eye fixations out of the task area (e.g. when participants looked at the researcher) were filtered out. The POG signal was measured and coded as 1, 2, or 3 according to the pre-defined range locations of each one of the three target blocks. Data points that were out of the range of the blocks were coded as 0. Additionally, fixations on the blocks shorter than 150 ms (three samples) were also filtered out. This number was based on previous studies where 100 ms was the minimum amount of time to declare a fixation (Majaranta & Bulling, 2014).

Two algorithms were written: longest fixation-based and last fixation-based algorithms. These algorithms used eye fixations to predict which block the user was going to knock down once the movement started. The longest fixation-based algorithm had the complete scanning intervals as input. This algorithm made predictions based on the block that had the most and the longest fixations before the movement started. The last fixation-based algorithm made predictions based on the blocks participants fixated at four samples before and four samples after the movement started. The accuracy in each of these samples was calculated to test the performance of the last fixation-based algorithm.

2. Learning Algorithms

The two learning algorithms used to predict which block the user wanted to knock down (target) were: Double Q-Learning and Multi-Layer Perceptron (MLP) Artificial Neural Network. These algorithms were intended to anticipate the user's intention of reaching. For training these algorithms, the final database (with all the participants data) consisted of 514 episodes (5 participants X 6 sets X 6 instructions X 3 blocks, minus 26 episodes excluded). The database was divided into training and test datasets, 90% and 10% respectively. The training dataset with 463 episodes was used to do the off-line training of the Double Q-learning and the MLP algorithms.

a) Double Q-learning algorithm

Double Q-learning is a reinforcement learning technique used to solve decision making problems, by finding an optimal control policy through repetitive trial and error interactions with the environment, in this case the play area and the robot (van Hasselt, 2010). The policy, a rule for the selection of an action on a given state, is learnt by taking actions that maximize the rewards (positive or negative feedback) generated by the environment. This algorithm, described in Algorithm 1, converges to the optimal policy in any finite Markov Decision Processes by employing updates to the estimated values Q_1 and Q_2 iteratively after each action a taken in a given state s.

The learning rate α was 0.1, the discount factor γ chosen was 1, and the exploration was ϵ-greedy with ϵ=0.05, and having these values constant. Lines 2 through 12 were repeated 10000 times, meaning that the algorithm ran about 4.6 million episodes.

In this study the possible actions a can be translated as which block is the user going to knock down, thus three possible actions: block 1, 2 or 3. In the context of the end-goal of the project, the action would refer to the target (toy) the robot should provide haptic guidance towards in order to help the children with physical disabilities reach it. The possible states *s* included the codification for the POG used in the fixation-based algorithm (four possible values: 0, 1, 2 and 3) and the movement onset variable (two possible values: 0 or 1). Therefore, the states were coded from one to eight using a tile coding approach. As each episode ended with one block being knocked over, the rewards were +1 throughout the episode when the algorithm predicted successfully the block that was going to be knocked down, and -1 otherwise.

Algorithm 1 Double Q-learning

Initialize: $Q_1(s,a)$ and $Q_2(s,a)$ to small (10e-3) random
numbers
Repeat (for each episode):
Initialize s
Repeat (for each state-step of the espisode):
Choose a from ϵ-greedy in $Q_1(s,a) + Q_2(s,a)$
Take action a, and observe R and s'
with 0.5 probability:
$Q_1(s,a) \leftarrow Q_1(s,a)+\alpha(R+\gamma\, Q_2(s',\mathrm{argmax}_a\, Q_1(s',a))- Q_1(s,a))$
Otherwise:
$Q_2(s,a) \leftarrow Q_2(s,a)+\alpha(R+\gamma\, Q_1(s',\mathrm{argmax}_a\, Q_2(s',a))- Q_2(s,a))$
s$\leftarrow$s'
until s is the terminal state

b) Multi-Layer Perceptron

A Multi-Layer perceptron (MLP) is an artificial neural network (ANN) that maps inputs in the form of sets to the appropriate outputs (Haykin, 2008). This neural network model is a popular method for real problem applications such as pattern classification, and speech and object recognition. In this study an MLP-ANN was trained to predict the block the participants wanted to knock down (target), based on the POG coordinates.

Being that the POG is a continuous signal, windows of approximately 0.25, 0.5, 0.75 and 1 seconds were used to process the signal. These values were chosen to explore the performance of the neural network when it had different lengths of input data. This means that for training the

MLP, the input consisted of the POG data for the actual time (t) and POG data for past times (t-windows size). Due to the sample rate of 17Hz the windows of 0.25, 0.5, 0.75 and 1 second were composed of 5, 9 13 and 17 samples, respectively. The targets for each episode were also included in the dataset as the desired output.

Figure 3-4 shows the structure of the MLP used for this study, where the input data were the windows with the POG values of the participants. The variables that were used to train the MLP were the POG x and y coordinates, and the movement onset variable. The size of the input layer of the network changed according to the window size. For example for the window of 0.25 s (w=5 samples), the size of the input was N=(5*2)+1=11, where five values were the x coordinate of the POG data, another five values were the y coordinate POG data variables, and the last value was the movement onset variable. The targets were the output signal of the neural network, where B1, B2 and B3 refer to block 1, 2 and 3 respectively. The MLP was trained using the Levenberg-Marquardt algorithm. Results with the four different windows sizes were compared to find out the best window size to train the algorithm.

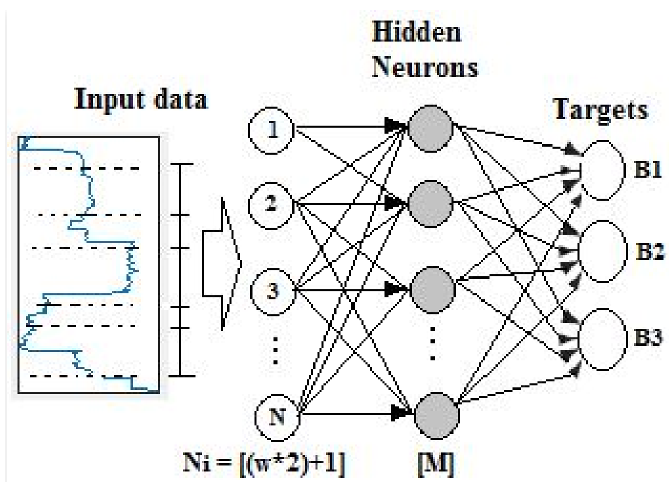

Figure 3-4 Structure of the Multi-layer perceptron with the POG data as input and the three blocks as target outputs.

c) Testing the algorithms

To calculate the accuracy of the learning algorithms, 51 (test dataset) episodes were used. For the Double Q-learning algorithm, the learned policy was applied to the test dataset with no further learning or exploration (i.e. $\alpha = 0.1$ and ϵ=0.05). For the MLP algorithm the neural network was applied to the testing set with no further learning.

The learning algorithms were given the POG coordinates and movement variable in a sample by sample basis. The accuracy of each algorithm at predicting the correct target at the time the movement started *t* was obtained. Additionally, the accuracy at predicting the target four samples before *t* and four samples after *t* was included.

II. Results

A characterization of the POG signal according to the performed task with the blocks was carried out. In this characterization it was found that 7.12 was the average number of saccades the participants made during one episode, and 5.61 and 2.29 were the average numbers of saccades participants made during the scanning and movement intervals, respectively.

Figure 3-5 shows the plot of the accuracy results for the last fixation-based, the Double Q-learning and the MLP algorithms. The longest fixation-based algorithm had 52% accuracy at predicting the target block. This means that participants focused their attention for a longer time on the block they wanted to knock down in only 52% of the scanning interval. As this algorithm achieved a low accuracy at predicting the target blocks, it was not compared to the other three algorithms due to its poor performance.

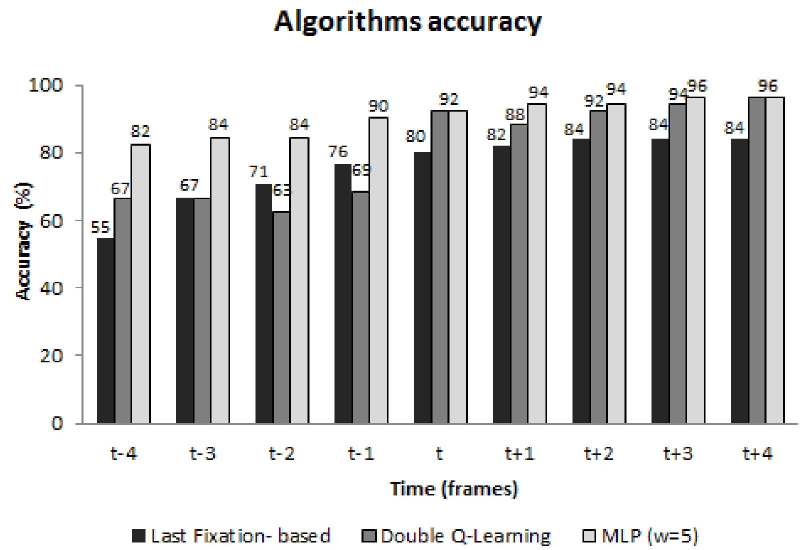

Figure 3-5 Algorithms' accuracy from four samples before and after the movement onset.

The average number of fluctuations during the scanning and the movement intervals were calculated for the implemented algorithms, and are shown in Table 3-1. *Figure 3-6* shows the predicted output obtained with the Double Q-learning and the MLP algorithms for two different episodes. During the scanning interval (1) of the first episode, it is possible to see that the target prediction of the MLP algorithm had seven fluctuations (changes in the target prediction), and the target prediction of the Double Q-learning had just three fluctuations.

The selection of the best window size to train the MLP algorithm was based on the average number of fluctuations during the scanning and movement intervals, and on the accuracy that each algorithm obtained when predicting the target at the time the movement interval initiated. Accuracy for the different window sample sizes went from 90% to 94%, with the 9-sample window having the highest accuracy. However, the average fluctuation for the 5-sample window was smaller than that for the 9-sample window, and the accuracy was 92%. Thus, it was decided to proceed with the 5-sample window.

Table 3-1 Results of the Average Number of Fluctuations in Predictions for each Algorithm

Intervals	**Average of Oscillations**		
	Last fixation-based		***Last fixation-based***
Scanning and Movement (Full Episodes)	6.92	Scanning and Movement (Full Episodes)	6.92
Scanning	5.80	Scanning	5.80
Movement	1.96	Movement	1.96

III. Discussion

Overall results indicated that the longest fixation-based algorithm made 52% accurate target predictions. Despite the low accuracy for predicting targets, it was still better than random, so this longest fixation variable could have a correlation with the selection of targets that is worth investigating. Perhaps be combined with other algorithms to improve their prediction performance.

The results of the last fixation-based algorithm were coherent with the literature of hand-eye coordination about target selection, which has identified that humans typically make a saccade to fixate on the target they have selected, slightly before or after the hand begins to move (Issen & Knill, 2012). *Figure 3-5* shows that fixations at the target before the onset of the end-effector movement (< t) was lower than after the movement occurred (> t), and increased as users moved towards the selected target. These results suggest that visual behavior for target selection during human-robot interaction (HRI) with a haptic system is similar to behavior when doing the task with the hands. This is because the end-effector of the environment-side robot acts as the human hand, following the same movements the user performs on the user-side robot.

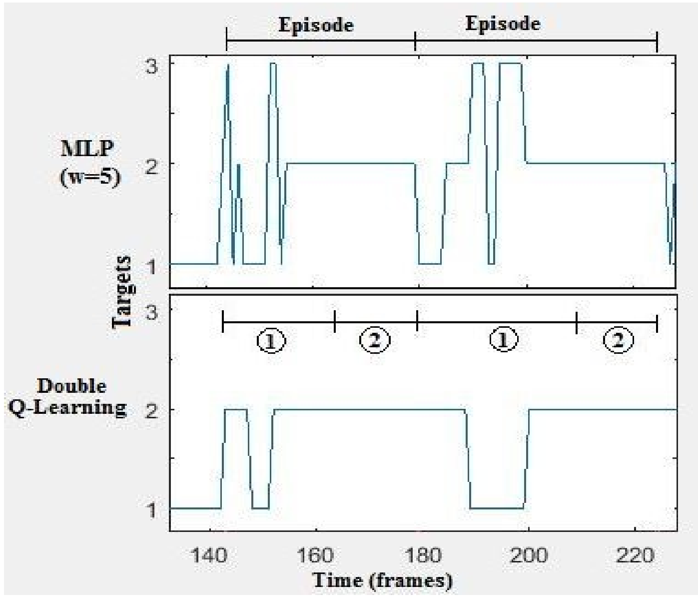

Figure 3-6 Output signal for the Multi-Layer Perceptron and the Double Q-learning algorithms during two episodes.

Figure 3-5 shows that the learning algorithms, Double Q-learning and MLP, usually had a better accuracy than the last fixation-based algorithm at predicting the target before and after the movement onset of the robot's end effector. However, the prediction performance for the MLP with the 5-sample windows outperformed the other two algorithms across all time-steps. It shows that the performance tended to increase as time progressed after the movement, likely due to the fact that the users were getting closer to the target. The MLP and the Double Q-learning results were consistent with a study that used a neural network with eye tracking and brain computer interfaces to predict targets, where the users selected the targets on a computer screen using the eyes as direct input method (Biswas & Langdon, 2014).

The learning algorithms showed a high accuracy for predicting the target once the user began to move. The accuracy was above 92%, except for the Double Q-learning at one sample (t+1) after the movement onset. The predictions 4 samples before participants began to move had an accuracy of above 63%. However, accuracy in general did not reflect the stability of the algorithms to predict the target from the beginning to the end of the episode, for that fluctuations of the output needs to be examined.

Table 3-1 shows that the last fixation-based algorithm was unable to sustain a stable prediction throughout the episode, because this was based only on the last fixation. In contrast, the learning algorithms had lower numbers of fluctuations during the scanning and movement intervals, meaning it was a more stable prediction. Double Q-learning had a more stable prediction response during the scanning interval than the MLP, there were only 2.94 compared to 5.37 fluctuations on average respectively. The MLP had the most stable response during the movement interval, with only 1.31 fluctuations on average. How the output fluctuated for the learning algorithms was illustrated in *Figure 3-6*. If haptic guidance was to be applied towards the target selected by the algorithms, haptic forces would be changing between targets, based on fluctuations, which can cause instability of the robot and discomfort for the users. One way to avoid this would be to only apply haptic guidance once the users start moving, because it is more stable. The ideal algorithm would have zero fluctuations from the moment it makes a prediction to the moment when the user reaches the target. However, the MLP with a 5-sample window appears to be the best option for the context of this task.

The task can be considered as semi-structured since there was a restriction of movement for the users (e.g. moving the robot back to the center position), and some of the questions only had one correct answered (e.g. spelling words), whereas others did not (e.g. favorite colors). In the context of unstructured play where children are constantly moving their toys to different locations and involving more than three toys at a time, scanning interval and movement onset may not be as simple to label. The possible scenario in which a child wants to move the robot to a location where there are no toys should also be taken into consideration. In addition to these considerations, future work will involve the integration of more input variables to increase the performance of the algorithm, for instance position of the robot with respect to the toys, velocity, and direction of the movement.

IV. Conclusions

The present study was the first step of a project in which the use of eye tracking information was proposed as an input to a haptic telerobotic system, with the purpose of assisting children with physical disabilities to reach the toys they want to play with. Haptic robots can exert forces to guide children to locations in a play environment. By analyzing visual behavior, it could be possible to predict which toys (targets) children want to reach, in order to provide a suitable guiding assistance for them.

This study analyzed the eye gaze of adults in a target selection task using a haptic telerobotic system. A fixation based approach and two machine learning algorithms were used to predict the targets they wanted to reach. Double Q-learning and a Multi-layer Perceptron (MLP)were compared to the last fixation-based approach, which was based on the last fixation made by the user before moving the robot. Results demonstrated that learning algorithms can predict the target the users wanted to reach with higher accuracy and robustness than the last fixation-based algorithm. Above 92% accuracy was obtained once the users started to move and above 63% four samples (235ms) before the users moved towards the target. Overall the MLP had a higher accuracy than the other algorithms.

Average numbers of fluctuations in the algorithm’s target prediction were calculated as a measure of stability. Analysis revealed that both learning algorithms did not have a stable prediction response after users started to move, with number of fluctuation above 1.3 on average.

However, both algorithms performed better than the last fixation-based algorithm. Fluctuations after the users started to move can be related to the fact that while users move towards the target, they make saccades to fixate on other objects possibly to think about their next target, or to make sure they do not knock down the other targets.

The results also suggest that these learning algorithms can potentially be used for the purpose of supporting play in children with physical disabilities to predict the guidance they need in order to reach a toy. However, more input variables will be required for more complex play as well as more targets since the task presented here only had three possible targets and the object locations were fixed. It is worth adding that these algorithms could be used for any application of robot control that involves the selection of targets.

Chapter 4
Preliminary Testing of a Telerobotic Haptic System and Analysis of Visual Attention during a Playful Activity

In this study, a telerobotic haptic system was developed with two haptic robots, one that is for a user and the other to interact with the environment. The goal of this study was to do preliminary tests of a haptic guidance method and the prediction of targets. Another goal was to explore and analyze the visual attention of the participants during the activity when eye-hand discoordination was induced. Five adults without disabilities played a whack-a-mole game using the robotic system, to assure that the robot works adequately before children with disabilities use it. The robots were programmed to induce eye-hand discoordination, so that haptic guidance would be required. A multi-layer perceptron neural network (the algorithm that performed the best in chapter 3) was implemented to predict the target moles that the participants had to reach, which in future versions, will control the activation of forbidden region virtual fixtures (FRVF) to guide the user towards the target moles.

I. Methods

A. System Description

The overall system consisted of two subsystems. The main subsystem included two PHANToM Premium 1.5A haptic robots (3D Systems, Inc., Rock Hill, SC, USA). One was placed in the environment (slave side) and the other in the user side (master side), for participants to control. The robots were programmed in bilateral teleoperation mode using a PID controller for position control. Hence, the two robots followed each other’s position (e.g., if the user-side robot moved 2 cm in the x-axis, then the environment-robot did the same). This allowed the user to have haptic feedback, i.e. feeling forces implemented at the user-side robot when the environment-side robot touched an object. The robots were programmed in R2016a Matlab/Simulink and using the Quarc V2.2 library (Quanser Inc., Markham, ON, Canada) on a Windows PC. This library provides the necessary Simulink blocks for accessing external robotic devices such as the PHANToM Premium.

The second subsystem included a low-cost Tobii EyeX eye tracking system (Tobii Technology, Stockholm, Sweden), and a LogiTech webcam (Logitech International S.A., Romanel-sur-Morges, Switzerland). The Tobii EyeX is a binocular eye tracker intended for gaming. It provides estimation in real time of the point of gaze (POG), with an operating distance of 540-800 mm, and has a nominal sampling rate of 55 Hz when controlled with the Matlab toolkit (Gibaldi et al., 2016). Gibaldi et al. (2016) stated that this gaming eye tracker can be used in research applications in which fixation parameters, saccadic and vergence eye movements, and smooth pursuit are needed. Another Windows PC and the R2016b version of Matlab software were used to acquire the signals from the eye tracker and the webcam. A simple code was written in Matlab to record the user's eye gaze and the video of the trials at a sampling rate of 16z. The sampling frequency was lower than the nominal due to the integration of the webcam and the eye tracker. Additionally, at the start of the trials, the program sent a signal to the robots' PC, which facilitated the synchronization of the data collected from both subsystems. Communication between the two computers was done using user datagram protocol (UDP).

For the activity, a Whack-A-Mole Arcade Game by Fisher-Price was utilized. The game consisted of five moles with LEDs and switch-buttons to measure when the player whacked the mole. The game was adapted with a microcontroller, an Arduino Leonardo, to light up the LEDs and measure when the moles were pressed. The microcontroller also communicated in real time with the robot's PC using serial communication. For synchronization purposes, the robot's PC sent to the microcontroller which mole to light up. The target moles were randomly generated but predefined before the session. Also, the microcontroller sent to the robot's PC in real time the moles that were pressed.

B. Haptic Guidance

The haptic guidance was based on FRVF, which were implemented to help the users reach the target moles. The FRVF had a spherical shape that kept the user inside the space. The radius of the sphere decreased as the user got closer to the target mole, preventing the user from getting away from the target. The radius was determined as the minimum distance between the position of the environment-side robot and the target mole, analyzed in a two-second window, plus 0.5cm. The 0.5cm was added so that the user did not feel like he/she was up against the FRVF all the time. Once the target mole changed, the initial radius was re-calculated with the current distance between the end-effector and the target mole. The center of the sphere was the target. *Figure 4-1* illustrates the guidance method. This design is based on the findings in a previous study (Sakamaki et al., 2017). In that study, a person with cerebral palsy tested the FRVF in a sorting task. The shape was a cylinder that went from the pick-up location to the drop-off location. Results indicated that the motion of the person with physical disabilities was not a straight line, instead the person did arc-like movements. Therefore, the FRVF opposed the natural and preferred motion. The proposed spherical FRVF in this study would allow the person to move with more freedom, except for not letting him/her move away from the target.

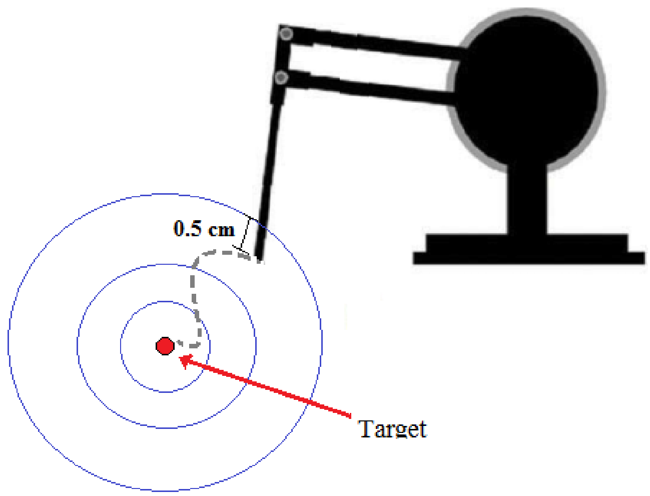

Figure 4-1 Illustration of the FRVF as guidance method. As the user gets closer to the target, the radius of the sphere-shaped FRVF decreases, preventing the user from moving away from the target.

Additionally, four FRVF were implemented as spheres with 4cm of radius, which surrounded the moles that were not lit up. In this case, the FRVF were used to keep the user outside that space and prevent him/her from whacking the wrong moles.

C. Procedure

Five adults without physical or visual disabilities, four females and one male, participated in the study. One of them wore glasses. Ethical approval was obtained from the Health Research Ethics Board – Health Panel at the University of Alberta.

The playful activity was carried out with three different robot conditions: typical teleoperation, inverted teleoperation, and inverted teleoperation with guidance. Typical teleoperation refers to the robots following each other's positions. "Inverted" teleoperation was done with the purpose of inducing eye-hand discoordination in the participants, so that they made involuntary movements. Inverted teleoperation refers to inverting the x and y-axis of the teleoperation and mirroring the z-axis. The environment-side robot moved in the x-axis according to the user-side robot's position in the y-axis, and in the y-axis according to the user-side robot's position in the x-axis. Also, when the user-side robot went upwards (z-axis) the environment-side robot moved downwards. Finally, inverted teleoperation with guidance refers to having the robots with inverted-axis teleoperation, but with the haptic guidance.

The order of the starting conditions was randomized. Three participants started with inverted teleoperation without guidance, and two started with guidance. This was intended to control for the learning effect. All participants played last using the robots with typical teleoperation, which served as control for comparing the results of the other two conditions.

Participants were asked to sit in front of the play area and rest their chin on a mounting arm, this was to record reliable data for the analysis of the visual attention and avoid recalibration due to head movements. In each robot condition, the participants played the whack-a-mole game until whacking a total of 60 moles. The moles were lit up one at a time, randomly, and without being repeated consecutively. After the mole was whacked, the next one was lit up. Participants were instructed to keep focus on the play area the entire time. *Figure 4-2* depicts the setup of the activity and the robotic system.

D. Data Collection and analysis

From each participant, the x, y, and z position of the environment-side robot, the eye gaze data, the moles that were whacked, and a video of the play area were recorded at 16 Hz. The data was processed to compose a dataset of 900 episodes. An episode consisted from the time the target mole was lit up until it was whacked. The final dataset was composed of 888 episodes, after excluding the episodes where the eye gaze was not detected (12).

During each condition, the time and trajectory length that the participants took to do the activity was measured. Additionally, the percentage of the episode period that the participants' eye gaze was on the target mole was measured. To compare the three conditions, the average of the measures was utilized. Also, the learning curve of each variable was plotted to visualize the improvements of the participants as the activity continued.

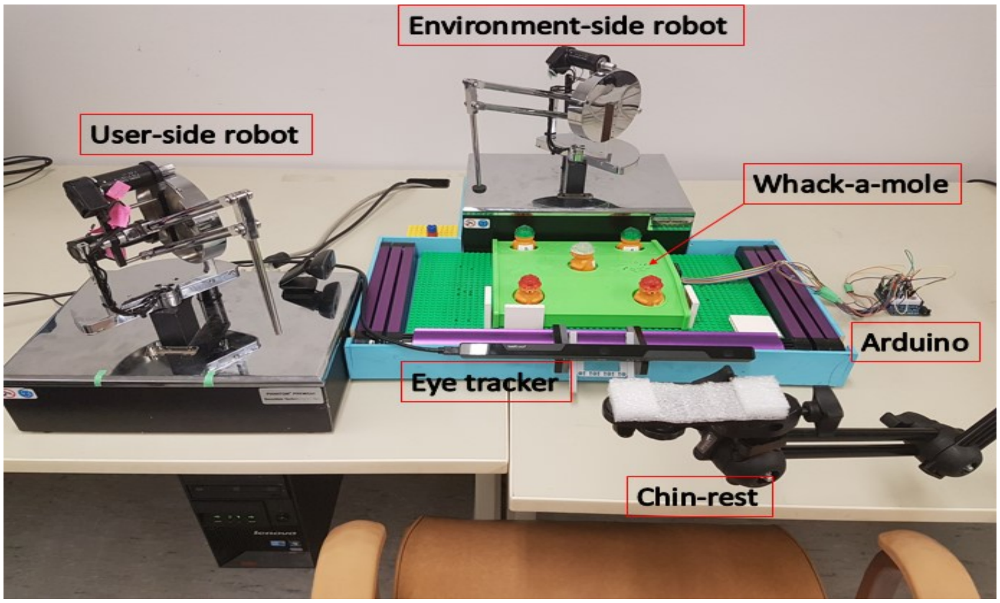

Figure 4-2 Setup for the whack-a-mole activity game.

E. Neural network: Target prediction

The MLP was used to predict the target (mole) the user was supposed to go to. The MLP was implemented as a classifier. The input variables were seven: the x, y and z coordinates of the environment-side robot, and the x and y coordinates of the POG of the left and right eyes of the user. Windows of 4 (0.25s), 8 (0.5s), 12 (0.75s), and 16 (1s) samples were created to train the four neural networks. The datasets were created using the actual time (t), and inputs for past times (t – windows size). The size of the input layer of the network changed according to the windows size: N=window size*input variables. The output labels were the target mole that was lit up for each episode. The structures of the MLPs (N-M-O) were 28-20-5, 56-56-5, 84-70-5, and 112-80-5, for the window sizes of 4, 8, 12, and 16 samples, respectively. *Figure 4-3* shows the structure of the

MLP neural network that was used in this study. Gradient descent was applied with a learning rate of 0.01 for the training.

To train the MLPs, four datasets of 77854, 75725, 71583, and 70099 training examples were composed to train the MLP of 4, 8, 12, and 16 samples, respectively. The datasets contained the data from all three conditions. The purpose was to train the MLP with a high variety of movements. The MLPs were trained offline using 80% and tested with the remaining 20% of the respective datasets. The accuracy was measured in four episode-intervals, only in the testing sets. Interval 1 had the data of only the first 50% of each episode. Interval 2 had the data from the 25% until the 75% of each episode. Interval 3 had the data from the 50% until the 100% of each episode. Interval 4 was the entire episode. To explore the stability of predictions, the average number of fluctuations in the prediction signal was measured for the episodes in the testing set. Results with the four different window sizes were compared, to find out the best window size that would lead to better predictions of the target mole the users were supposed to whack.

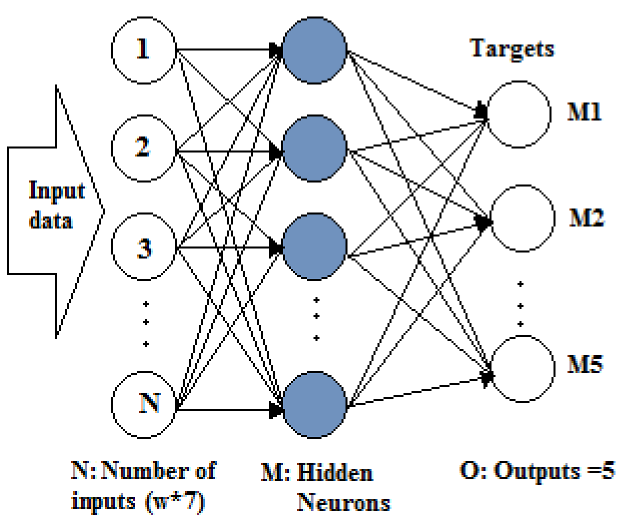

Figure 4-3 Structure of the fully connected multi-layer perceptron neural network used for target prediction. The input data was the x,y,z coordinates of the environment-side robot end-effector, and the x,y coordinates of each eye. Window sizes (w) of 4, 8, 12 or 16 samples of the input data were tested. The output layer had five output neurons, one for each mole.

II. Results

All participants played the game successfully. They had 100% success in whacking the correct moles with the three robot conditions. Table 4-1 presents the average time spent by the participants to complete the activity, and, the average trajectory length traveled by the participants. Finally, Table 4-1 lists the average percentage of visual attention that participants devoted to the target mole.

Participants had great difficulties to control the environment-side robot with inverted teleoperation, having the *x* and *y* axis inverted and the *z* axis mirrored. The most challenging was understanding that if the user-side robot moved to the right, the environment-side robot would move towards the front. For this reason, they spent more than three times the time spent with typical teleoperation and travelled more than twice the distance.

Table 4-1 Average performance in the whack-a-mole game with the three robot conditions

	Time (min)	**Trajectory Length (m)**	**Visual attention on the target (%)**
Typical	2.34±0.46	15.68±2.09	61.01±22.72
Inverted with Guidance	9.07±5.43	40.65±21.56	43.22±22.89
Inverted with no Guidance	8.84±3.36	36.67±9.12	44.10±21.56

The variance of the results is because three participants started with inverted teleoperation and without guidance, while two participants started with guidance. The three participants that started without haptic guidance had lower times and trajectory lengths in the second trial when they played the game with guidance. In contrast, the participants that started with haptic guidance had better results in the second trial, without haptic guidance. In terms of the percentage of visual attention on the target, this measure was similar for trials with and without the haptic guidance.

Figure 4-4 depicts the learning curve in terms of time spent in whacking each mole for the participant with the scores closest to the average. This participant started with inverted teleoperation with haptic guidance. A similar behavior was observed in terms of trajectory length. However, a trend was not observed in terms of percentage of visual attention on the target mole, therefore it was not drawn.

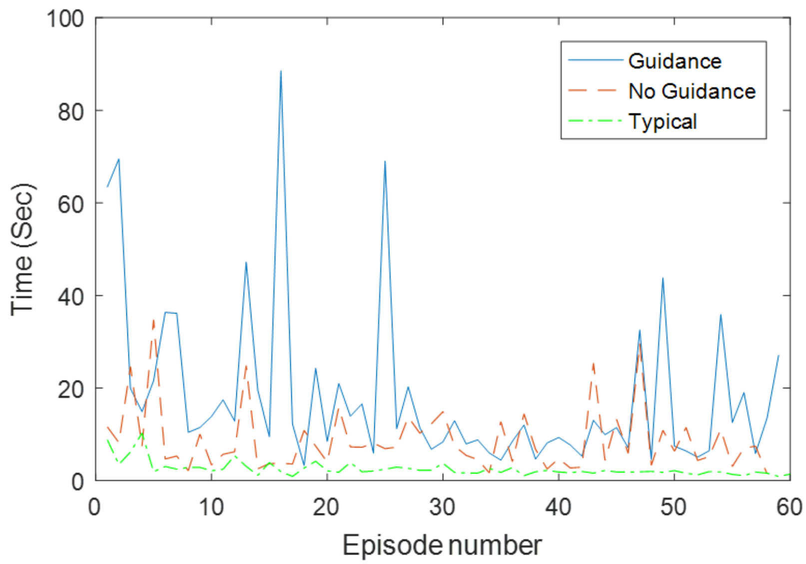

Figure 4-4 Learning curve of the participant with scores closest to the average.

Table 4-2 lists the accuracy of the MLPs predictions with the different window sizes. The accuracy is reported for each of the intervals of the episode, and for the complete episode. The average number of fluctuations (changes in the target prediction) per episode were 7.91, 5.42, 4.06, and 3.48, for the 4, 8, 12 and 16 sample windows, respectively.

Table 4-2 accuracy of the target predictions for each window size of the MLP neural networks

Window size	**Accuracy (%)**			
	Interval 1	***Interval 2***	***Interval 3***	***Entire episode***
w=4	46.3	60.3	66.4	56.2
w=8	45.7	66.5	78.4	62.6
w=12	50.7	78.6	89.3	70.6
w=16	48.9	74.8	**90.2**	70.7

III. Discussion

As expected, the participants performed better with typical teleoperation compared to the other two conditions. The inversion served its purpose to confuse the able-bodied users and thus make demands on the guidance aspects of the robotic system. We had expected that the inverted results would be in favor of when haptic guidance was applied. The participants who started with guidance took longer than when they did the activity without guidance. A possible reason why the participants did not perform better with the haptic guidance is because the FRVF were designed only to restrict the user from moving away from the target and not actively guide (push) him/her towards it. It was observed that participants often followed the sphere walls without getting closer to the target.

In terms of visual attention, when participants had the inverted teleoperation, they fixated at the target mole around 44% of the duration of the episodes. The rest of the time, the users were mainly looking at the effector and the possible obstacles (the moles that were not lit up). By inverting and mirroring the axis, the participants' eye-hand coordination was diminished, thus making their movements less reliable. Therefore, they had to rely more on visual feedback from the environment-side robot's end-effector, compared to typical teleoperation. We had expected that the haptic guidance would help the users gain more control and improve their coordination. As the guidance method did not improve the user's performance in the activity, we will test this hypothesis in a future study using a different form of haptic guidance.

When people without disabilities move their hand to reach an object, their visual attention is on the target and never on the hand (Johansson, Westling, Bäckström, & Flanagan, 2001). Studies about upper-limb artificial prosthesis users have found that during reaching and grasping tasks, skilled users fixate at the target object a higher percentage of the movement period compared to inexperienced users (Bouwsema, Kyberd, Hill, van der Sluis, & Bongers, 2012). The lack of full proprioception feedback (i.e., awareness of the movement and position of the body parts) from the prosthesis requires the user to rely on visual feedback. In this study, participants fixated at the target mole less during the inverted teleoperation than in typical teleoperation, as they could not rely on the proprioception feedback provided by the haptic user-side robot. In a sense, the less control a person has over the telerobotic system, the less he/she will look at the target mole. More research is required to understand the visual behavior in children with physical disabilities when they use robots.

It was observed that users improved by spending less time and travelling less distance as the activity progressed (e.g., *Figure 4-4* for one participant's time). This behavior was not observed for the visual attention data. However, if participants had more time to use the robot, we would expect their visual attention on the target to increase. This will reflect their skill level at controlling the robot, as happens with prosthetics users (Bouwsema et al., 2012).

Regarding the prediction of targets, from Table 4-2 it is possible to observe that the performance increases with the increase of the window size. Accuracy increases, and the average number of fluctuations decreases with the window size. This is because with a bigger window size the input for the MLP increases, thus providing more information leading to more accurate predictions. Additionally, higher accuracy was achieved when the MLP was tested using the last part of the episodes (interval 3). This means that the MLP had a better accuracy when participants were approaching the target. The lowest accuracy was obtained in the first half of the episodes (interval 1), On average, the performance of the MLP of 16 samples was lower than in our previous study, where the MLP achieved 94% (Castellanos et al., 2017). Possible reasons are that in the whack-a-mole game there were more targets, the participants were not asked to move back to a starting position, and eye-hand discoordination was induced.

The MLP did not provide stable predictions throughout the episodes. With a window of 16 samples the MLP had the fewest fluctuations (3.48). However, if haptic guidance was applied towards the targets predicted by the MLP, haptic forces would be applied rapidly in different directions based on fluctuations towards different targets making the user uncomfortable or causing the teleoperation to be unstable. Using FRVF guidance has less instability issues in the teleoperation system than if we applied attraction forces towards the target. In case that the prediction of the MLP fluctuates to wrong targets, the FRVF would not allow the user to go to their desired target but it would not pull them towards the wrong target, which has more potential of causing instability in the teleoperation system.

In this study, the data of the five participants in the three robot conditions were grouped together. In other words, we tested the one-size-fits-all approach. Perhaps the MLP could perform better if it was trained specifically for each person, or for each robot condition. However, this would require a bigger dataset from each person, therefore, the person would need to play more using the robotic system. Also, the MLP would need to be trained differently if the toys are not static and if the users can move their heads freely.

For future work we will improve the haptic guidance by implementing the FRVF in a cone shape or potential force fields, so that better performance can be achieved. Also, the system will be evaluated with the haptic guidance being directed by the output of the neural network. We will explore different options to improve the performance of the MLP neural network for predicting targets, including training a neural network for each person and for each condition, and including more input variables. Also, the best strategy will be identified to apply the haptic guidance in case the MLP does not reach 100% accuracy. In the upcoming stage of the project, we will be testing the system with adults and children without disabilities. We can gain insight about the possible demands of the system and take them into account for subsequent trials by children with physical disabilities.

IV. Conclusions

Telerobotic systems could help children with physical disabilities to play if they have difficulties reaching their toys. The user-side robot could be mounted on the child's wheelchair, for example, thus increasing the opportunities to interact with their world.

In this study, five adults without disabilities tried a telerobotic haptic system to play a whack-a-mole game. Inverting the axis of the teleoperation system induced eye-hand discoordination in the participants, which allowed us to test the FRVF guidance. The haptic guidance that was implemented in this study did not improve the performance of the participants. The FRVF prevented the participants from getting further away but did not help them to get closer. The guidance needs to be more active so that it can guide the users towards the targets easier and faster.

In terms of visual attention, participants had their eye gaze on the target longer in typical teleoperation than in inverted teleoperation. The proprioception feedback from the haptic interface was confusing, hence the users had to rely more on visual feedback while watching the environment-side robot's end effector and other moles. We hypothesize that the less control a person has over the teleoperation system, the less they will look at the target. However, these results are exploratory and cannot be generalized, more research required.

The user's eye gaze and the environment-side robot's position can be used for the prediction of target toys that the user wants to reach. We implemented and compared four MLP neural networks with inputs of window sizes of 0.25, 0.5, 0.75, and 1s. The MLP with a window size of 1s performed the best in the whack-a-mole game. The accuracy for prediction of targets increased as the robot end effector got closer to the target mole. On average, the MLP had a 70.7% accuracy throughout the testing episodes. This accuracy may cause a robotic system to be unstable if haptic guidance was applied with the network's output. Also, as the guidance affects the movements of the users, this may affect the MLP's performance. For future work, before testing the MLP and the guidance together, we will improve the accuracy by including more input variables such as velocity and direction of the robot during the tasks.

Chapter 5
Comparison of an Attentive User Interface and an Explicit Eye Input Interface for Controlling a Telerobotic Haptic System

This article describes the development of a telerobotic haptic system that uses eye gaze for the activation of haptic guidance. Two eye gaze interfaces were tested. First, an attentive user interface that predicts the target toy that users wanted to reach was tested. The interface used a neural network for recognizing the patterns of the user's hand movements performed on the user-side robot and the user's point of gaze (POG). The neural network was chosen following the findings of the study in chapter 4. The second interface tested was a dwell time-based explicit eye input interface. The systems were tested with 10 adults without impairments in a whack-a-mole game. The purpose was to assure the safety of the system and to test its performance before children use it. The performance of the two interfaces was compared.

I. 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 develop the objectives of this study. Each session took between 30 and 60 minutes and were scheduled one month apart.

A. Participants

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

B. Materials

1. Robotic system

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 microcontroller Arduino Leonardo.

The telerobotic haptic system included two PHANToM Premium 1.5A haptic robots (3D Systems, Inc., Rock Hill, SC, USA). One of the robots was placed in the environment (slave side) where it interacted with the moles in the game. The other haptic robot was placed in the user side (master side), for participants to control. The robots were programmed in bilateral teleoperation mode using a PID controller for position control. In a second computer, the Tobii EyeX (Tobii Technology, Stockholm, Sweden) stationary eye tracking system was interfaced to acquire the POG signal. 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). *Figure 5-1* shows the main components and the setup of the robotic system and the activity. More details about the system are described in (Castellanos-Cruz et al., 2018).

2. Haptic Guidance

FRVF with the shape of a cone were 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 further away from it. The design takes into account the findings of a previous study (Sakamaki et al., 2017), 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 the bin where she was supposed to drop them. The findings of that study indicated that the 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 *Figure 5-2A*. The origin of the cone was always at 1cm behind the robot's end-effector. This distance was set so that the user did not feel like he/she was up against the cone at all times. The orientation and the origin of the cone changed as the user moved closer to the target. The angle of the cone was set to 30 degrees. If the user attempted to go outside the cone, forces were applied perpendicular to the cone surface. The force was proportional to the distance between the robot end-effector's position and the cone surface, $F = K * |distance|^2$. K was the constant that regulated the haptic guidance force. The k-force constant was set to 50 for this study for both eye gaze interfaces. As pressing the moles was somewhat difficult, a force field was implemented to attract the robot end-effector towards a virtual vertical line over top of the mole, as illustrated in *Figure 5-2A*. The force 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 user was within 1.5cm away from the virtual line extending vertically through the mole (the mole's x and y coordinates). Once the force field was activated, the cone-shaped guidance was deactivated even if the user was further than 1.5cm away. The cone-shaped guidance was activated again, and the force field deactivated only 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-force constant was 10. In contrast, the constant was set to 50 for the explicit interface. The reason for the different constants is discussed later.

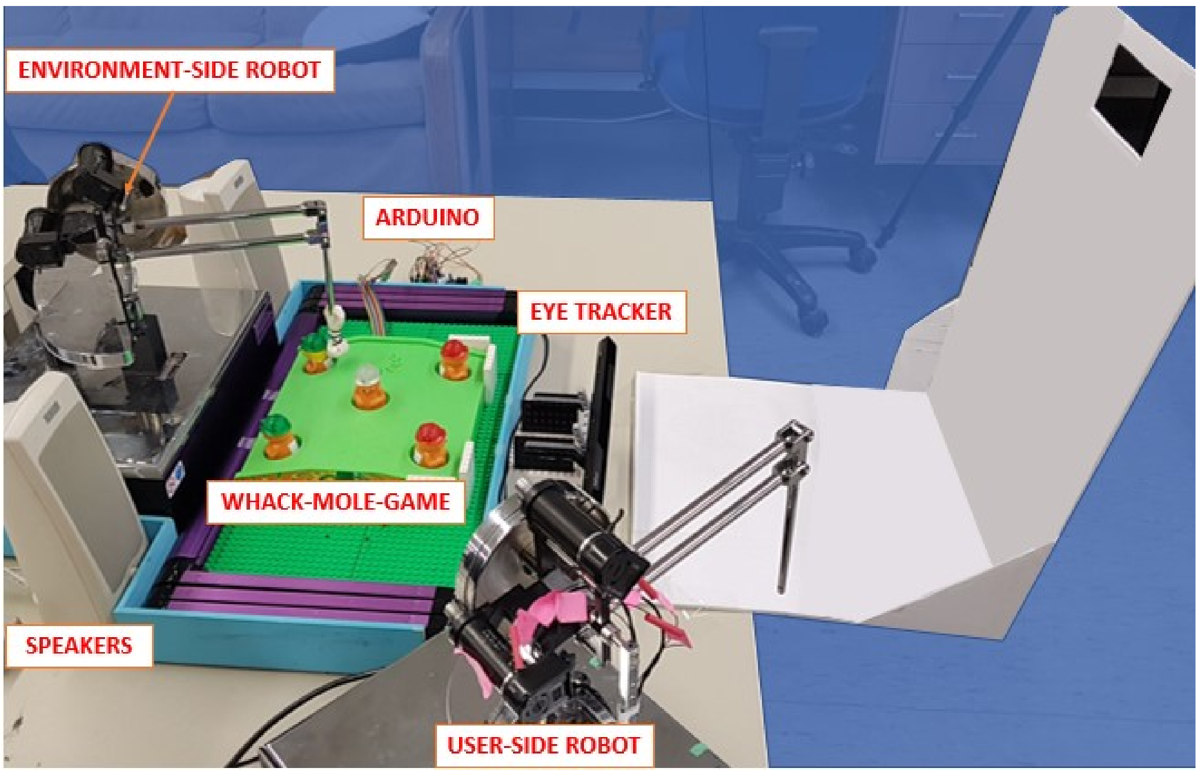

Figure 5-1 Setup of the robotic system and the activity.

3. Eye gaze interfaces

a) Attentive user interface

This interface consisted of integrating 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.

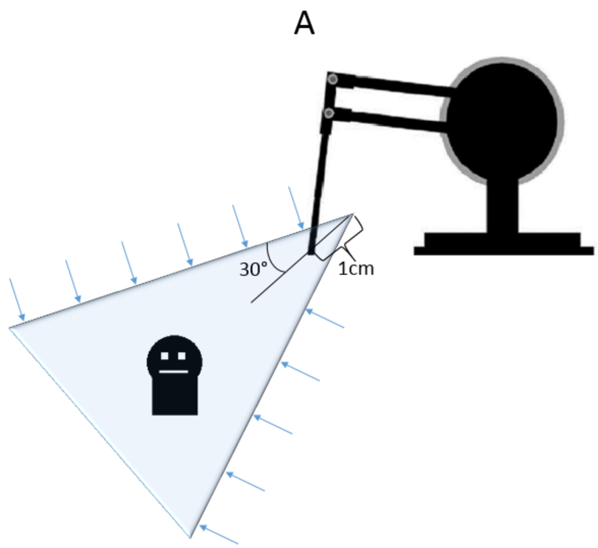

Figure 5-2 Illustration of the haptic guidance: A) cone-shaped FRVF for guiding the user towards moles, B) artificial potential field for whacking the moles.

The performance of training three MLP neural networks with three combinations of input variables was compared: 1) the POG x, and y coordinates, 2) the POG x and y coordinates, and the environment-side robot's x,y,z-position, and 3) the POG x and y coordinates, the environment-side robot's x,y,z-position, 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. To decide which structure to use in session 2, a 5-fold cross-validation was performed. The training was done using the scaled conjugate gradient method, having $\sigma = 5.0e\text{-}5$ and $\lambda = 5.0e\text{-}7$.

The accuracies of the neural network with the three variable combinations were 84.91 ± 5.02%, 90.06 ± 2.44%, and 90.13 ± 2.27, 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. The neural network was trained one more time the using the complete dataset and was integrated it in the Simulink code for activating the haptic guidance in session 2.

b) Explicit eye input interface

This interface was designed using dwell times, which is the most common approach for controlling robots. This interface worked in a two-step process to activate the haptic guidance towards any of 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 what mole he/she had selected. The user could repeat the two-step process in case he/she wanted to go to another mole.

C. Procedure

1. Session 1

The first session was carried out with the purpose of collecting the data needed to train a multilayer perceptron (MLP) neural network of the attentive interface. Participants were asked to use the robot interface with their non-dominant hand to whack 60 moles. Three moles were lit up at a time, 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 each mole. The moles were lit up randomly. To simulate involuntary movements on the users, the y and z axis 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, thus the haptic guidance applied in the second session would be engaged. Participants had the chance to get familiarized with the system by whacking each mole twice.

2. Session 2

The second session consisted of three parts, these 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 *Figure 5-1*. Participants did the activity by looking through the hole 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. They whacked a total of 60 moles. This part showed how the mirroring of the y and z axis affected the users.

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. 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.

D. Data collection and analysis

From session 1, the user's POG and the environment-side robot's x,y,z position 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 position. 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 moles that 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 (Castellanos-Cruz et al., 2018), 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).

From 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. The measure of jerkiness can reflect how the participants' movements were affected by the mirroring of the teleoperation and the haptic guidance. Jerkiness was measured by using the Dimensionless Jerk formula and taking the negative logarithm to improve the sensitivity (LDLJ) (Balasubramanian, Melendez-Calderon, Roby-Brami, & Burdet, 2015; Hogan & Sternad, 2009):

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

where V is the velocity at which the environment-side robot was moving. LDLJ is a valid measure for measuring jerkiness of movements (Balasubramanian et al., 2015). 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 is performed after 300ms at the start of each episode, as humans do not react instantly to light. Kiselev, Espy, & Sheffield (Kiselev, Espy, & Sheffield, 2009) 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.

From 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 errors when the users selected other moles before selecting the intended one during each episode. 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. Linear mixed-effects models were used to create a model for each measure, controlling for the effects of each participant. For instance, the model for time was:

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

where condition refers to the conditions of interest for comparing, for example, “without guidance” and “with guidance”. 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.

II. Results

Table 5-1 lists the means and standard deviations for time, distance, and jerkiness, for the trials with normal teleoperation (from session 2 part 1) and when the y and z axis were mirrored (results obtained from session 2 part 2 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 5-1 Linear mixed-effects models for comparing the results of normal and mirrored teleoperation

	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 5-2 lists the means and standard deviations of time, distance, and jerkiness, when the participants did the activity in the "without guidance" and with guidance during part 2. It also lists the results of the linear mixed-effects model for comparing the results of the two conditions. For part 2 of session 2, 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 5-2 Linear mixed-effects models for comparing the results of "without guidance" and "with guidance" of the attentive interface.

	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 5-3 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 a mean accuracy of 86.43% (SD=15.58). The explicit interface had a mean accuracy of 100% (SD=0) and the number of errors made by the participants when selecting a mole was 11, which were committed 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.

Table 5-3 Linear mixed-effects models and results for the comparison of the attentive and explicit interfaces

	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

III. Discussion

Inverting the axis of the teleoperation increased the difficulty to control the robotic system compared to normal teleoperation. According to Table 5-1, participants spent significantly more time, travelled longer distances, and had jerkier movements. These results reflect that mirroring of the axis indeed induced confusion or involuntary movements in the participants. Participants moved 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, at some extent, a simulation of the movements performed by a person with physical impairments who experiences involuntary movements.

According to Table 5-2, 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 green has, to some extent, a similar movement (e.g. the direction) trajectory as going to the mole in the middle.

There were many differences found between the attentive and explicit interfaces from results in Table 5-3. Regarding time, the participants spent significantly more time to whack each mole with the explicit interface than when using the attentive interface. Theoretically, the users would spend one second (two dwell times) longer than the attentive interface but it was about 1.5 seconds in 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 by using the attentive interface. The main reason is that force of the potential field (*Figure 5-2B*) was higher for the explicit interface, therefore, providing more help to whack the moles. Also, the movements to reach and whack the moles were significantly jerkier using the attentive interface than the explicit interface, because of the accuracy achieved by the neural network. This was why all participants commented that the haptic guidance was sometimes against their movement. Additionally, this was another reason why seven out of ten preferred the explicit interface over the attentive interface.

The accuracy of the explicit interface was 100% because the participants could change their selection if they had selected the wrong target, but at the cost of spending more time to whack each mole. There were 11 errors and were made by five of the ten participants, which represents less than the 2% of the total number of episodes. The low number of errors reflect that the adults did not experience difficulties using the explicit interface, and another reason why seven out of ten preferred this interface. 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 errors will be higher and the accuracy lower. Furthermore, as five adult participants thought it was bothersome or tedious to transition their gaze, it is possible that children may refuse to use the explicit interface. It is still needed to research how children would perform with the explicit interface, especially children under five years old who lack some of the cognitive skills required to control robots (Poletz et al., 2010).

Regarding the k-force constant of the potential field (*Figure 5-2B*), the attentive interface had a lower value than the explicit interface. Having a value of 50 for the k-force 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-force constant had to be decreased to a value (10) where the neural network could respond appropriately, allowing the user to move towards other moles but still be supported to whack the moles. In contrast, these 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-force constant to provide more support for whacking the moles. Regarding the cone-shaped guidance, the k-force 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 and allowing the neural network to be responsive.

A significant finding of this paper was understanding that haptic guidance affected the movements performed by the user i.e., position, velocity and direction, which were input variables of the neural network. In consequence, the haptic guidance influenced the input and output of the neural network. Therefore, different and additional input variables should be considered, as well as different machine learning algorithms.

The 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 effect is also reported in other studies with explicit interfaces

(Majaranta & Bulling, 2014). Despite this effect, seven out of ten participants preferred the explicit interface.

There were some limitations of this study. Only adults without physical impairments tested the system, and not children, who are the target population. However, the results of this study were very helpful to understand that a different approach will be needed to develop an attentive user interface. 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.

IV. Conclusions

The attentive interface offers several advantages over the explicit interface, including lower times to complete the activity and the user's eyes do not get as tired as with the explicit interface. Using the attentive user interface was significantly faster and can be easier and 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. Nevertheless, this type of interface may work in technology that does not provide haptic feedback, e.g., controlling a robot with switches.

Overall, we believe that the attentive interface would be easier and more intuitive to use by children than the explicit interface and can contribute to a more natural interaction between the user, the robot, and the environment. More research needs to be done to create attentive eye gaze interfaces because they can potentially be faster and less complex to use. Next stages of the project will be directed to improve the attentive interface and tested with children with physical impairments.

Chapter 6
Comparison of eye gaze interfaces for controlling haptic robots that support play in children with physical impairments

Chapter 5 introduced some advantages and disadvantages that an attentive interface can have over an explicit interface when adults without physical disabilities use those interfaces while controlling the telerobotic haptic system. However, the literature of eye gaze interfaces is limited and does not report in depth how successful are children when using explicit eye input interfaces (refer to chapter 2 for the papers found on this topic). Children and individuals with disabilities may have trouble using explicit interfaces due to the complexity of controlling it added to operating a robot. Users must think about voluntarily controlling their eye movements (i.e. fixating during a dwell time or doing eye gestures) and at the same time think about controlling the movements of the robot. An explicit interface may not be as accurate with children and individuals with disabilities as it was with adults.

This chapter contains two studies that were carried out with child participants, and participants with physical disabilities. In the first study the participants' eye gaze was examined with the purpose of developing an algorithm for the attentive interface to activate the haptic guidance of a telerobotic haptic system. The second study had the purpose of comparing the attentive user interface with an explicit eye input interface when these activated the haptic guidance of the telerobotic haptic system.

Study 1

In this study, the eye gaze data of children and participants with physical disabilities was collected while they controlled a telerobotic haptic system to reach objects. The data was used to determine an algorithm that could predict the object the user wanted to reach with the robot, so that the algorithm could be used in second study (for the attentive user interface to activate the haptic guidance of a telerobotic haptic system towards the predicted target).

I. Methods

A. Participants

Nine typically developing children participated in this study. Their ages ranged from three years and one month to four years and 10 months (48.3 ± 7.3 months). None of the children had any known physical or visual impairments. Additionally, a child with hemiplegic cerebral palsy participated in this study. He was seven years and four months old. His right limbs were affected, and he had difficulties grasping and reaching objects with his right hand. He was classified as Level I in the Gross Motor Function Classification System (GMFCS) Expanded and Revised (Palisano, Rosenbaum, Bartlett, & Livingston, 2007), which means that he can walk but with limited balance and coordination. He was classified as Level III in the Manual Ability Classification System (MACS, (Eliasson et al., 2006), which means he has difficulties handling objects (with his right hand). He had corrected-to-normal vision (i.e. wore glasses) and had attention deficit hyperactivity disorder, as reported by his parent. Also, a 52-year-old female adult who had quadriplegic cerebral palsy participated in this study. She has difficulties handling objects due to poor motor control and spastic movements. She was classified as Level IV in the GMFCS scale, meaning that she can perform self-mobility by using a powered wheelchair. In addition, she was classified as Level III in the MACS scale, meaning that she has difficulties handling objects. She has alternating amblyopia; thus, her eyes are turned in different directions, but her dominant eye was the left one. Consent was obtained from the children's parents and verbal assent was obtained from the children prior to starting the trials. The adult provided consent for her participation. Ethical approval was obtained from the Health Research Ethics Board – Health Panel at the University of Alberta.

B. Materials

The materials for this study were the same as in chapter 5. The robotic system had two PHANToM Premium 1.5A haptic robots in teleoperation mode. One of them was placed in the environment (environment-side robot) and followed the movements performed by the user on the other robot (user-side robot). The system also included a Tobii EyeX eye tracking system to measure the x and y coordinates of the point of gaze (POG) for the left and right eyes of the user. The sampling frequency of the eye tracker was 40Hz. The robots and the eye tracker were programmed in R2016a Matlab/Simulink.

The activity chosen was the whack-a-mole game that was used in chapter 5, but in this study, it was limited to the three moles that were closest to the environment-side robot. This was to avoid having to move the robot's end effector to the two moles at the limits of the robots' workspace where instability in the teleoperation was more prone to happen. Figure 6-*1* illustrates the setup of the robotic system and the activity. Children looked through the hole of a stand, illustrated in Figure 6-*1*, to avoid losing the view of the eyes due to large head movements. The adult did the activity without the stand. The distance between the participant's eyes and the eye tracker was approximately 65cm, and about 90cm to the rear moles of the game.

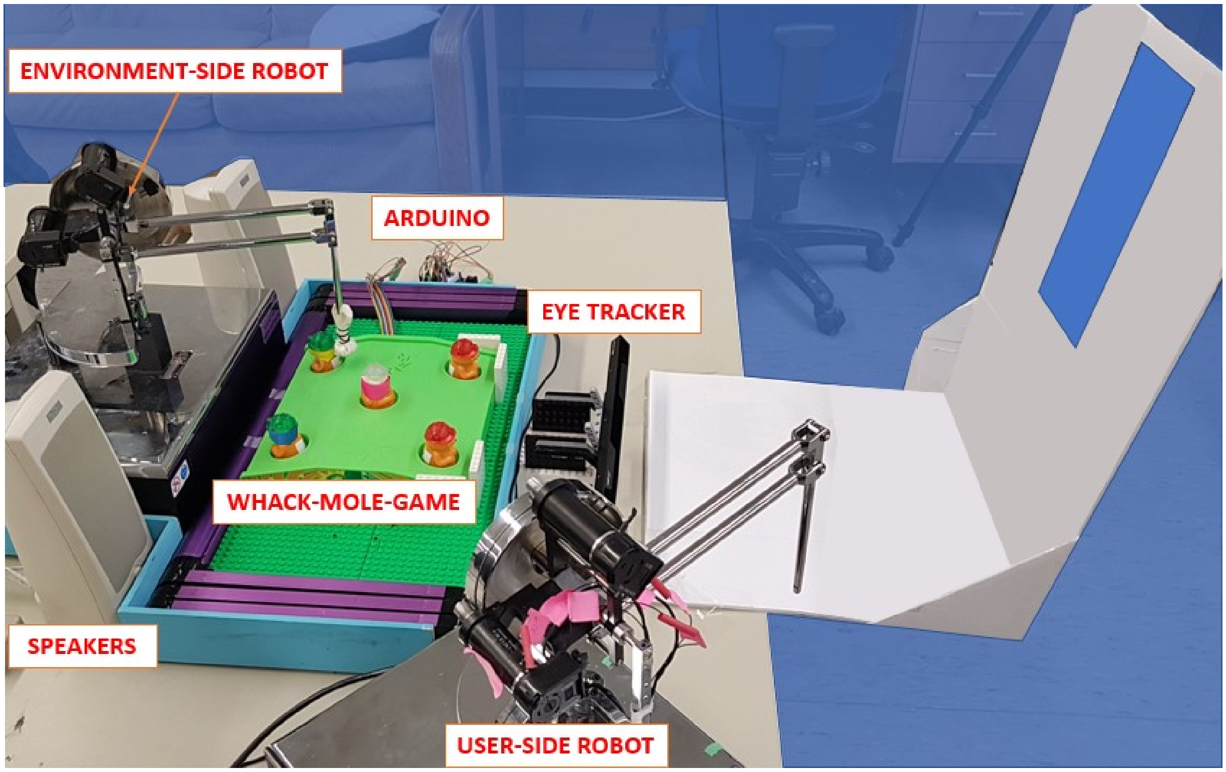

Figure 6-1 Setup of the robotic system and the activity.

C. Procedure

There was only one session for this study that lasted about 20 minutes. The typically developing children and the adult with CP held the robot interface with their dominant hand, while the child with CP held the robot interface with his affected right hand. Because of the adult's eye condition, she did the activity wearing an eye patch on her right eye so that the eye tracker could measure her left eye's POG.

Participants played the whack-a-mole game without the activation of the haptic guidance. One mole was lit up at a time randomly, and after whacking it, a different mole was lit up one second later. The typically developing children whacked 60 moles. The child and the adult with CP whacked 180 and 120 moles, respectively. They took breaks after whacking 60 moles. The participants with CP played the game longer than the typically developing children because they represented a small sample, thus, more data was needed from each of them. Before starting the activity, participants had the chance to get familiar with the system by whacking each mole twice. To keep children engaged, author 1 motivated the children by congratulating them after whacking each mole, e.g. “wow, that was amazing!”, and every few moles the researcher said an engaging sentence such as “He/she is a good player, but can he/she really win at this game? Of course, he/she can!”

D. Data collection and analysis

The participants’ POG and the environment-side robot's x,y,z position was collected, and the session was video recorded. The POG x and y coordinates for the left and right eyes were averaged. After the session, the POG and robot’s position data were synchronized and divided into episodes. An episode was the time interval from the moment a mole was lit up until the participant whacked it. Episodes were excluded from analysis if the participant anticipated and moved from the mole he/she had whacked towards the other two moles without waiting for the one second delay to light up one of the other moles. Exclusion of those episodes was based on if the position of the environment-side robot’s end effector was more than 8cm away from the target mole once it was lit up, i.e. approximately half of the distance between the rear moles. Also, episodes were excluded from the analysis if the eye gaze data was lost due to head movements or if participants were looking outside the eye tracker's workspace (e.g. when children looked at their parents or lifted the robot’s end-effector too high). Episodes were also excluded if users exceeded the force limits of the robots (i.e. 8.5N), since that deactivated the robots’ motors for a few seconds. The remaining number of episodes that were included in analysis were 258 for the typically developing children, 52 for the child with CP, and 39 for the adult with CP. The POG and robot’s position data were examined for understanding the participants’ eye-robot coordination (i.e., coordination of eye movements with the movements of the robot) to implement an algorithm to predict the mole that participants wanted to whack with the robot. The accuracy of the predictions throughout each episode was measured.

II. Results and discussion

All the participants were able to control the robots to complete the activity without help from the researcher, and children seemed to enjoy playing the game using the robots. In this study, children as young as three years of age were capable of controlling the telerobotic system to complete the activity. The environment-side robot followed the movements of the user's hand performed on the user-side robot, allowing the environment-side robot to act as the user's hand. That might have made the robotic system intuitive for the children because it might have seemed like they were moving their own hand. In fact, participants needed very little time to get familiarized with the system to be able to whack the moles.

From the videos, it was possible to observe which types of haptic guidance used by adults in chapter 5 should be applied for the participants in study 2. The typically developing children and the child with disabilities did not have trouble reaching the moles, but sometimes they had some difficulties to push straight down on the moles. Therefore, it was recommended the only guidance needed for the children in study 2 was the artificial potential field above the mole (described in study 2 below). This guidance could help users to push down the moles easier, especially for the child with CP who could not whack the moles as hard as the typically developing children. In the case of the adult with CP, she also had trouble whacking the moles, but she also had difficulties reaching the moles due to spastic movements. Thus, it was recommended that she would also need the cone-shaped forbidden region virtual fixture (described in study 2 below) to help her reach the moles.

Examination of the participants' eye gaze and the robot's position revealed that most of the time the participants' POG was close to the mole they wanted to whack while they moved the robot toward it. Thus, it was determined that an attentive user interface could be developed to predict the mole that participants wanted to whack with the robot by first measuring the distances between the user's POG and each mole, then assigning the mole with the least distance as the predicted mole. In this study, the accuracy of such predictions was 95.84% (SD=9.24%) for the typically developing children, 93.97% (SD=9.64) for the child with CP, and 90.54% (SD=9.76) for the adult with CP. It is important to note that these accuracies were obtained without the activation of the haptic guidance towards the predicted moles. When activating haptic guidance, it would be expected that users might feel that the guidance is opposing their movements, as happened in the study in chapter 5, because the accuracy is not 100%. Nevertheless, this approach has a high accuracy, it is simple, it has low computational expense, and it does not require training as with the neural network in chapter 5. Thus, this approach was used in study 2.

Study 2

The objective of this study was to test the performance of the attentive user interface that was designed in study 1, and to compare it with an explicit eye input interface.

I. Methods

A. Participants

Five typically developing children participated in this study, four of whom participated in study 1. Their ages ranged from three years and 11 months to four years and 10 months (52.8 ± 3.9 months). There were three females and two males. None of the children had any known physical or visual impairments. Additionally, the child and the adult with CP who participated in study 1 also participated in this study. Consent was obtained from the children's parents and verbal assent was obtained from the children prior to starting the trials. The adult provided consent of her participation. Ethical approval was obtained from the Health Research Ethics Board – Health Panel at the University of Alberta.

B. Materials

1. Robotic system and set up

The materials were the same utilized in study 1, including the setup of the robotic system and the activity illustrated in Figure 6-*1*. However, in this study the robots were programmed to apply haptic guidance. Details about the guidance is described in chapter 5, but in short, there were two types of haptic guidance: a cone shaped forbidden region virtual fixture (FRVF) and an artificial potential field (refer to Figure 5-2 for the illustrations). The cone-shaped guidance was designed to help the user reach the moles. It allowed the user to move the robot end-effector closer to the target mole and prevented the user from moving further away from it. The potential field guidance was implemented to help the user whack the moles once the environment-side robot's end effector was close to them. When the robot's end effector was 1.5cm away from the target mole the potential field guidance attracted the robot end-effector towards a virtual vertical line that passed through the mole, this way it helped the user to push straight down on the mole. Forces were applied towards the x and y coordinates of the target mole, but forces were not applied in the z-axis so that the user could move up and down. For all participants the k-force constant was 50 for both types of guidance.

2. Eye gaze interfaces

a) Attentive user interface

The attentive interface was implemented as suggested in study 1 of this chapter. The interface first measured the distances between the user's POG and each mole, then the mole with the least distance was assigned as the predicted mole to which the guidance was directed.

b) Explicit eye input interface

The explicit interface design from chapter 5 was pilot tested with three typically developing children who participated in study 1 but did not go on to participate in study 2. They were two four-year olds and one three-year old. Children were not able to use the interface without constant prompting by researchers. Children also lost their engagement in the game due to the complexity of the interface. That explicit interface had two steps, first to fixate on a change-selection-spot (the middle light of the eye tracker), and then to fixate on the desired mole.

For this study, the interface was resigned to have only one step: looking at a single dwell-spot. Each of three lights in the eye tracker corresponded to each of three moles in the game. Fixating at a light for a minimum of 0.5s activated the guidance towards the respective mole. Once a mole was selected the computer spoke out loud the mole ID (i.e., blue, pink, or yellow) to let the participants know what mole he/she had selected. Figure 6-*2* illustrates how each dwell-spot on the eye tracker corresponded to each mole.

C. Procedure

There was one session which consisted of two parts with a five-minute break in-between. The session took between 20 to 40 minutes.

1. Part 1 – Attentive user interface

This part was focused on testing the attentive user interface while it activated the haptic guidance. This part was carried out using an experimental crossover design, comparing two conditions: when the guidance was activated by the interface (“with guidance”) and not having the guidance (“without guidance”). The attentive interface activated only the artificial potential field guidance for the children (since they only required haptic guidance to whack the moles). In the case of the adult participant, the attentive interface also activated the cone-shaped guidance (the guidance depicted in Figure 5-2*A*) to help her reach the moles since she had spastic movements Before starting the activity, the eye tracker was calibrated with respect to the three moles. In the case of children, it was necessary to call their attention to the moles by pointing to the moles using the environment-side robot´s end effector.

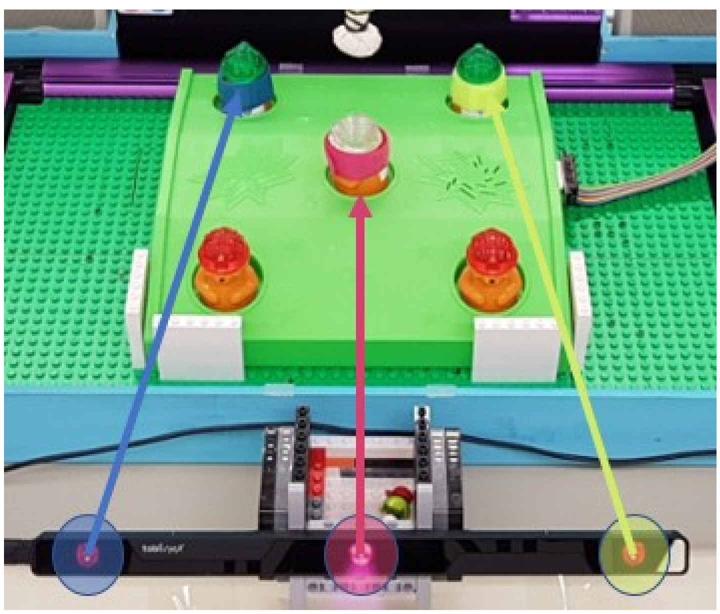

Figure 6-2 Illustration of how the dwell-spot of the explicit eye input interface corresponded to each mole.

For this part, all participants were asked to play the game to whack 54 moles in total. One mole was lit up at a time, and after whacking it, a different mole was lit up immediately. The 54 moles were divided into six sets of nine moles, in which the conditions “with guidance” or “without guidance” were alternated, with the first condition randomly assigned for each participant. Participants did three sets in the “with guidance” condition and three in the “without guidance” condition. For the adult with CP, a short break was given between sets to ask her whether the most recent set was easier than the previous set, but she was not told if guidance was on or off. At the end of the trial, she was asked if her eyes felt tired. Her responses were recorded by the researcher into the research notes. Children were encouraged as in study 1 (e.g., e.g. “wow, that was amazing!”). To keep children engaged to continue to part 2, at the end of this part the computer spoke out loud “(Name of the child), that was awesome, you are the best! I have one more challenge for you, do you want to try?”

2. Part 2 - Explicit eye input interface

This part had the purpose of testing the explicit eye input interface while it activated the haptic guidance. This part of the study only had a “with guidance” condition in order to compare it to the “with guidance” condition of the attentive interface. The explicit interface activated the artificial potential field guidance to help all participants whack the moles, and it also activated the cone-shaped guidance for the adult with CP.

Before starting the activity, an explanation of how to control the robot and the explicit interface was given to the all participants. To get familiar with how the explicit interface worked, all participants whacked 10 moles that were lit up randomly. In the case of the children, researchers pointed out the spots where they had to fixate, and then showed them that the system called out the color of the mole they had selected. The children were asked to whack the mole and to feel the haptic guidance. After whacking the mole, they were shown that the haptic guidance would not let them move towards other moles unless they fixated at the dwell-spots that corresponded to the other moles, this was because the potential field guidance was activated. After familiarization, the children's understanding of the interface was tested by asking them where they had to look if they wanted to whack each mole, e.g. "if you want to whack the blue mole, where do you look?"

For the experiment, participants whacked 45 moles using the explicit interface. Prompting to look at the dwell-spots (eye tracker's lights) was given to the children if they tried to go to the lit-up mole by overcoming the potential field haptic guidance without first fixating at a dwell spot. Children were also encouraged in the activity as in study 1. At the end of the trial, the adult with CP was asked if her eyes felt tired and which interface she preferred and the reasons why. Her responses were recorded into the research notes.

D. Data collection and analysis

The participants' POG and the environment-side robot's position were recorded, and the session was video recorded in both parts of the session. After the session, the POG and robot's position data were synchronized and divided into episodes. In part 1, episodes were excluded when the eye gaze was lost due to head movements or the robot's force limits were exceeded. In the "without guidance" condition, 124 episodes were included for the typically developing children, 22 for the child with CP, and 23 for the adult with CP. In the "with guidance condition", 124 episodes were included for the typically developing children, 21 for the child with CP, and 22 for the adult with CP.

The accuracy of the predictions made by the attentive interface in the "with guidance" condition for each episode was calculated after the participants' POG was closer to the target mole than the other two moles, until the target mole was whacked. The accuracy was measured as the percentage of time in which the output of the attentive interface corresponded to the target mole.

The episodes of each condition were processed after the session to obtain the time the participants took to whack each mole and the distance travelled by the end effector of the robot. Also, the distance that the robot's end effector travelled before the participants' POG was closer to the target mole than the other two moles was measured (referred herein as anticipation distance). Measuring anticipation distance could indicate how the guidance affects the participant's eye-robot coordination during the beginning of the movements. The anticipation distance was measured only in the XY plane of the environment-side robot. For the adult with CP, the jerkiness of the movements was measured using the Log Dimensionless Jerk (LDLJ) formula (Balasubramanian et al., 2015; Hogan & Sternad, 2009) as in chapter 5 of this thesis. This jerkiness measure was used to examine how the adult's movements were affected by the haptic guidance: the lower the value of LDLJ the jerkier the movements.

The results of time, distance, anticipation distance, and jerkiness were compared between the "without guidance" and the "with guidance" conditions. Linear mixed-effects models were used for statically comparing the results obtained from the typically developing children. Independent t-test was applied for the statistical comparisons of the results obtained from the child and the adult with CP, separately. Both statistical tests were performed using a 95% confidence level.

In part 2, episodes were excluded when the eye gaze was lost due to head movements or the robot's force limits were exceeded. The episodes where the participants were prompted were also excluded, although, the percentage of episodes where prompting occurred was calculated. The remaining episodes were 200 for the typically developing children, 19 for the child with CP, and 35 for the adult with CP.

From these episodes the success rate of using the explicit interface for selecting the correct moles was calculated. Success rate was measured as the percentage of episodes in which the participants' fixated at the dwell spot that corresponded to the target mole.

The time the participants took to whack each mole and the distance travelled by the robot's end-effector was calculated in each episode. For the adult with CP, the jerkiness of her movements was also measured using the LDLJ formula.

Results of time, distance, and jerkiness were compared between the attentive and explicit interfaces. Linear mixed-effects models were used for statically comparing the results obtained from the typically developing children. Independent t-test was applied for the statistical comparisons of the results obtained from the child and the adult with CP, separately. Finally, the body and head movements that the participants did when using both eye gaze interfaces were compared by observing the videos of both parts.

II. Results

In part 1, all the participants were able to control the robots to whack all moles without help. During the "with guidance" condition, the accuracy of the predictions of the attentive interface was 96.06% (SD=8.48) for the typically developing children, 94.65% (SD=9.8) for the child with CP, and 89.03% (SD=15.51) for the adult with CP. The adult with CP commented that the "without guidance" and "with guidance" conditions had the same difficulty. In the "with guidance" condition, she felt that the cone-shaped haptic guidance was sometimes against her movements when she was trying to reach the target mole, but that the guidance of the artificial potential field was helpful to whack the moles.

Table 6-1 lists the average times, distances travelled, anticipation distances, and jerkiness to whack each mole, during the “without guidance” and “with guidance” conditions for the attentive interface. Table 6-1 also lists the results of the linear mixed-effects model for comparing the results of the two conditions for the typically developing children. Also, it lists the results of the t-test for the comparisons of the conditions for the participants with CP.

Table 6- 1 Statistical comparisons of the “without guidance” and “with guidance” conditions when participants used the attentive user interface

Variable	Group	Without guidance	With Guidance	β	p-value	t	Degrees of freedom
Time (s)	TDC	1.35 ± 0.40	1.53 ± 0.33	0.181	0.000	4.204	248
	Child CP	2.04 ± 1.03	1.80 ± 0.77	--	0.393	0.860	41
	Adult CP	2.64 ± 1.47	2.40 ± 1.22	--	0.561	0.58	43
Distance (m)	TDC	0.231 ± 0.04	0.228 ± 0.03	-0.002	0.648	-0.456	248
	Child CP	0.649 ± 0.25	0.631 ± 0.27	--	0.903	0.122	41
	Adult CP	0.615 ± 0.33	0.517 ± 0.25	--	0.277	1.101	43
Anticipation distance (m)	TDC	0.044 ± 0.03	0.025 ± 0.02	-0.018	0.000	-5.308	248
	Child CP	0.036 ± 0.01	0.023 ± 0.01	--	0.011	2.646	41
	Adult CP	0.040 ± 0.02	0.025 ± 0.02	--	0.031	2.217	43
Jerkiness	Adult CP	-21.48 ± 2.50	-20.56 ± 2.05	--	0.185	-1.345	43

TDC – Typically developing children

Table 6-2 lists the percentage of episodes where the participants required prompting and the success rate of using the explicit interface to select the correct moles in Part 2. None of the children changed their selection when it was wrong. They moved towards the mole as soon as the computer spoke out the color of a mole, even if they selected the wrong mole. In addition, children tried to move towards the other moles before gazing at the dwell-spots, however, the haptic guidance did not allow them.

Table 6-2 Percentage of episodes that participants required prompting and success rate of using the explicit eye input interface for selecting the correct target moles

Individual	Percentage of episodes that required prompting (%)	Success rate (%)
TDC1	18.37	85
TDC2	5.13	89.19
TDC3	43.9	65.22
TDC4	0	90
TDC5	0	80
Child with CP	52.50	63.16
Adult with CP	0	100

TDC – Typically developing child

Table 6-3 lists the average times, distances travelled, and jerkiness of the movements when the participants used the attentive and explicit interfaces in the "with guidance" condition (note that the results of the attentive interface listed in Table 6-3 are the same listed in Table 6-1 for the “with guidance” condition). The table also lists the results of the statistical tests for comparing the results of the two eye gaze interfaces. The adult with CP commented that the guidance activated by the explicit interface was helpful to reach and whack the moles when using the explicit interface, and she did not feel the forces were against her movements, as with the attentive interface. She felt that her eyes were tired after using the explicit interface, unlike when using the attentive interface. Overall, she said she preferred using the attentive interface because it was not tiring, and it was easier and faster than using the explicit interface.

Table 6-3 Statistical comparisons of the attentive user interface and the explicit eye input interface

Variable	Group	Attentive interface	Explicit interface	β	p-value	t	Degrees of freedom
Time (s)	TDC	1.53 ± 0.33	3.79 ± 1.89	2.2684	0.000	13.84	325
	Child CP	1.80 ± 0.77	3.70 ± 1.50	--	0.000	5.108	38
	Adult CP	2.40 ± 1.22	5.61 ± 2.08	--	0.000	6.526	55
Distance (m)	TDC	0.228 ± 0.03	0.376 ± 0.19	0.1537	0.000	9.644	325
	Child CP	0.631 ± 0.27	0.811 ± 0.34	--	0.086	1.759	38
	Adult CP	0.517 ± 0.25	0.608 ± 0.32	--	0.274	1.103	55
Jerkiness	Adult CP	-20.56 ± 2.05	-20.28 ± 1.97	--	0.618	-0.500	55

TDC – Typically developing children

From the videos, it was possible to observe that all children moved their trunk and heads more when using the explicit interface than when they used the attentive interface. All children turned and inclined their heads in the direction of the dwell spots to select the respective moles, which was unnecessary given the distance they were from the eye tracker. In the case of the adult with CP, she also inclined her head instead of just moving the eyes and sometimes this caused the eye tracker to lose the view of her eye, therefore, she had to lift her head up for the system to be able to record her POG.

III. Discussion

According to Table 6-1, when using the attentive user interface typically developing children significantly spent more time to whack the moles when the guidance was activated ("with guidance" condition) than when it was not ("without guidance" condition). Examination of the children's eye gaze and the environment-side robot's position during both conditions revealed that children moved the robot towards the target mole while they were still looking at the previous mole. Thus, in the "with guidance" condition the guidance was still activated towards the previous mole, preventing them from moving towards the target mole, causing the increase of time to whack the moles.

In the case of the participants with CP, they spent less time to whack the moles when the guidance was activated, although, it was not statistically significant, probably due to the sample size. The child with CP experienced difficulties grasping the user-side robot's end effector tightly and he could not whack the moles as hard as the typically developing children. The artificial potential field guidance in the "with guidance" condition made it easier for him to whack the moles because the potential field guidance helped him to keep the robot over top of the mole so that he just had to push downwards. The potential field guidance also helped the adult with CP to whack the moles when she experienced spastic movements while trying to hit them. However, she did not think that the cone shaped FVRF was helpful to reach the moles because the guidance opposed her movements sometimes.

There was no significant difference between the distance travelled by the robot's end-effector when the participants played the game in the “without guidance” and the “with guidance” conditions. However, the distance travelled by the adult with CP was about 10cm less with the guidance than without the guidance. One possible reason is that the haptic guidance helped to reduce the range of her spastic movements and made her movements less jerky. Her movements were less jerky with the guidance than without it. However, it was not statistically significant, possibly due to the low sample size.

In terms of the anticipation distance, all participants anticipated significantly less "with guidance" than “without guidance". This result reflects how the haptic guidance activated by the attentive interface affected the eye-robot coordination of the participants. Examination of the participant's eye gaze and the robot's position revealed that the participants fixated at the target mole before (case 1) or after (case 2) moving the robot towards it. This is consistent with findings about eye-hand coordination studies indicating that humans typically fixate at a target object slightly before (case 1) or after (case 2) the movement onset of the hand (Issen & Knill, 2012). When case 2 happened during the "with guidance" condition, the haptic guidance prevented the participants from moving towards the target mole. It is possible that when case 2 occurred, participants may have perceived that the guidance was against their movements, as commented upon by the adult with CP.

The attentive user interface achieved high accuracies of 96.06% and 94.65% when the children with and without disabilities used it. The interface did not reach 100% because the children's visual attention was possibly sometimes on the environment-side robot's end effector, rather than always on the target mole. The accuracy of the interface was lower for the adult with CP, 89.03%, and a possible reason is that she experienced involuntary movements that might have driven her visual attention away from the target mole to the environment-side robot's end effector, and in those instances the attentive interface would have made the guidance go towards the closest mole to the end effector. This likely contributed to her comment that the guidance was sometimes against her movements to reach the moles.

Children had some difficulties using the explicit eye input interface despite the interface only requiring one step to operate it. Three out of five typically developing children required promoting to look at the dwell-spots in 5.13 to 43.9% of the episodes, and the child with CP needed prompting in 52.5% of the episodes. When prompting was not required, typically developing children used the explicit interface to select the target moles with a success rate between 65.22% and 90%, and the child with CP's success rate was 63.16%. Children were able to whack the moles despite selecting the wrong moles and this was why none of the children corrected their selections when they were wrong. The reason that they could whack the moles was because the potential field guidance was only activated when the robot's end effector was less than 1.5cm away from the selected mole (correct or wrong). For example, if they had selected the pink mole (mole in the middle of Figure 6-*2*) but the target mole was the blue mole (left), they could whack the blue target mole as long as they did not get close to the pink mole (1.5cm). If the cone-shaped guidance had been activated for the children, then they would possibly have corrected their selections, because, the cone-shaped guidance would have guided them to the selected (non-target) mole. In the case of the adult with CP, she did not have trouble understanding how to use the explicit interface, thus she did not need prompting and her success rate was 100%.

The fact that children required prompting and did not have 100% success at using the explicit interface to select the target mole supports the argument that it can be complex for children to use explicit interfaces. Children had to transition their gaze between the play area and the dwell-spots, and they also had to think about which dwell-spot corresponded to the mole that was lit up and think about moving the robot. With the attentive interface, children only had to think about controlling the robot to complete the activity. In fact, it was perceived that they needed to concentrate more with the explicit interface than with the attentive interface.

The explicit interface not only required the users to change their visual behavior (i.e. maintaining focus on the dwell spots) but it also changed the way they moved their head and trunk. All participants moved their heads in the direction of the dwell-spots, and sometimes this caused the eye tracker to fail at measuring their POG. Participants also moved their head and trunk, which can affect the calibration of the eye tracker or can cause the eyes to be outside the eye tracker's workspace.

Participants spent more time and moved the robot longer distances when using the explicit interface than the attentive interface. Of course, the length of the dwell took time, but children also took additional time to remember that they had to fixate at the dwell-spots, and the adult with CP took additional time because she had to adjust the position of her head for the system to be able to record her POG.

On average, all participants travelled longer distances to whack the moles using the explicit interface than when using the attentive interface. However, statistical significance was only achieved for the results of the typically developing children. The longer distance with the explicit interface was due to the movements the participants made with the robot while fixating at the dwell-spots. In the case of children, they were constantly moving while they were trying to overcome the haptic guidance to go to the target mole. The adult with CP moved the robot while fixating at the dwell spots because she could not keep the robot's end-effector completely steady.

In terms of jerkiness, there was no statistical difference between the jerkiness of the movements of the adult with CP while using the explicit and attentive interfaces. However, she mentioned that the guidance was more helpful while using the explicit interface than the attentive interface. The reason was likely because once she selected a mole, the haptic guidance applied by the explicit interface was directed to towards the selected mole the entire time and did not oppose her movements as it sometimes did with the attentive interface.

The accurate guidance with the explicit interface was at the cost of spending more time and travelling longer distances. Additionally, the adult experienced tiredness in her eyes when using the explicit interface, whereas she did not feel tiredness in her eyes tired when using the attentive interface. Overall, the adult with CP said she preferred the attentive interface because it was easier and faster to use.

The usability of the eye gaze interfaces may be affected because of the stationary eye tracker. With this eye tracking system, the user had to be near, and if the user moved or inclined her/his head, re-calibration may be required, or it is possible that eye tracker loses the view of the eyes. However, these limitations could be addressed by using a wearable eye tracker.

The attentive and explicit interfaces may not work in other activities. The attentive interface may not work in other activities where there are more toys or where the 3D POG (i.e., x, y coordinates and depth of the POG) is required to discriminate between objects that are at different distances but in the same line of vision. Also, the attentive interface may not work as well as it did in this study if the objects or toys are closer to each other. The complexity of the explicit eye input interface may rise and lead to different results if there are more objects in the environment because more dwell-spot would need to be added.

There were some limitations of this study. Due to the sample size the results of this study cannot be generalized to all children, nor to the different types of cerebral palsy or physical impairments. The comparisons between the two eye gaze interfaces may be biased because all participants tested the attentive interface first and then the explicit interface. In the case of the child with CP, his performance at using the explicit interface may have been affected by his attention deficit disorder. Also, this study had only a single intervention, and results could improve as the participants gain more experience with the robotic system and the interfaces. For future work, it will be necessary to recruit more children, and test the interfaces in other activities and using a wearable eye tracker.

IV. Conclusions

This study showed that the eyes can be used to predict what object a person wants to reach with a telerobotic haptic system with high accuracy. An attentive user interface can be implemented to apply haptic guidance towards the predicted object without requiring the user to change his/her visual behavior as required with explicit eye input interfaces. This study also showed that children were not 100% successful at using the explicit eye input interface and required prompting to use it. The attentive interface implemented provided more advantages than the explicit interface, such as lower times to complete the activity and less distances travelled by the robot. The adult participant with spastic cerebral palsy perceived that the attentive interface was easier and faster to control the robotic system than with the explicit interface. The attentive interface activated the guidance according to the user's eye-robot coordination for the robotic system of this study, and this contributed to a more natural and intuitive interaction between the user, the robot, and the environment than the explicit interface. However, more research is required to test these interfaces in other activities and with a larger sample size of people with physical disabilities.

Chapter 7
General discussion and conclusion

Most of the work of this thesis was dedicated to developing an attentive user interface, the main objective of this thesis. The purpose of the attentive interface was to predict the target object a user wanted to reach and to activate haptic guidance to guide the user towards the predicted object. Eye gaze has been used to predict the target objects that a person wants to reach or select when interacting with a computer (Novak, Omlin, Leins-Hess, & Riener, 2013; Biswas & Langdon, 2014) but not in a physical environment. The novelty of this thesis lies on the fact that the prediction of the target object was done in the physical world and using a stationary/remote eye tracker. Also, users of the developed system interacted with the environment directly and not through a computer screen.

The study in chapter 3 demonstrated that it was possible to predict the target object when participants (adults without disabilities) were using a telerobotic system in a task with three target blocks. The accuracy of the predictions after the participants started moving the robot towards the target block was above 80%. In that task the participants had to start in the same position for every episode, thus the episodes could be separated into two parts: using the eyes to scan the target objects while the robot was at the starting position and moving the robot towards the target. In the context of play, children are constantly moving, therefore, separating the episodes into "scanning" and "moving" may not be possible. Therefore, it was necessary to implement an algorithm that could predict the target object without knowing beforehand if the user was "scanning" or "moving". In chapter 4 a more flexible task was used, a whack-a-mole game, in which the participants (adults without disabilities) moved without imposing a start position, although, the game indicated which mole the user should whack. Since the episodes were not separated into scanning and moving intervals, instead of using a variable that indicated the movement onset as in Chapter 3, the x, y, z position of the environment side robot's end effector was used. In addition to those variables, the neural network (the algorithm that performed the best in chapter 3) made the predictions based on the POG x, y coordinates. The neural network was trained with data collected when the participants used the two haptic robots in three conditions: typical teleoperation, inverted teleoperation without guidance, and inverted teleoperation with guidance. The inversion of the teleoperation was to try to simulate involuntary movements that people with physical disabilities may experience and thus, make it so the non-disabled participants might make use of the haptic guidance. The neural network reached the highest accuracy (70.7%) when it was trained with a one-second window. However, the accuracy was lower than in chapter 3, suggesting that the network was not able to respond appropriately when the data from all the conditions were grouped together for training. To increase the accuracy of the predictions, it was suggested to train the neural network for each condition. It was also proposed that in future work with people with physical disabilities, the neural network would need to be trained for each person, since everyone would have different eye-robot coordination.

Chapter 5 showed that the accuracy improved (86.43%) when the neural network was trained only for one condition (mirrored teleoperation), and when more input variables such as the velocity and direction of the environment-side robot's end effector were included, in addition to the position. However, the experiments revealed that with this accuracy the haptic guidance applied by the attentive interface sometimes opposed the movements of the participants. It was also found that the haptic guidance affected the participants' movements and the position of the environment-side robot, which was an input variable of the neural network. As a consequence, the neural network could not recognize appropriately the movements performed by the user on the robotic system. It was concluded that the attentive interface should only use input variables related to the user's eye gaze. Despite this result, it is important to highlight that machine learning algorithms such as neural networks can learn the patterns of the user's eye-robot coordination and predict the target object that a user wants to reach with the robot.

Chapter 6 took a different approach for the attentive interface and it replaced the neural network with an algorithm that did not have the robot's end effector position as input. The algorithm was based on the eye-robot coordination observed in the trials conducted in study 1 in Chapter 6. From that data it was determined that the guidance could be directed towards the mole that was closest to the user's POG. In study 2 the attentive interface used that algorithm to activate the haptic guidance. The predictions of the attentive interface had a high accuracy: for children the accuracy was above 94%, and for the adult with CP it was 89.03%. The attentive interface did not reach 100% accuracy because the participant's visual attention was not always on the target mole, and possibly on the environment-side robot's end effector as well. In the case of the adult with CP the accuracy was lower than for the children because she experienced spastic movements that likely drove her attention away from the target mole to the robot's end effector. Due to the imperfect accuracy of the attentive interfaces, adults with and without physical disabilities felt that the guidance activated by the attentive interface was sometimes against their movements. The accuracy of the predictions needs to be improved otherwise users may reject this interface (suggestions to improve accuracy are given below in future research). More research is required to determine what a minimum acceptable accuracy would be so that the user does not feel that the guidance is opposing his/her movements.

The second objective of this thesis was to compare the attentive and explicit eye input interfaces. When all participants (adults without physical disabilities, children with and without disabilities and the adult with CP) used the attentive interface they spent less time to whack each mole than the explicit interface. When the children and the adult with CP used the attentive interface, they travelled shorter distances to whack each mole than when they used the explicit interface. Additionally, adults without disabilities and the adult with CP felt that their eyes were less tired with the attentive interface than the explicit interface.

Even though the attentive interface was faster, seven out of ten adults without physical disabilities preferred the explicit interface over the attentive interface. One of the reasons given was that the guidance activated by the attentive interface was sometimes against their movements. In contrast, the guidance activated by the explicit interface was 100% accurate because it was directed to the mole that the participants pointed at with their gaze. However, the adult with CP had difficulties using the explicit interface because she inclined her head and the eye tracker lost the view of her eye, therefore she had to adjust her position leading to an increase in time to whack the moles. The adult with CP preferred the attentive user over the explicit interface because she perceived it as easier and faster to use.

All adults were capable of voluntarily controlling their eye movements to fixate at the dwell-spots while also thinking of controlling the robot. This argument is supported by the fact that all the adults had 100% success rate at using the explicit interface for selecting the correct moles. The adult with CP had a 100% success rate at selecting targets using the interface and did not require prompting, but she did have to move her head for the interface to register her eye gaze. In the case of the children, they needed to be prompted to transition their gaze to the dwell-spots and when they were not prompted they did not have 100% success rate at using the interface to select the correct mole. From the videos, it looked like the children had to concentrate more with the explicit interface than with the attentive interface. The explicit interface was indeed challenging because they had to remember to transition their eye gaze between the play area and the dwell-spots and fixate during the dwell time.

When children and the adult with CP looked at the dwell spots, they made head and trunk movements that were not observed when they used the attentive interface, such as inclining their head towards or in the direction of the dwell spots. This behaviour was not observed with the adults without disabilities, most likely because they did the activity looking through a hole that was smaller than the one children used, and it prevented them from doing head movements.

The telerobotic haptic system developed for this thesis was easy and intuitive to use, and children as young as three years of age were capable of using it to play the whack a mole game. It was easy and intuitive because the environment-side robot acted as the user's hand by following the movements performed by the user on the user-side robot. Children did not need help or prompting to control the robotic system for completing the activity. If children used other robots that have a higher cognitive demand the results of success rate and prompting of the explicit interface could have been different. It is possible that they may need more prompting and they may have a lower success rate than the results reported in chapter 6.

In this thesis two different explicit interfaces were developed as described in chapters five and six. The interfaces were designed using dwell times, as it was the approach more often used in the literature of eye gaze interfaces for controlling robots. The explicit interface of chapter 5 required two steps to activate the guidance towards a mole, first fixating at a dwell-spot outside the play area, and then fixating at the mole they wanted to whack. However, three typically developing children were not able to use the interface by themselves and required constant prompting from the researcher. The explicit interface developed for chapter 6 only required one step, fixating at the dwell-spot that corresponded to each mole. The limitation of this interface is that if more toys or objects are introduced in the activity, then more dwell-spots need to be created, and this may increase the complexity of the explicit interface. It will be necessary to test the explicit interfaces with more children of different ages and in other activities with different targets.

An additional finding of this thesis was that the cone-shaped guidance worked better than the spherical-shaped guidance. According to the results of the study in chapter 4, the spherical-shaped guidance did not improve the performance (time and distance) of the participants in the game. Participants often followed the walls of the sphere and did not get closer to the target mole. In contrast, the cone-shaped guidance improved the performance (time and distance) of participants in the studies of chapters five and six. The cone encouraged the participants to get closer to the target mole. However, this thesis did not compare the cone-shaped guidance to other types of guidance such as trajectories or artificial potential fields, which may lead to a higher performance. It is still required to explore what type of haptic guidance will best support people with physical disabilities and their movement intentions.

This thesis contributes to the understanding of eye-robot coordination that may happen when people use telerobotic haptic systems, such as the one developed for this thesis. From eye-hand coordination studies, we know that people fixate at the target object slightly before or after their hand starts moving (Issen & Knill, 2012). In chapters 3 and 6, it was possible to observe a similar behavior with the robotic system. Another fact about eye-hand coordination is that when people without physical disabilities move their hand to reach an object, they fixate at the object and never at the hand (Johansson et al., 2001). However, this was not the case when adults used the telerobotic haptic system.

Results suggests that what people are attending to might not be the what they want to act upon. In chapter 4 it was observed that adults fixated at the target object 61.01% (SD=22.72) of the time that it took them to reach and whack the mole. The remaining percentage of the time participants were probably looking at the environment-side robot's end effector or the other non-target moles. The results of the studies in chapter 6 revealed that the POG was closer to the target object than to the other objects (moles) 96.06% (SD=8.48) of the time for the typically developing children, 94.65% (SD=9.8) for the child with hemiplegic CP, and 89.03% (SD=15.51) for the adult with spastic CP.

Understanding the eye-robot coordination can contribute to develop attentive interfaces that predict the target object that the user wants to reach with the robot. It was possible to use information learnt in chapter 6 about the eye-robot coordination for implementing the attentive interface that activated the guidance towards the mole that was closest to the user's POG.

A. Limitations

The limitations of this thesis include the sample sizes for each study, especially for individuals with physical disabilities. The findings of this thesis cannot be generalized for all children nor to all the population with physical disabilities. The comparisons between the attentive interfaces and the explicit interfaces may have been biased by a learning effect because of the order in which they were tested. A longer washout period or a cross-over design would have controlled for the possible order or learning effects. Another limitation of this thesis was that the calibration accuracy of the eye tracker was not measured. The reason was that the exact locations of the moles with respect to the eye tracker´s frame of reference were unknown. This is not an issue when the eye tracker is mounted on a computer screen, the position of each pixel of the screen is approximately the same with respect to the eye tracker´s frame of reference. For future research it will be necessary to develop a method to calculate the calibration accuracy of eye trackers for applications in a physical environment.

B. Future work

For future work it will be necessary to test the attentive user and explicit interfaces in other tasks and with other robots. The whack-a-mole game was a structured activity, and the game indicated to the participants which mole they had to whack. Also, the game had the moles in fixed positions and these were relatively far from each other. The attentive interfaces may not work in activities where the objects are closer to each other, or if the objects can be moved with the robot. The eye-robot coordination may be different in other tasks and with other kind of robots. For this reason, the attentive interfaces may not work well for all playful activities. In the case of activities where the toys are not fixed, it will be necessary to integrate a camera and computer vision techniques to detect the objects and their positions. Also, the eye tracker should be replaced by eye tracking glasses for people who do not find them uncomfortable to wear. The Tobii EyeX is a stationary eye tracking system and requires the users to be close to it, and it cannot measure the POG when the user moves the head out of its field of vision. For those users who do not like wearing eye tracking glasses, it will be necessary to place the eye tracker at a fixed distance from the user, for example, on the user´s wheelchair. To avoid losing the view of the user´s eyes, a possible solution is to have multiple eye trackers in different positions. Additionally, an algorithm for measuring the depth of the POG may also be necessary for activities where the objects are positioned at different distances but on the same line of vision of the user.

It will be necessary to do more research to improve the accuracy of the attentive interface. Perhaps adding more inputs from the user and the robotic system could improve the accuracy of the predictions. For example, the accuracy could be improved by having signals from the user’s muscle and brain activity as inputs, this way the interface could know more about the user's movement intentions. Also, a force sensor could be used to detect when the guidance is pushing against the user's movements, that way the system could know when the prediction was wrong. In the case of explicit interfaces, it will be necessary to determine at what age children can successfully use this type of interfaces. Also, it will be necessary to research how to make explicit interfaces easier to control for children, for example, not having dwell spots outside the play area.

C. Conclusions

In conclusion, this thesis demonstrated that the POG was a significant predictor of the target object that people wanted to reach with the robot. Results demonstrated that attentive interfaces can be faster to use than explicit interfaces. Children were able to use the attentive interface to complete the activity without help or prompting, whereas children required prompting and were not 100% successful at using the explicit interface. Attentive interfaces can contribute to a more natural interaction between the user, the robot and the environment, meaning that the user does not have to change his/her visual behavior or the way he/she uses the robot. Also, attentive interfaces may reduce the cognitive load of having to think about voluntarily controlling the eye movements, allowing the user to focus on controlling the robot to complete the activity. However, attentive interfaces may not be 100% accurate in their predictions and this may lead to rejection of the interface. Therefore, it is still necessary to improve the accuracy by adding more input variables about the user, this could allow the interface to get more insight about the user's movement intentions.

References

Adams, K. D., Rios Rincón, A. M., Becerra Puyo, L. M., Castellanos Cruz, J. L., Gómez Medina, M. F., Cook, A. M., & Encarnação, P. (2017). An exploratory study of children's pretend play when using a switch-controlled assistive robot to manipulate toys. *British Journal of Occupational Therapy*, *80*(4), 216–224. https://doi.org/10.1177/0308022616680363

Amantis, R., Corradi, F., Molteni, A. M., Massara, B., Orlandi, M., Federici, S., … Mele, M. L. (2011). Eye-tracking assistive technology: Is this effective for the developmental age? Evaluation of eye-tracking systems for children and adolescents with cerebral palsy. *Assistive Technology Research Series*, *29*, 489–496. https://doi.org/10.3233/978-1-60750-814-4-489

Arai, K., & Yajima, K. (2011). Robot arm utilized having meal support system based on computer input by human eyes only. *Journal of Human Computer Interaction (IJHCI)*, *2*(1), 1–9.

Atashzar, S. F., Jafari, N., Shahbazi, M., Janz, H., Tavakoli, M., Patel, R. V, & Adams, K. (2017). Telerobotics-assisted platform for enhancing interaction with physical environments for people living with cerebral palsy. *Journal of Medical Robotics Research*, *2*(2), 1–15. https://doi.org/10.1142/S2424905X17400013

Balasubramanian, S., Melendez-Calderon, A., Roby-Brami, A., & Burdet, E. (2015). On the analysis of movement smoothness. *Journal of NeuroEngineering and Rehabilitation*, *12*(1), 1–11. https://doi.org/10.1186/s12984-015-0090-9

Banerjee, A., Datta, S., Das, P., Konar, A., Tibarewala, D. N., & Janarthanan, R. (2012). Electrooculogram based online control signal generation for wheelchair. Paper presented at 2012 International Symposium on Electronic System Design (ISED) (pp. 251–255). https://doi.org/10.1109/ISED.2012.12

Baptista, P. M., Mercadante, M. T., Macedo, E. C., & Schwartzman, J. S. (2006). Cognitive performance in Rett syndrome girls: A pilot study using eyetracking technology. *Journal of Intellectual Disability Research*, *50*(9), 662–666. https://doi.org/10.1111/j.1365-2788.2006.00818.x

Barbuceanu, F., Antonya, C., Duguleana, M., & Rusak, Z. (2011). Attentive user interface for interaction within virtual reality environments based on gaze analysis. Paper presented at 14th International Conference on Human-Computer Interaction: Interaction Techniques and Environments (Vol. Part II, pp. 204–213). https://doi.org/10.1007/978-3-642-21605-3

Barea, R., Boquete, L., Mazo, M., & López, E. (2002). System for assisted mobility using eye movements based on electrooculography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *10*(4), 209–218. https://doi.org/10.1109/TNSRE.2002.806829

Ben Taher, F., Ben Amor, N., & Jallouli, M. (2015). An extended eye movement tracker system for an electric wheelchair movement control. Paper presented at the IEEE/ACS International Conference on Computer Systems and Applications, AICCSA. https://doi.org/10.1109/AICCSA.2015.7507101

Biswas, P., & Langdon, P. M. (2014). Multi-modal target prediction. In C. Stephanidis & M.

Antona (Eds.), *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access. Lecture Notes in Computer Science* (Vol. 8513). Springer, Cham. https://doi.org/10.1007/978-3-319-07437-5_30

Borgestig, M., Sandqvist, J., Parsons, R., Falkmer, T., & Hemmingsson, H. (2016). Eye gaze performance for children with severe physical impairments using gaze-based assistive technology—A longitudinal study. *Assistive Technology*, *28*(2), 93–102. https://doi.org/10.1080/10400435.2015.1092182

Bouwsema, H., Kyberd, P. J., Hill, W., van der Sluis, C. K., & Bongers, R. M. (2012). Determining skill level in myoelectric prosthetics use with multiple outcome measures. *Journal of Rehabilitation Research and Development*, *49*(9), 1331–1348.

© [2018] IEEE. Reprinted, with permission, from Castellanos-Cruz, J. L., Gómez-Medina, M. F., Tavakoli, M., Pilarski, P. M., & Adams, K. D. (2018). Preliminary Testing of a Telerobotic Haptic System and a Neural Network to Predict Targets during a Playful Activity. Paper presented at the 7th International Conference on Biomedical Robotics and Biomechatronics (pp. 1280–1285). Enschede, Netherlands: IEEE. https://doi.org/10.1109/BIOROB.2018.8487612

© [2017] IEEE. Reprinted, with permission, from Castellanos, J., Gomez, M. F., & Adams, K. (2017). Using machine learning based on eye gaze to predict targets: An exploratory study. Paper presented at the IEEE Symposium series on computational intelligence. Honolulu, HL, USA: IEEE. https://doi.org/10.1109/SSCI.2017.8285207

Chen, X., Ragonesi, C., Galloway, J. C., & Agrawal, S. K. (2011). Training Toddlers Seated on Mobile Robots to Drive Indoors Amidst Obstacles. *IEEE Transactions on Neural Systems and Rahabilitation Engineering*, *19*(3), 271–279.

Chin, C. A., Barreto, A., Cremades, J. G., & Adjouadi, M. (2008). Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities. *Journal of Rehabilitation Research and Development*, *45*(1), 161–174. https://doi.org/10.1682/JRRD.2007.03.0050

Choi, K. S., & Lo, K. H. (2011). A hand rehabilitation system with force feedback for children with cerebral palsy: two case studies. *Disability and Rehabilitation*, *33*(17–18), 1704–1714. https://doi.org/10.3109/09638288.2010.535091

Cook, A., Encarnação, P., & Adams, K. (2010). Robots: Assistive technologies for play, learning and cognitive development. *Technology and Disability*, *22*(3), 127–145. https://doi.org/10.3233/TAD-2010-0297

Cook, A., Howery, K., Gu, J., & Meng, M. (2000). Robot enhanced interaction and learning for children with profound physical disabilities. *Technology & Disability*, *13*(1), 1–8.

Cook, A. M., Adams, K., Volden, J., Harbottle, N., & Harbottle, C. (2011). Using Lego robots to estimate cognitive ability in children who have severe physical disabilities. *Disability and Rehabilitation: Assistive Technology*, *6*(4), 338–346. https://doi.org/10.3109/17483107.2010.534231

Cook, A. M., Bentz, B., Harbottle, N., Lynch, C., & Miller, B. (2005). School-Based Use of a

Robotic Arm System by Children with Disabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *13*(4), 452–460. https://doi.org/10.1109/TNSRE.2005.856075

Cui, J., Tosunoglu, S., Roberts, R., Moore, C., & Repperger, D. W. (2003). A review of teleoperation system control. Paper presented at the Conference on Recent Advances in Robotics *(FCRAR)* (pp. 1–12). Florida.

Deitz, J. C., & Swinth, Y. (2008). Accessing Play Through Assistive Technology. In L. D. Parham & L. S. Fazio (Eds.), *Play in Occupational Therapy* (2nd ed., pp. 395–412). St. Louis, USA: Mosby Elsevier.

Demain, S., Metcalf, C. D., Merrett, G. V, Zheng, D., & Cunningham, S. (2013). A narrative review on haptic devices: relating the physiology and psychophysical properties of the hand to devices for rehabilitation in central nervous system disorders. *Disability and Rehabilitation: Assistive Technology*, *8*(3), 181–189. https://doi.org/10.3109/17483107.2012.697532

Dhas, B. N., Samuel, P. S., & Manigandan, C. (2014). Use of computer access technology as an alternative to writing for a pre-school child with athetoid cerebral palsy - A case report. *Occupational Therapy in Health Care*, *28*(3), 318–332. https://doi.org/10.3109/07380577.2013.874063

Dybdal, M. L., Agustin, J. S., & Hansen, J. P. (2012). Gaze input for mobile devices by dwell and gestures. Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA ’12 (pp. 225–228). https://doi.org/10.1145/2168556.2168601

Eliasson, A. C., Krumlinde-Sundholm, L., Rösblad, B., Beckung, E., Arner, M., Öhrvall, A. M., & Rosenbaum, P. (2006). The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Developmental Medicine & Child Neurology*, *48*(07), 549–554. https://doi.org/10.1017/S0012162206001162

Encarnação, P., Leite, T., Nunes, C., Nunes da Ponte, M., Adams, K., Cook, A., … Ribeiro, M. (2017). Using assistive robots to promote inclusive education. *Disability and Rehabilitation: Assistive Technology*, *12*(4), 352–372. https://doi.org/10.3109/17483107.2016.1167970

Fager, S., Bardach, L., Russell, S., & Higginbotham, J. (2012). Access to augmentative and alternative communication: New technologies and clinical decision-making. *Journal of Pediatric Rehabilitation Medicine*, *5*(1), 53–61. https://doi.org/10.3233/PRM-2012-0196

Fejtova, M., Figueiredo, L., Novak, P., Stepankova, O., & Gomes, A. (2009). Hands-free interaction with a computer and other technologies. *Universal Access in the Information Society*, *8*, 277–295. https://doi.org/10.1007/s10209-009-0147-2

Fenson, L., & Schell, R. (1985). The origins of exploratory play. *Early Child Development and Care*, *19*(1–2), 3–24. https://doi.org/10.1080/0300443850190102

Ferland, F. (2003). Le jeu et l’enfant. In F. Ferland (Ed.), *Le modèle ludique* (3rd ed., pp. 15–22). Montréal: Les Presses de l’Université de Montréal.

Frisoli, A., Loconsole, C., Leonardis, D., Bann, F., Barsotti, M., Chisari, C., & Bergamasco, M.

(2012). A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *42*(6), 1169–1179. https://doi.org/10.1109/TSMCC.2012.2226444

Garvey, C. (1990). *Play*. Cambridge,MA: Harvard University Press.

Gibaldi, A., Vanegas, M., Bex, P. J., & Maiello, G. (2016). Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. *Behavior Research Methods*. https://doi.org/10.3758/s13428-016-0762-9

Hansen, J. P., Alapetite, A., MacKenzie, I. S., & Møllenbach, E. (2014). The use of gaze to control drones. Paper presented at the Symposium on Eye Tracking Research and Applications - ETRA ’14 (pp. 27–34). Florida. https://doi.org/10.1145/2578153.2578156

Haykin, S. (2008). Introduction. In *Neural Networks and Learning Machines* (3rd ed., Vol. 3, pp. 1–46). New York: Pearson. https://doi.org/978-0131471399

Hayward, V., Astley, O. R., Cruz-Hernandez, M., Grant, D., & Robles-De-La-Torre, G. (2004). Haptic interfaces and devices. *Sensor Review*, *24*(1), 16–29. https://doi.org/10.1108/02602280410515770

Hemmingsson, H., Ahlsten, G., Wandin, H., Rytterström, P., & Borgestig, M. (2018). Eye-Gaze Control Technology as Early Intervention for a Non-Verbal Young Child with High Spinal Cord Injury: A Case Report. *Technologies*, *6*(1), 12. https://doi.org/10.3390/technologies6010012

Henry, A. (2008). Assessment of play and leisure in children and adolescents. In L. D. Parham & L. S. Fazio (Eds.), *Play in Occupational Therapy for Children* (2nd ed., pp. 95–193). St. Louis, USA: Mosby Elsevier. https://doi.org/10.1016/B978-032302954-4.10005-4

Hogan, N., & Sternad, D. (2009). Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of Motor Behavior*, *41*(6), 529–534. https://doi.org/10.3200/35-09-004-RC

Hornof, A., Cavender, A., & Hoselton, R. (2004). Eyedraw: A System for Drawing Pictures with Eye Movements. Paper presented at the 6th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 86–93). Atlanta, GA. https://doi.org/10.1145/1028630.1028647

Hyrskykari, A., Majaranta, P., Aaltonen, A., & Kari-Jouko, R. (2000). Design issues of iDict: A gaze-assisted translation aid. Paper presented at the Eye Tracking Research and Applications Symposium (pp. 9–14). https://doi.org/10.1145/355017.355019

Issen, L. A., & Knill, D. C. (2012). Decoupling eye and hand movement control: Visual short-term memory influences reach planning more than saccade planning. *Journal of Vision*, *12*(1), 1–13. https://doi.org/10.1167/12.1.3.Introduction

Jacob, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems*, *9*(3), 152–169. https://doi.org/10.1145/123078.128728

Jafari, N., Adams, K., & Tavakoli, M. (2016). Haptics to improve task performance in people with disabilities : A review of previous studies and a guide to future research with children with disabilities. *Journal of Rehabilitation and Assistive Technologies Engineering*, *3*, 1–13. https://doi.org/10.1177/2055668316668147

Jafari, N., Adams, K., Tavakoli, M., & Wiebe, S. (2017). Development of an assistive robotic system with virtual assistance to enhance play for children with disabilities: A preliminary study. *Journal of Medical and Biological Engineering*, 1–13. https://doi.org/10.1007/s40846-017-0305-6

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-Hand Coordination in Object Manipulation. *Journal of Neuroscience*, *21*(17), 6917–6932.

Kang, J., Logan, S., Galloway, J. C., & Agrawal, S. K. (2014). A chase-game to teach children on a robot to follow moving objects. Paper presented at the IEEE International Conference on Robotics and Automation (pp. 234–239). Hong Kong, China: IEEE. https://doi.org/10.1109/ICRA.2014.6906615

Karlsson, P., Allsop, A., Dee-Price, B.-J., & Wallen, M. (2017). Eye-gaze control technology for children, adolescents and adults with cerebral palsy with significant physical disability: Findings from a systematic review. *Developmental Neurorehabilitation*, *21*(8), 497–505. https://doi.org/10.1080/17518423.2017.1362057

Keates, S., Langdon, P., Clarkson, J., & Robinson, P. (2000). Investigating the use of force feedback for motion-impaired users. Paper presented at the ERCIM Workshop "User Interfaces for All." Florence, Italy.

Kiselev, S., Espy, K. A., & Sheffield, T. (2009). Age-related differences in reaction time task performance in young children. *Journal of Experimental Child Psychology*, *102*(2), 150–166. https://doi.org/10.1016/j.jecp.2008.02.002

Klein, T., Gelderblom, G. J., de Witte, L., & Vanstipelen, S. (2011). Evaluation of short term effects of the IROMEC robotic toy for children with developmental disabilities. Paper presented at the IEEE International Conference on Rehabilitation Robotics. https://doi.org/10.1109/ICORR.2011.5975406

Kronreif, G., Prazak, B., Kornfeld, M., Hochgatterer, A., & Fürst, M. (2007). Robot assistant "PlayROB" - User trials and results. Paper presented at the 16th IEEE International Workshop on Robot and Human Interactive Communication (pp. 113–117). https://doi.org/10.1109/ROMAN.2007.4415063

Li, S., Zhang, X., Kim, F. J., Donalisio da Silva, R., Gustafson, D., & Molina, W. R. (2015). Attention-aware robotic laparoscope based on fuzzy interpretation of eye-Gaze patterns. *Journal of Medical Devices*, *9*(4), 041007. https://doi.org/10.1115/1.4030608

Majaranta, P., & Bulling, A. (2014). Eye tracking and eye-based human-computer interaction. In S. Fairclough & K. Gilleade (Eds.), *Advances in Physiological Computing* (pp. 39–65). London: Springer-verlag. https://doi.org/10.1007/978-1-4471-6392-3

Man, D. W. K., & Wong, M. L. (2007). Evaluation of Computer-Access Solutions for Students With Quadriplegic Athetoid Cerebral Palsy. *The American Journal of Occupational Therapy*, *61*(3), 355–364.

Marchal-Crespo, L., Furumasu, J., & Reinkensmeyer, D. J. (2010). A robotic wheelchair trainer: design overview and a feasibility study. *Journal of Neuroengineering and Rehabilitation*, *7*(40), 1–12. https://doi.org/10.1186/1743-0003-7-40

Marti, P., & Iacono, I. (2011). Learning through play with a robot companion. *Everyday Technology for Independence and Care*, *29*, 526–533. https://doi.org/10.3233/978-1-60750-814-4-526

Miguel Cruz, A., Ríos Rincón, A. M., Rodríguez Dueñas, W. R., Quiroga Torres, D. A., & Bohórquez-Heredia, A. F. (2017). What does the literature say about using robots on children with disabilities? *Disability and Rehabilitation: Assistive Technology*, *3107*(July), 1–12. https://doi.org/10.1080/17483107.2017.1318308

Missiuna, C., & Pollock, N. (1991). Play deprivation in children with physical disabilities: The role of the occupational therapist in preventing secondary disability. *The American Journal of Occupational Therapy*, *45*(10), 882–888. https://doi.org/10.5014/ajot.45.10.882

Møllenbach, E., Hansen, J. P., & Lillholm, M. (2013). Eye movements in gaze interaction. *Journal of Eye Movement Research*, *6*(2), 1–15. https://doi.org/10.16910/jemr.6.2.1

Najafi, M., Sharifi, M., Adams, K., & Tavakoli, M. (2017). Robotic assistance for children with cerebral palsy based on learning from tele-cooperative demonstration. *International Journal of Intelligent Robotics and Applications*, *1*(1), 43–54. https://doi.org/10.1007/s41315-016-0006-2

Nakanishi, M., & Mitsukura, Y. (2013). Wheelchair control system by using electrooculogram signal processing. Paper presented at the Korea-Japan Joint Workshop on Frontiers of Computer Vision (pp. 137–142). https://doi.org/10.1109/FCV.2013.6485476

Novak, D., Omlin, X., Leins-Hess, R., & Riener, R. (2013). Predicting targets of human reaching motions using different sensing technologies. *IEEE Transactions on Biomedical Engineering*, *60*(9), 2645–2654. https://doi.org/10.1109/TBME.2013.2262455

Okimoto, A. M., Bundy, A., & Hanzlik, J. (2000). Playfulness in children with and without disability: Measurement and intervention. *American Journal of Occupational Therapy*, *54*, 73–82. https://doi.org/10.5014/ajot.54.1.73

Palisano, R., Rosenbaum, P., Bartlett, D., & Livingston, M. (2007). Gross motor function classification system expanded and revised. *Retrieved from Www.Canchild.Ca*.

Păsărică, A., Andruseac, G. G., Adochiei, I., Rotariu, C., Costin, H., & Adochiei, F. (2016). Remote control of an autonomous robotic platform based on eye tracking. *Advances in Electrical and Computer Engineering*, *16*(4), 95–100.

Poletz, L., Encarnação, P., Adams, K., & Cook, A. (2010). Robot skills and cognitive performance of preschool children. *Technology and Disability*, *22*(3), 117–126. https://doi.org/10.3233/TAD-2010-0296

Prazak, B., Kronreif, G., Hochgatterer, A., & Fürst, M. (2004). A toy robot for physically disabled children. *Technology and Disability*, *16*(3), 131–136.

Ríos-Rincón, A. M., Adams, K., Magill-Evans, J., & Cook, A. (2016). Playfulness in children with limited motor abilities when using a robot. *Physical & Occupational Therapy In Pediatrics*, *36*(3), 232–246. https://doi.org/10.3109/01942638.2015.1076559

Robins, B., Dautenhahn, K., Ferrari, E., Kronreif, G., Prazak-Aram, B., Marti, P., … Laudanna, E. (2012). Scenarios of robot-assisted play for children with cognitive and physical disabilities. *Interaction Studies*, *13*(2), 189–234. https://doi.org/10.1075/is.13.2.03rob

Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., … McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. *Computer Graphics Forum*, *00*, 1–28. https://doi.org/10.1111/cgf.12603

Sakamaki, I., Adams, K., Gomez Medina, M. F., Castellanos Cruz, J. L., Jafari, N., Tavakoli, M., & Janz, H. (2017). Preliminary testing by adults of a haptics-assisted robot platform designed for children with physical impairments to access play. *Assistive Technology*. https://doi.org/10.1080/10400435.2017.1318974

Schulmeister, J., Wiberg, C., Adams, K., Harbottle, N., & Cook, A. (2006). Robot assisted play for children with disabilities. Paper presented at the 29th Annual RESNA Conference Proceedings (pp. 1–5).

Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. Paper presented at the SIGCHI conference on Human factors in computing systems (pp. 281–288). https://doi.org/10.1145/332040.332445

Stagnitti, K., & Unsworth, C. (2000). The importance of pretend play in child development: An occupational therapy perspective. *British Journal of Occupational Therapy*, *63*(3), 121–127. https://doi.org/10.1177/030802260006300306

Tanta, K. J., & Knox, S. H. (2014). Play. In J. Case-smith & J. C. O'Brien (Eds.), *Occupational Therapy for Children and Adolescents* (7th ed., pp. 483–495). St. Louis, USA: Elsevier.

Udhaya Kumar, S., & Vinod, V. M. (2015). EOG based wheelchair control for quadriplegics. Paper presented at the ICIIECS - EEE International Conference on Innovations in Information, Embedded and Communication Systems. https://doi.org/10.1109/ICIIECS.2015.7193165

United Nations. (1989). *Convention on the Rights of the Child*. Retrieved from https://www.unicef.org/crc/

van den Heuvel, R. J. F., Lexis, M. A. S., Gelderblom, G. J., Jansens, R. M. L., & de Witte, L. P. (2015). Robots and ICT to support play in children with severe physical disabilities: a systematic review. *Disability and Rehabilitation: Assistive Technology*, 1–14. https://doi.org/10.3109/17483107.2015.1079268

van Hasselt, H. (2010). Double Q-learning. In *Advances in Neural Information Processing Systems*. Vancouver, BC; Canada.

Wästlund, E., Sponseller, K., Pettersson, O., & Bared, A. (2015). Evaluating gaze-driven power wheelchair with navigation support for persons with disabilities. *Journal of Rehabilitation Research and Development*, *52*(7), 815–826. https://doi.org/10.1682/JRRD.2014.10.0228

Watson, G. S., Papelis, Y. E., & Hicks, K. C. (2016). 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)*.

Wijesoma, W. S., Kang Say Wee, Ong Choon Wee, Balasuriya, A. P., Koh Tong San, & Kow Kay Soon. (2005). EOG based control of mobile assistive platforms for the severely disabled. Paper presented at the IEEE International Conference on Robotics and Biomimetics - ROBIO, 490–494. https://doi.org/10.1109/ROBIO.2005.246316

World Health Organization. (2007). *International Classification of Functioning and Disabilities, Children and Youth version (ICF_CY)*. Geneva: World Health Organization.