Leveraging Human Intuition for Learning Optimal Grasps for Objects

- Human-Robot Collaboration -

Master's Thesis

Christian Winther Rønnest

Aalborg University
Robotics

P10 Master's Thesis project

M.Sc. Robotics, Department of Electronic Systems

Fredrik Bajers Vej 7B, 9220 Aalborg Ø, Denmark

This report is set in palatino font using LaTeX Overleaf

Published June 1, 2023

AALBORG UNIVERSITY
STUDENT REPORT

Robotics
Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7B
9220 Aalborg Ø
https://www.robotics.aau.dk

Title:
Leveraging Human Intuition for Learning Optimal Grasps for Objects

Theme:
Human-Robot Collaboration

Project Period:
Spring Semester 2023

Project Group:
1059A

Participant(s):
Christian Winther Rønnest

Supervisor(s):
Dimitris Chrysostomou
Strahinja Dosen

Copies: 1

Page Numbers: 47

Date of Completion:
June 1, 2023

Abstract:

Using collaborative robots in a workshop environment can alleviate strains on workers and increase productivity, but improvements are still required for workers to want to work with them.
Pick-and-place tasks are one of the most widespread applications for robots, but they are still not as efficient as human workers when it comes to selecting grasps for different objects. In this thesis, a method is presented that leverages the human worker's intuition of optimal grasps to teach the robot which grasp to use for different objects.
A Convolutional Neural Network has been developed to classify between 6 different grasps with an accuracy of 87.22%, while objects are recognised using ORB features with an accuracy of 56.61%.
The method has been tested with 9 test subjects using a UR3 robot, and their feedback of the system in regards to comfort and trust in the implementation resulted in a score of 5.89 on a 7-point Likert scale, showing that there is an interest in collaborating with a robot in this manner.

Contents

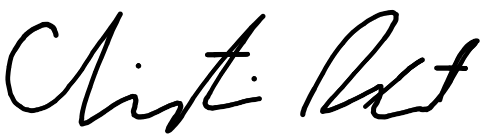
Preface

This thesis presents the work done on the 4th semester of the MSc. Robotics education at Aalborg University. While the goal for this project was to have a fully working implementation in regards to grasp classification, EMG control, robot communication and object recognition, the scope of the project had to be reduced over time due to unforeseen complications.

This reduced project, however, still shows the core concept that was explored and lays a foundation for improving the Human-Robot Collaboration in regards to a human user teaching optimal grasps to a robot system.

The author would like to thank the Supervisors for their valuable insights and suggestions, which helped improving the final implementation greatly compared to the initial solution.

Aalborg University, June 1, 2023



Christian Winther Rønnest
<cwra18@student.aau.dk>

Chapter 1

Introduction

The use of robots in industries provide a range of benefits compared to using human workers [42]. The company-oriented benefits include things like reduced production time and cost, higher precision, and extended production time, as a robot manipulator provide fast, precise movements in a constant cycle. From a worker perspective, the robots take over most of the monotonous and dangerous tasks that may put the worker's health at risk.

Using industrial robots may increase the efficiency and turnover for the company, but each robot requires safety features and fencing that meets current safety standards in order to be applied in the workshop [36]. This means that each robot cell take up some area of floor space that could have been used for other equipment or stations instead, while workers may become worried that their jobs will be replaced by the robots [19].

As a solution to these problems, collaborative robots (cobots) can be used in the workshop [1], as long as the tasks are not heavy duty. These robots require less space and safety equipment, freeing up more space for other processes, and allowing humans to work alongside with the robots.

This opens for a new kind of work condition in that the human workers need to collaborate with the robots in order to ensure the production flow. This collaboration can be in the form of the robot performing some task and sending the product to a worker for further processing, much like an industrial robot, or the robot may be used as a tool by the human worker, such as in a collaborative lifting task, that may alleviate stresses on the human worker as it helps carrying heavy objects.

In both cases, object manipulation is often required, and grasping objects are one of the most widespread tasks for robots to perform, even though the robot system does not have any intuition on how to grasp an object optimally.

This thesis presents a method for a human user to teach a robot system which grasp to use for a certain object using an EMG grasp recognition method and a visual object recognition method.

Chapter 2

Problem Analysis

The implementation of this thesis focuses on robot manipulation of objects in collaboration with a human user. This area of research is an extension of the notion of collaborative robots, with the ultimate goal of humans and robots working together and alongside each other. This thesis investigates the human and robot collaboration in an industrial object manipulation task, with the human user controlling the interaction. Before the implementation is presented, the current state of research within Human-Robot Collaboration and robot grasping is investigated.

2.1 Human-Robot Collaboration

Collaboration between a human user and a robot is a field of ongoing research. In essence, Human-Robot Collaboration (HRC) is any task where a human user and a robot works together to reach a certain goal. With the continuous growth of the industrial automation market [35], the number of processes and tasks where humans and robots interact with each other increases. This allows for the execution of a wide variety of tasks, such as drilling, lifting, sawing, and hand-over. Incorporating a robot to do these tasks allows for increased accuracy during execution and better ergonomics for the human workers, as the handling of heavy objects and/or tools is greatly decreased.

For successful HRC, though, the implementation of a cobot alone is not enough: The robot needs to be able to react according to the human user in order for the collaboration to be successful. This requires some sort of communication between the participants. For human-human collaboration, this can be done using gestural, tactile, and verbal cues, which the other participant can use to determine the other's intention.

Thus, in order to obtain proper HRC, external devices are required. This can be in the form of cameras, Inertial Measurement Units (IMU's), or biological signals from the user, such as ElectroMyoGraphy (EMG).

Gowtham et al. [17] implemented control of a 5 Degree-of-Freedom (DoF) robot manipulator using EMG signals and accelerometers for detecting arm postures and gestures, while Kaplanoglu et al. [24] implemented a gesture-based EMG control of a 7 DoF collaborative robot. EMG was also used by Zhang et al. [53] along with an IMU to make a robot follow the human operator during a collaborative sawing task, and Lin and Peng [30] implemented a teleoperation approach with the equipment, while Chico et al. [6] utilised the same equipment to control the position and orientation of a robot manipulator in another approach.

Treussart et al. [47] used EMG signals to control an exoskeleton for assisting in carrying heavy loads. DelPreto and Rus [9] created a collaborative lifting framework using only the EMG signals from the biceps and triceps muscles to control a robot manipulator, while Bednarczyk et al. [3] used EMG to enable the robot to distinguish forces from the human operator from other external forces, and Li et al. [29] used EMG to enable the human user to regulate the robot's compliance and force regulation.

A method for teleoperating a robot manipulator based solely on EMG signals were devised by Dwivedi et al. [11], and combining fiducial markers with EMG signals, Dwiwedi et al. [12] also created a control system for teleoperation of a dexterous robotic arm hand system. Boru and Erin [5] used EMG and a Leap Motion Sensor for teleoperating a robot manipulator.

Another approach to HRC has been implemented by Wang et al. [50], that used EMG and IMU signals for the robot to determine the best approach in a human-to-robot and robot-to-human hand-over.

HRC has already been applied to some degree, for example in the form of active prostheses, where there is some form of autonomy in the prosthesis, so the human user does not have to control every movement of the prosthesis.

This is for example seen in the work of Satya Shree et al. [43] where they used EMG signals to control a bionic hand for different types of grasps. Perera et al. [33] used EMG and vision to increase the probability of correct grasp classification depending on the object to be grasped, and Zandigohar et al. [52] used the same method, including eye tracking as well.

For lower limb prostheses, the robot system can be trained to detect direction intentions and robustness during walking, as implemented by Khiabani and Ahmadi [26], by applying LDA to EMG signals of the subject's lower limb muscles. In the work done by Clark and Amor [7], the HRC is implemented in an imitation learning strategy used to predict biomechanical variables in order to determine the correct strategy for minimising stress on the body of the user. Karulkar [25] developed a predictive two-stage estimation framework relating the human user's gait speed and features, which can be used in a lower limb exoskeleton to realise change in direction or speed.

Another central aspect in HRC is the overall control framework for the system. The overall control of the system lies with the human user. The collaboration can vary from the robot passively following or responding to commands from the human user, to the robot actively predicting the intention of the user. These collaboration approaches have each been implemented in different ways:

In some papers, the robot is directly controlled by the human user, such as in teleoperation as implemented in [11, 12], where the user's arm configuration determine the configuration of the robot manipulator, meaning that the robot is effectively imitating the user's movements. Another approach to teleoperation is done by [17, 24, 6], where arm and hand gestures directly control the robot's movements.

In other papers, the control is more indirect, such as in [3, 53], where the robot is compliant to forces exerted on it by the human user, based on the measured EMG signals from the user's upper limb, or as in [50], where the signals are used to provide information about the object and the cues given by the human operator. For lower limb prostheses and exoskeletons, the control is usually based on intention prediction, based on measurements of the human joints or of biological signals like EMG.

Most of the latest research done in HRC thus include the use of biological signals from the human user, with EMG being particularly widespread.

2.2 Myoelectric control

As mentioned on several occasions in Section 2.1, EMG signals have been widely used as a means of Human-Robot Interaction, and, by extension, Human-Robot Collaboration. By using myoelectric signals, the interaction between the human user and a robot can appear seamless, similar to that of Human-Human Interaction.

EMG signals are electric impulses that can be measured from muscle tissue by the use of electrodes [31]. In essence, the control of the human body is done by electric signals sent from the brain to the relevant skeletal muscle groups through the nervous system. By using electrodes, these signals can be measured and processed by devices, which enables the use of these signals for inputs to control a device such as a prosthesis, robot arm, or a mobile robot.

The acquisition of myoelectric signals can be divided into two approaches: Surface EMG (sEMG) and Intramuscular EMG (iEMG). Both approaches have both pros and cons, so the determination of which method to use for a certain application is based on an evaluation of these.

In order to acquire myoelectric signals of the best quality, iEMG is used. This approach involves inserting an electrode under the skin of the subject and directly onto or into the muscle tissue, from which signals are wanted [48]. This secures the electrode a stable connection to the specific muscle and better measurement of the electric signals. Thus, iEMG is also suitable for applications, where specific muscles are of interest, as the electrodes can be inserted as deeply into the body as required [48]. This is especially useful for determining the EMG signals produced from the different muscles responsible for the extension and contraction of the individual fingers on the human hand, as they are placed deep inside the forearm [22, 23].

The other approach, sEMG, is non-invasive, allowing for reading myoelectric signals from the body without needing to inject electrodes into the muscle tissue. As such, sEMG has been used extensively in research and in commercial use, as the implementation is not as extensive as that of iEMG [8]. Because of this, sEMG has already been adapted to personal and industrial use, exemplified by prostheses and Human-Robot Collaboration.

However, sEMG control is not yet widespread, because of the many ways the measurements can be impacted, compared to that of iEMG. The electrodes used for sEMG is placed on the surface of the human body, i.e. the skin, which disturbs the measurement of the myoelectric signals from the muscles. [31, 48]

Additionally, the skin tissue is elastic, allowing it to stretch and contract depending on the motion of the body. This, however, displaces the surface electrodes placed on the skin, resulting in disturbance of the signals due to, for example, motion artifacts. [48]

Because the measurements are made on the surface of the body, the most significant signals measured are those of the muscles right beneath the skin. As such, signals measured from muscles located deeper in the body, such as, for example, Flexor Digitorum Profundus and Flexor Pollicis Longus, are not as distinguishable, meaning that it is harder to discern individual finger movements using sEMG signals than, for example, wrist movements. [31]

In the context of HRC, the use of sEMG is particularly widespread, however, as it is the primary means of acquiring EMG signals in most of the latest research. The utilisation of these signals, however, are varying greatly:

The implementation by [17] was done on an Arduino UNO controller using EMG signals and two accelerometers. The control recognised wrist rotation, left, right, up, and down motions, and the accelerometers were used to determine the X and Y directions of the movement, determining the position of the robot manipulator.

In [24], 7 hand gestures were classified based on EMG signals using a Finite State Machine classification method. This allowed for the control of a 7 DoF cobot, with each gesture representing different movements for different parts of the robot, such as gripper open/close or left/right arm position.

In [53], the forces of the human operator were extracted from the EMG signals of the upper part of the upper limb, which were used in a Long Short-Term Memory (LSTM) neural network to estimate the three-dimensional forces of the upper limb, which were used to make the robot compliant while the human user pulled or pushed on the end-effector.

The solution of [30] used 3 IMUs for human arm posture recognition on the shoulder, elbow, and wrist, while the EMG signals were used for hand gesture recognition to control a 3 finger robotic gripper with the gestures open, close, and pinch. The classification was done using a Cascade SVM, yielding a recognition rate between the four gestures of 99%.

Alternatively, the same equipment could be used and implemented as in [6], where the hand gestures were used in another way than to recognise a grasp. The fist gesture was used to activate and deactivate a paint gun at the virtual robot's end-effector, while the wrist flexion and extension were used to switch between 3 different orientation references on the robot, using the pinch gesture to confirm the reference selection. The EMG signals were classified by an SVM with a mean performance of 81.3%. The IMU was used to control the end-effector position, with roll, pitch, and yaw controlling movement along the X, Z, and Y axis, respectively.

The EMG signals were implemented as assistance in [47], as a means of gravity compensation in an exoskeleton to assist in carrying an unknown load. Signals from the biceps and triceps muscles were used to estimate the intensities of the muscles, and whether the user was flexing or extending the elbow. The implementation reached the ideal torque of the exoskeleton with an error of 0.038 ± 0.012 N m.

In [9], EMG signals from the biceps and triceps of the human arm were used for controlling a robot in a collaborative lifting task by classifying the hand height of the user and the muscle contractions. A feedforward neural network was used to perform the classification, resulting in an overall 68.8% classification accuracy during tests.

EMG signals from 8 muscles on the upper arm region were used in [3] to calculate the forces and wrench exerted on the end-effector of a robot manipulator by the human user, allowing the user to move around the end-effector.

In [29], the upper limb EMG signals were used in the control system to determine whether the user wanted to move the robot end effector around. Based on a use case of ultrasound scanning, the robot regulated the pressure on the tissue on its own, allowing the user to focus on other tasks, but the user may adjust the end effector by exceeding a force threshold in the control system, making the robot compliant.

The method for teleoperation implemented in [11] was based on a Random Forest (RF) classification of EMG signals of the upper arm, chest and back muscles, that was converted into the goal pose of a robot manipulator. The implementation yielded a correlation between the human motion and the pose decoding of 92.8%, with a Normalised Mean Square Error (NMSE) of 73.4%. A practical test of cleaning a whiteboard showed a task completion of 90.6%.

In [12], the EMG signals were used to classify arm and hand postures and gestures, while fiducial markers placed on the shoulder and hand of the human subject were used to determine the trajectory of the robot based on tracking from a head-mounted camera. The implementation included the identification of three different grasps as well as muscle co-contraction and rest used for controlling the grasp of the dexterous gripper mounted on the robot manipulator. Three classifiers were tested for the study: Linear Discriminant Analysis (LDA), Random Forest (RF), and Support Vector Machine (SVM), with the RF providing the highest prediction accuracy of $> 97\%$, compared to the others yielding an accuracy of $> 93\%$.

In [5], a Leap Motion Sensor was used for controlling the robot arm positioning and trajectories, while EMG was used to control the robot tool. Using the infrared cameras of the Leap Motion Sensor, the finger positions were measured, from which the motions could be calculated and used for controlling the positioning of the robot. The classification of EMG data was tested on three different classifiers: SVM, RF, and K-Nearest Neighbour (KNN), with RF proving the most efficient classifier.

In [50], the EMG and IMU signals from a human user's forearm were used to represent high-level intentions in the HRC, so the robot was not directly controlled by the user. Six different high-level commands were defined by EMG signals, while the IMU signals were used as triggers for initiating the handover. To determine the intention of the user, Hidden Markov Models (HMM) were used, and a Directed Acyclic Graph-SVM (DAG-SVM) classification was implemented to determine the attributes of the object to be handed over, used for the robot to actively decide to reject or accept the object.

For controlling a bionic arm, the solution done in [43] used EMG signals from 8 electrodes and trained a computer to recognise 4 hand gestures: Hand open, hand close, circular grasp, and fine squeeze. SVM, KNN, and a combination of SVM and KNN were used for classification, with the SVM-KNN model giving the best performance of 96.33% accuracy.

The work in [33] combined vision and EMG signals for determining the correct grasp of an object. The EMG signals were classified using and Artificial Neural Network (ANN), while a pre-trained You Only Look Once version 3 (YOLOv3) neural network was used for object detection. Object recognition was used for determining grasping probabilities of 6 grasp patterns, based on a survey done on human subjects, allowing the system to rule out false positives in the EMG grasp classification. The implementation achieved a mean True Positive Rate (mTPR) of 73.1%.

In [52], vision and EMG were used together as well, but with a slightly different approach: The EMG-based grasp classification was based on the extra-trees method, and the vision-based grasp classification was based on YOLOv4. The detected object located closest to the gaze of the user was the target object of the vision-based grasp classification system. The probabilities of predetermined grasps were determined for both EMG and vision, as well as fused, yielding an accuracy of 41.85%, 81.46%, and 92.93%, respectively.

The implementation of the EMG detection varies throughout the research in that the electrodes may be placed on certain areas of the human user's body [9, 53, 12], or a series of electrodes may be arranged on a wearable device [17, 24, 6, 5, 50].

A wearable device is practical for commercial use, as it requires minimal time to set up, compared to the separate electrodes that have to be placed on selected parts of the body. Because of this, the wearable device might more likely be accepted as a workplace tool due to its simplicity and user-friendliness.

2.3 Robot grasping

Grasping is one of the basic types of manipulation a robot can do, and it is used extensively throughout industries. Pick and place tasks are one of the most common types of tasks that a robot manipulator can perform, and often with greater speed and accuracy compared to a human worker. Additionally, industrial robot manipulators can carry larger loads than humans, allowing for the utilisation of robots in handling heavy or large objects more efficiently compared to human workers [38].

The task of grasping an object is however more complex than it may appear at first glance. While it is simple to program a robot to move its end-effector to a certain point and close its attached gripper, this implementation only allows for picking up an object at that specific location. This approach may be suitable for industrial tasks where the robot works within a fixed cell, but this approach makes the robot inflexible to change in the environment. In order for the robot to be able to adjust to its environment, the robot needs to be autonomous [37].

For a successful HRC, the robot should be able to adapt to a dynamic environment, which includes grabbing objects that are not placed at a fixed position in the robot's workspace. The robot should thus be able to grab an object with intuition similar to a human operator's.

In order for a robot to successfully grasp objects in a 'human' manner, two factors need to be considered: The location of the object, and the optimal grasp for picking up this object.

Locating objects to be grasped is an intuitive task for humans, but implementing it in a robot system efficiently has been, and still is, subject to research [15]. Humans rely on a range of modalities for locating objects in the environment, such as vision, proprioception, and haptic perception. These same modalities can be mimicked in a robotic system. Proprioception is integral for the robot's operation and is already implemented on a robot manipulator. However, if the robot should be able to locate an object, proprioception cannot be used alone, and thus, vision and/or haptic perception modalities need to be implemented.

Vision has been the most used modality for object localisation in robotics, as it can provide much information about the object and the environment, such as shapes, sizes, distances, and locations.

In a review paper done by Fan et al. [15], the state of the art work for object localisation is investigated, with the use of Convolutional Neural Networks (CNNs) being prevalent in newer papers, but the use of conventional computer vision strategies such as background subtraction and shapes and edges are still applied in industrial settings, due to the lack of large datasets and Graphical Processing Units (GPUs) for effective implementation of CNNs.

Shahid et al. [44] had a robot approach and grasp an object correctly based on Reinforcement Learning (RL) done in a simulation environment. The robot grasped the object successfully in every trial with the object at a fixed position, and with extra training, the robot successfully grasped the object at new locations in every trial.

Fang et al. [16] also used vision for object localisation, where a stereo camera was used. By colour segmenting the shoulder, wrist, and gripper of the robot, as well as the target object, the areas were located on the image plane, and the stereo camera was then used to calculate the object and joint locations within the workspace. These locations are used in two neural networks (NNs) for the robot to reach the object. The first NN is based on Radial Basis Functions (RBFs) and is used for rough reaching movements toward the object location. When the gripper gets close to the object, a Brain Emotional Nesting Network (BENN) is used to approach the object.

Once the object has been localised, the optimal grasping strategy needs to be found in order for the robot to successfully grasp the object. This entails determining the optimal direction to approach the object and determining the best grasp for the object, if the end effector is equipped with a $>$ 2-fingered gripper, or the best contact points for gripping the object, if the gripper has 2 fingers or rely on, for example, vacuum for gripping.

Shi et al. [45] used a CNN to classify different daily objects into 4 different grasp patterns for a dexterous hand prosthesis by using Red-Green-Blue-Depth (RGB-D) images, reaching a classification accuracy of 98%.

Berscheid et al. [4] implemented a Fully Convolutional Neural Network (FCNN) combined with a model-based controller for learning to estimate the 6 DoFs for the optimal grasp policy for objects in a bin-picking task. The solution had a success rate of 92% for picking known objects in a dense clutter.

It may also be relevant for the robot to adjust the force of the grasp so as to not destroy or deform a fragile or delicate object. Such an implementation has been made by Wen et al. [51], where a NN was used to estimate the forces of a human user based on EMG signals, which was then used on a robot gripper to grasp delicate objects with the estimated force. Every grasp were successful during the tests.

Tieck et al. [46] used a Spiking Neural Network (SNN) to mimic the compliance control of an anthropomorphic robot hand to that of a human hand when handling delicate objects using sphere, cylinder, and pinch grasps. The implementation combined two control loops, with one controlling the motion primitives of the robot hand, and the other being the compliance controller to adjust the grasp force.

Autonomous grasping is often difficult to implement efficiently, especially if it should mimic a human user. Therefore, in a HRC context, it is often practical to include the human in the control loop, as this reduces the complexity of the robot program. The simplest approach is to build the system around teleoperation, where the human user controls the robot manipulator's movements, as done by [5, 12]. Another approach is to have the human user provide the grasp type for the robot using EMG signals, as in [43, 33], while the object localisation can be done by detecting the human user's gaze, as presented in [52].

One could also implement a solution similar to that of Kim et al. [27] that created a pipeline where one CNN was used to locate a hand in real-time images, while another CNN was used to classify the grasp of the hand. This could remove the need to use any biological signals in the control system, which would enable the HRC to resemble more closely the collaboration between humans.

Chapter 3

Project Scope

Based on the findings in Chapter 2, the concept of myoelectric control in HRC is widely used. This may be in combination with other external sensors, such as cameras, in order to provide an efficient control. Additionally, the term of Human-Robot Collaboration can be used for two implementations: A robot manipulator performing tasks alongside a human user; and an active prosthesis assisting in the tasks of the human user.

This thesis focuses on implementing a novel approach of HRC on a robot manipulator. By the help of a human user, the robot is to learn the optimal type of grasp for objects, which it can then use for autonomous manipulation or for cooperative lifting tasks with the human user.

The implementation thus have 2 states:

1. The learning state where the robot learns features of a new object and correlates the optimal grasp to it as provided by the human user.
2. The collaboration state where the robot grasps the objects with the learned grasps and moves it around based on the human user's commands.

This implementation is based on the argument that humans have a keen sense of determining the optimal grasp of an object, which is a major challenge for robots. This work seeks to utilise this ability for teaching the optimal grasps for a robot manipulator.

Because grasping is central in this implementation, forearm surface EMG signals are used as the primary control interface between the human user and the robot, as it provides information about the grasp configuration of the human hand.

In order for this implementation to be viable in a real world HRC scenario, the work in this thesis will be assessed based on the following requirements:

1. The implementation must have an EMG grasp recognition accuracy of at least 90%, based on the results presented by the State of the Art in Section 2.2
2. The implementation must be able to correctly recognise learned objects at least 95% of the time, as this will prove a viable base for future implementations

3. The implementation must be able to localise the object within a margin of 5% in order to ensure stable manipulation of the object
4. The implementation must be able to grasp an object without dropping it with a success rate of at least 90% to prove that the chosen grasp can be correctly applied in a real-world environment
5. The implementation must score at least 4.5 out of 7 on a Likert scale in terms of the perceived comfort and trust by human test subjects, as this is a good basis for determining its suitability for HRC

This implementation is developed focusing on a use case in an industrial environment. It is therefore meant to be used as an assistance to factory workers during lifting tasks, actively helping them. As such, the implementation has to be comfortable for human users, which is best measured in a test scenario and using a questionnaire with Likert scales for measurement.

A Likert scale offers a range of choices, typically between 5 and 7, for a given statement, allowing for a more nuanced answer to a statement rather than 'yes' or 'no' [2]. The scale can be formulated to measure a range of different attitudes such as agreement, frequency, acceptability, priority, and many others, spanning from one extreme to the other, such as an agree/disagree scale [49].

As such, this is a widely used method for quantifying otherwise qualitative data, which can be used to determine aspects such as the comfort of human users around a robot during HRC.

It must be noted, that in this project, the different grasps cannot be executed by the robot manipulator, as such dexterous grippers are not accessible for the project. As such, the physical grasping will not be a part of this implementation, and the choice of grasp will be simulated as a text display.

Chapter 4

Solution Concept

This chapter describes the concept of the implementation and provides an overview of the 2 states as well as how these states are implemented. A concept description of the solution's 2 states is presented in Section 4.1, giving an overview of the task approach and system architecture. For a successful HRC, the implementation consists of several parts and modalities, and the software and hardware used for the implementation are described in Sections 4.2 and 4.3.

4.1 Concept

Before giving a detailed description of the different parts used in the implementation in this thesis, an overview of the general execution of the two states is given. This provides a reference for how the different parts are used in the solution and how they interact with each other.

4.1.1 Learning state

For every object new to the robot, the learning state needs to be executed before the collaboration with the human user can begin. This essentially consists of the robot learning the correct grasp for a given object based on input from the human user. The architecture of the learning state is pictured in Figure 4.1.

Using a camera, the robot learns the type of the object by comparing the image to a library of reference images, and based on EMG signals from the forearm of the human user, the robot learns which grasp to pair with this object. This information is stored in a Comma-Separated Values (CSV) file, which serves as the grasp database for the robot in the collaboration state.

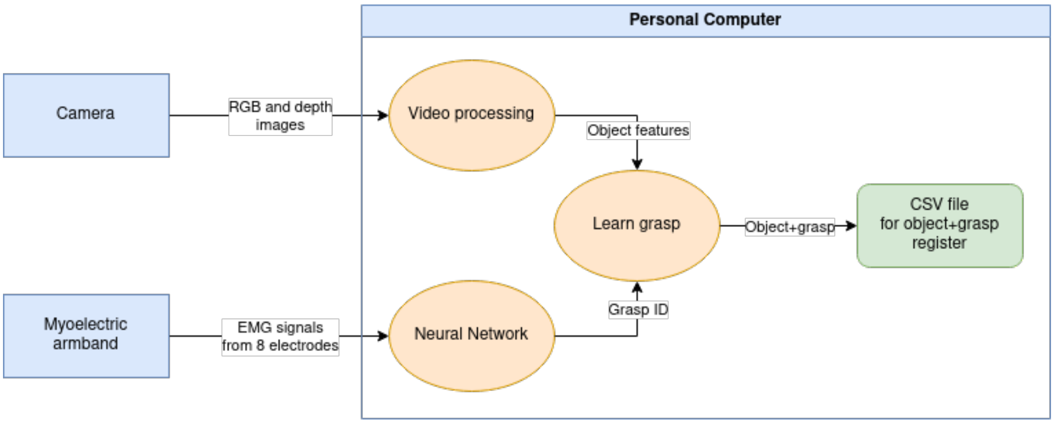


Figure 4.1: A diagram of the architecture of the learning state. The blue boxes represent hardware, the yellow ellipses represent programs, and the green rounded box represents a file.

4.1.2 Collaboration state

When the robot has learned how to grasp an object, the collaboration state can be used, the system architecture of which can be seen pictured in Figure 4.2.

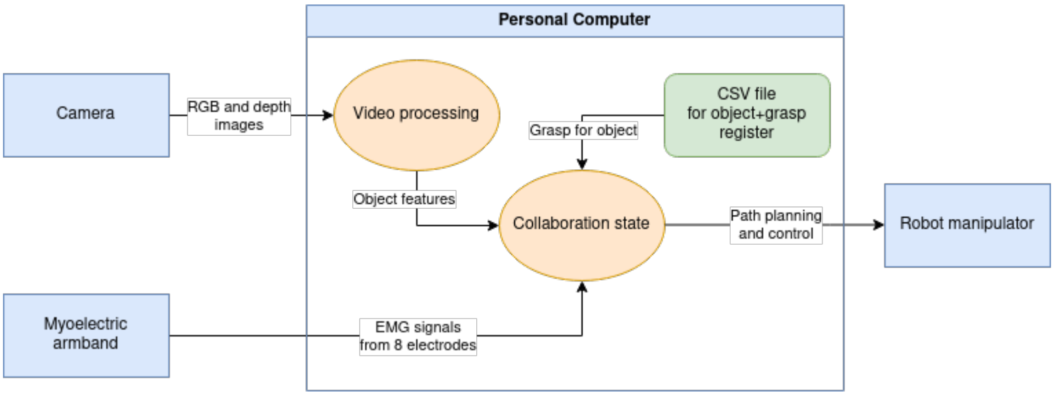


Figure 4.2: A diagram of the architecture of the collaboration state. The blue boxes represent hardware, the yellow ellipses represent programs, and the green rounded box represents a file.

The process consists of the robot detecting the object using vision and finding the corresponding object in the CSV file based on the object type, where the robot can find information about which grasp to use, as learned in the learning state.

The robot picks up the object with the learned grasp when the human user gives the cue based on EMG signals, and the user then gives another cue for the robot to proceed with the manipulation task.

When the object has been moved to the target place, the robot lets go of the object and moves to its home position, ready for the next task.

4.2 Software

The solution has been implemented using the Robot Operating System (ROS) [39], so the implementation consists of programs developed specifically for this solution, as well as freely accessible software for interfacing with the camera, the EMG armband, and the robot. This section gives an overview of the software used in this implementation.

4.2.1 Object detection

For both the learning and the collaboration states, information about the object is required. Features such as the location, size and shape of the object are relevant information for recognising the object and thus determine the optimal grasp.

For this project, a RealSense D455 camera [21] has been used, and as such, both Red-Green-Blue (RGB) images and depth images are used for extracting features.

Because the solution is implemented in ROS, the realsense2_camera ROS node [10] is used for interfacing with the camera. The object detection module then extracts a Binary Large Object (BLOB) from the RGB image after blurring and binarising the image, as well as doing some morphological opening. From this image, features such as the X and Y position of the object's centre are extracted and sent via an object message for further use, while the object recognition is done by extracting Oriented FAST and Rotated BRIEF (ORB) features from the processed images from the camera stream and comparing it with a processed reference image of objects to determine the most likely match. This recognition is used to determine the object type, which is also published in the object message.

The depth image is used to determine the distance between the camera and the object, which is useful for determining the height of the object as well as assessing how far down the robot should move in order to reach the object. This information is also sent via the object message.

4.2.2 EMG interface

The interface for collaboration between the robot system and a human user is done using a myoelectric armband. As such, a communication interface between the myoelectric armband and ROS is required for obtaining the myoelectric signals. For this project, the ros_myo package [34] is used that sets up a ROS node which allows for transferring the EMG signals as ROS messages.

The package can be used to obtain EMG measurements from the myoelectric armband, but one can also obtain IMU information from the armband, and predefined gestures and which arm is being used, if the armband is calibrated before use.

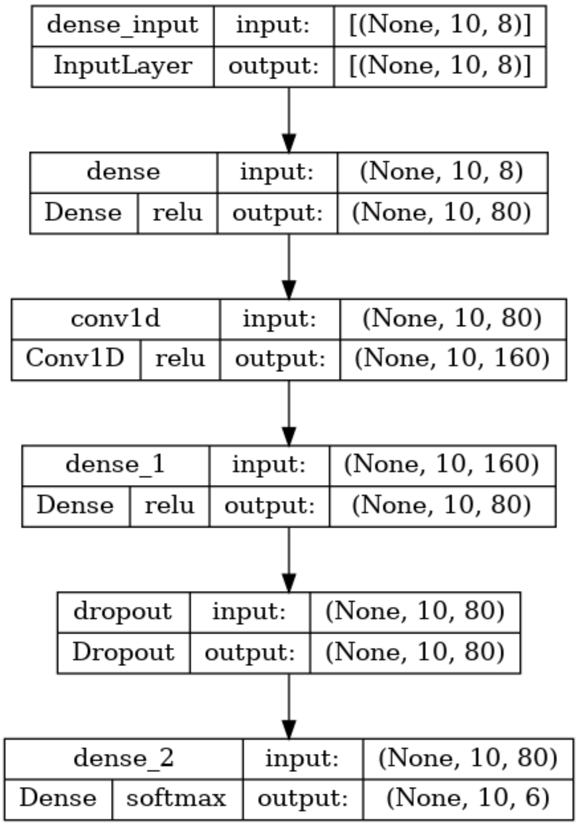


Figure 4.3: An overview of the developed neural network.

4.2.3 Grasp detection

In order for the implementation to be feasible, a method for grasp detection is needed. Based on the literature discussed in Chapter 2, Neural Networks are widely used for classifying grasps based on EMG signals and/or their features. As such, a novel Neural Network has been developed for this implementation to determine the grasps based on the EMG signals from the human user's forearm.

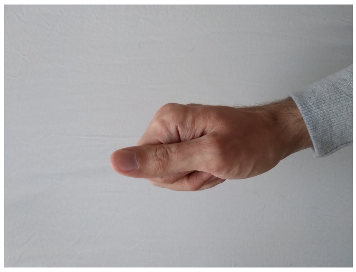
The developed network is a Convolutional Neural Network consisting of an input layer, and output layer, as well as 4 hidden layers. The input is an array of 10 means, each of 10 EMG measurements containing with measurements from 8 electrodes, while the output can be one of 6 possibilities, with 5 different grasps and a detection of no grasp.

More specifically, the network and this project is based on the classification of 5 different grasp: Power/fist grasp, key grasp, pinch grasp, tripod grasp, and spherical grasp. A visual depiction of these grasps can be seen in Figure 4.4.

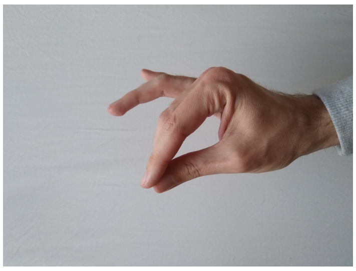
A graphical representation of the developed neural network can be seen in Figure 4.3. The layers of the neural network comprises of an initial dense layer of 80 nodes with ReLU activation, which feeds a convolutional layer of 160 nodes with a 3×3 kernel and a stride


(a) No grasp


(b) Power/fist grasp


(c) Key grasp


(d) Pinch grasp


(e) Tripod grasp

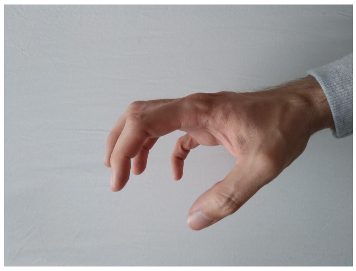

(f) Spherical grasp

Figure 4.4: The 6 classes used in the CNN as seen from above.

of 1, along with at ReLU activation function. The network then consists of a dense layer of 80 nodes with ReLU activation, followed by a dropout layer with a dropout rate of 30%. This is connected to the final dense layer of 6 nodes with the Softmax activation function.

The output is an array of grasp probabilities for each of the EMG measurements, so the most probable grasp for each of these measurements are used to determine the most probable grasp for the entire input array based on the grasp prediction most prevalent in the array of 10 means.

4.2.4 Learning grasps

Using the object detection and grasp detection, the learning state is implemented to combine the results of these two nodes. This node only runs once, as opposed to the others that run continuously. This is based on the assumption that the the learning is only done on 1 object at a time.

The program is visualised in the flow diagram in Figure 4.5. First, the program accesses a CSV file to determine the number of already learned object, so that the new object can be given a unique ID in the file.

The object message and grasp message are then received, and the script loops for 4 times in order to ensure that data from both messages are received, as well as the human user has determined the grasp for the object.

The object features and the grasp are then written to the CSV file, which serves as the database of the learned objects and grasps for the robot system.

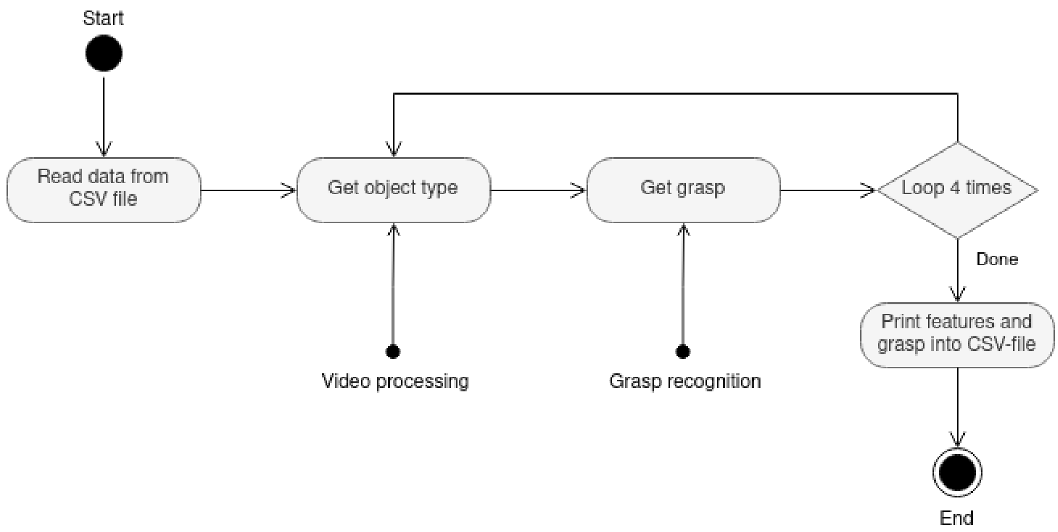


Figure 4.5: A flow diagram of the learning program. The video processing and grasp recognition nodes are inputs from the other processes required for the program to run.

4.2.5 UR driver for ROS

The solution is implemented using a UR3 robot from Universal Robots (UR) [41], so a ROS to UR interface is required for external control from the PC. This can be obtained using the Universal_Robots_ROS_Driver [14] installed on the PC and the externalcontrol-1.0.5.urcap [13] installed on the robot controller.

Once the communication has been established between the robot and the PC, the robot can be controlled using ROS with the ROS-industrial universal_robot package [32], which allows a ROS PC to control a UR manipulator using the MoveIt planner.

4.2.6 Collaboration

Once the system has learned features of an object along with the corresponding grasp as demonstrated by the human user, the collaboration program can be executed.

The robot receives data about the object on the table from the camera, providing the location and type of the object, and it also subscribes to the ros_myo node for the user to give prompts to the robot.

The collaboration program is visualised in a flow diagram in Figure 4.6. At the start of the program, all relevant controllers for the robot manipulator are set up.

When the object type is received from the object message from the camera, the program accesses the CSV-file containing information about the learned objects and the corresponding grasps to find an object of the received type. If the object type is found in the CSV file, the program finds the corresponding learned grasp for the object.

When the object has been identified, the robot waits for the human user to prompt

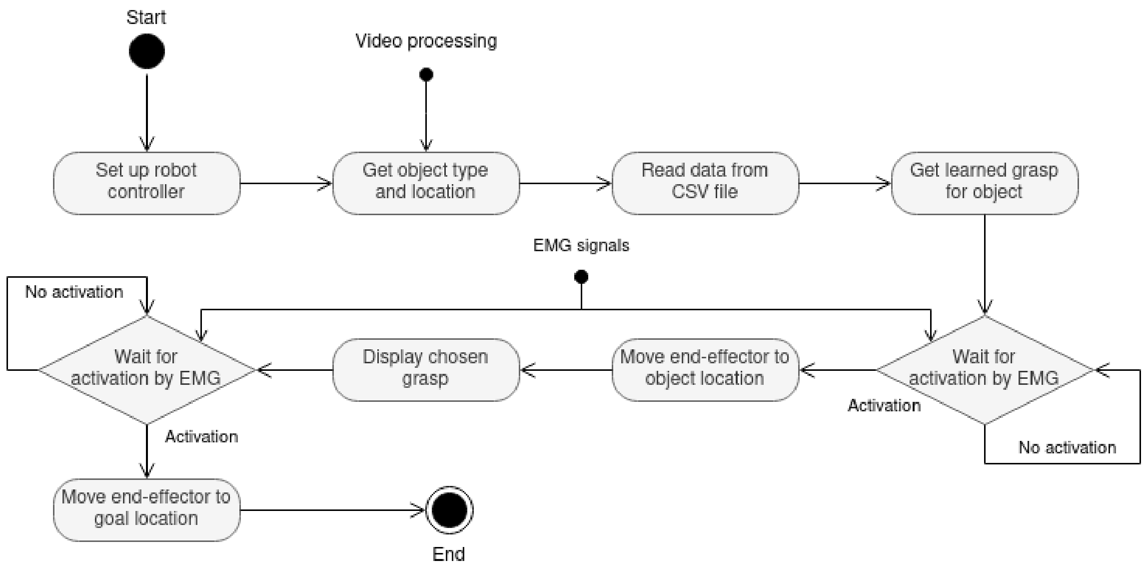


Figure 4.6: A flow diagram of the collaboration program. The video processing and EMG signals nodes are inputs from the other processes required for the program to run.

the robot to start the task execution. This is done using the EMG signals from the user's forearm.

The robot then moves to the object location and picks up the object with the corresponding grasp, and with a new prompt from the user, the robot proceeds with the task and moves the object to a predetermined goal position.

4.3 Hardware

The collaboration is implemented using a UR3 robot from Universal Robots, a myoelectric armband for collecting muscle signals from the user, and a camera for object recognition and localisation. The setup used for testing the implementation can be seen in Figure 4.7, using the hardware described in this section.

4.3.1 Camera

For object recognition, the Intel RealSense D455 depth camera [21] is used (see Figure 4.8). It is a camera equipped with an RGB module as well as a stereo depth module, which has a depth error of less than 2% at a distance of 4m. The camera is efficient in distance ranges between 0.6m and 6m.

The depth imaging has a field of view of $87° \times 58°$, with a maximum resolution of 1280×720 and a frame rate of up to 90 fps.

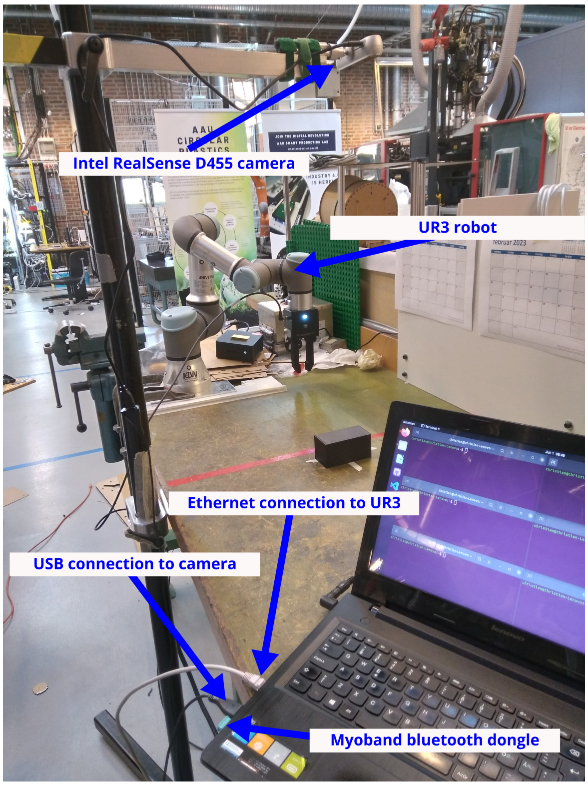


Figure 4.7: The test setup with the different hardware components and their connections to the PC.

Figure 4.8: The Intel RealSense D455 camera[20].

Figure 4.9: The Myo armband from Thalmic Labs [28].

The RGB imaging has a field of view of $90° \times 65°$, with a maximum resolution of 1280×800 and a frame rate of up to 30 fps.

For this application, an image resolution of 640×480 is used for both the colour and the depth imaging, as this provides faster processing and is sufficient for demonstration purposes.

The camera interfaces with a PC using a USB-C cable for both power supply as well as data transfer.

The camera is also equipped with an IMU, that allows for improved depth perception, which is useful for mobile robotic applications, but for this implementation, the camera is static, so the IMU data with not be relevant.

4.3.2 Myoelectric armband

For this project, the Myo armband [18] developed by Thalmic Labs is used, which comprises of 8 surface EMG electrodes joined together in a plastic casing (see Figure 4.9) for fitting around the forearm of a human user. The armband can be stretched and retained to fit on the forearm of a wide variety of users.

The Myo armband is furthermore equipped with a 9 DoF Inertial Measurement Unit (IMU), which comprises of a 3-axis gyroscope, 3-axis accelerometer, and a 3-axis magnetometer.

The signals are processed by the armband's ARM Cortex M4 processor, which publishes the EMG measurements via Bluetooth Low-Energy at a rate of 200Hz. The connection is done using a USB dongle which can be inserted in the PC.

The Myo Armband is powered by a rechargeable lithium-ion battery, which is recharged using a micro-USB cable.

4.3.3 Robot

The UR3 robot [41], seen i Figure 4.10 is a 6 DoF robot manipulator designed for use in collaborative applications. It comprises of 6 rotational joints and has a reach of 500mm with a repeatability of $\pm$0.1mm. It is the smallest model of manipulators from Universal Robots, weighing 11kg and having a payload of up to 3kg.

Figure 4.10: The UR3 robot from Universal Robots[40].

As such, the UR3 is not the type of robot that would be used in a practical application, given the small range and relatively low payload, but for demonstration and testing purposes, this manipulator is sufficient.

Chapter 5

Testing

In order to verify the developed solution described in Chapter 4, the different parts and the implementation as a whole needs to be tested. Therefore, 3 tests have been defined for this project, focusing on different aspects of the solution.

The test described in Section 5.1 is used to assess the performance of the developed neural network, in order to determine how well it performs based on some of the State of the Art implementations within the field of HRC described in Chapter 2.

In Section 5.2, the performance and stability of the developed object recognition implementation is assessed, while in Section 5.3, the entire solution is tested as a whole, with test subjects using the implementation and assessing their experience of it in terms of comfort and trust, which are central aspects to consider in HRC.

5.1 Grasp detection performance

The neural network has been trained on a dataset of $9,000$ averages of 10 EMG measurements, resulting in 150 measurements of each grasp classification from 6 individuals. The network has been trained for 40 epochs with a batch size of 30 and validated on a separate test dataset containing $1,800$ averages of 10 measurements. The test dataset contains 30 averages of each grasp classification from the same 6 individuals used in the training dataset.

The CNN is trained using the Nesterov-accelerated Adaptive Moment Estimation optimiser (Nadam), which is done using the Categorical Cross-Entropy loss function. In Figure 5.1, the loss convergence of the CNN can be seen for both the training set and the validation set.

Along with the loss function, the accuracy of the model is tracked as well, which can be seen in Figure 5.2 for both the training and the validation set.

The model finished the training with an accuracy of 93.39% on the training dataset and 84.22% on the validation set, with the corresponding losses being 0.1761 and 0.5981 on the training and validation datasets, respectively.

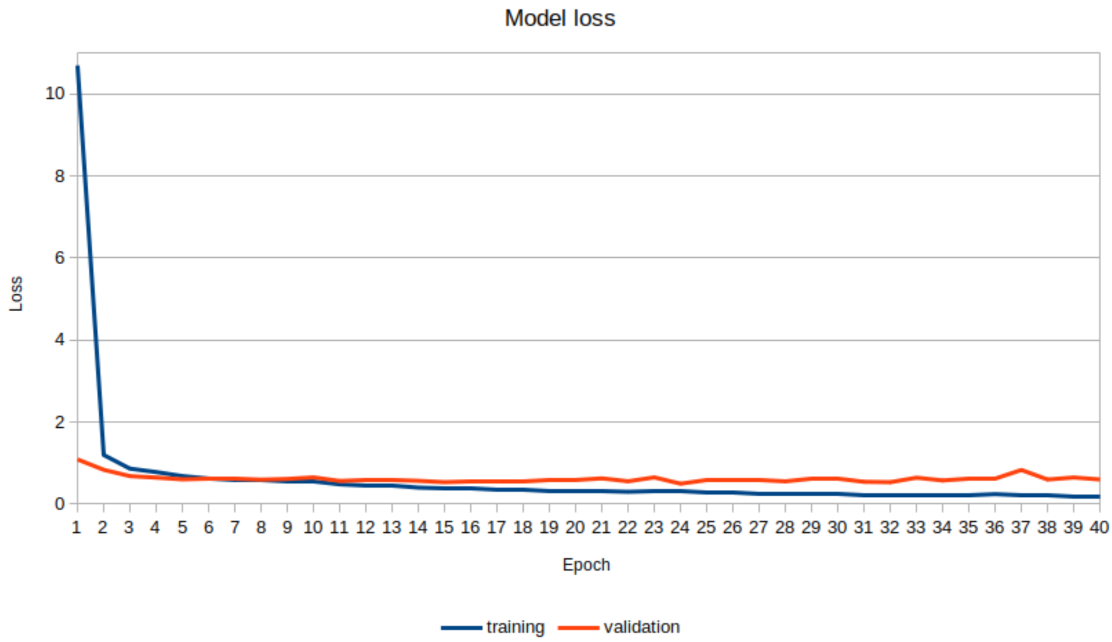


Figure 5.1: The training and validation loss for the CNN over the 40 epochs of training.

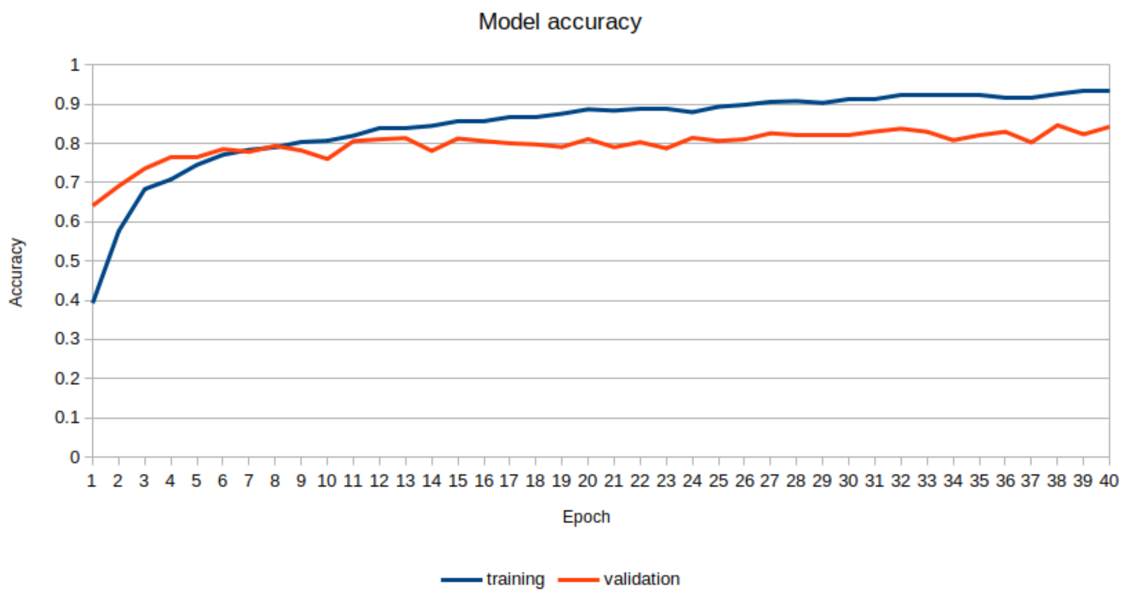


Figure 5.2: The training and validation accuracy for the CNN over the 40 epochs of training.

To further assess the accuracy of the model, the CNN is applied to the test dataset to predict the different grasps, from which a more detailed overview of the model's performance can be obtained.

5.2 Object recognition

As the object recognition is an essential part for the implementation, the solution's performance in this regard should be evaluated. As such, this will be tested using a range of different objects with different features.

Furthermore, it is also needed to investigate the performance of the object recognition with the object placed on different areas of the image with respect to the camera, as this has an influence on the camera's perception of the object due to perspective distortion.

Additionally, it is relevant to investigate how the lighting of the environment affects the object recognition.

The objects that will be used in the tests are a black 3D-printed box of $10 \times 5 \times 5$ cm, a black water bottle, and a sheet of black cloth. Both the box and the water bottle can be either standing or lying down, giving 5 different types of objects in total. The objects are purposefully black in order to ensure object detection for the tests.

5.3 Comfort and trust in the implementation

This implementation is meant to be used by a person in a collaboration with the robot.

Therefore, in order to assess a human user's perception of the robot and implementation, test subjects are tasked with teaching the robot a grasp for at least 3 different object types in the learning state, and to see the robot move and interact with the object in the collaboration state.

In total, 9 test subjects have been used for this test. 5 of the test subjects were male, while 4 were female, while the age for both males and females were between 20 and 50 years.

Before the test subject begins teaching grasps, a dataset of 1.200 means of 10 EMG measurements is recorded, giving in total 20 instances of each grasp. The CNN is then trained on this dataset, and the test subject is allowed some time to familiarise him-/herself with the CNN's grasp prediction according to their hand gestures.

When the test subject is familiar with the grasp prediction, the test subject is asked to pick an object to put on the table under the camera and make a grasp with which the robot should pick up the object. Once it is prompted that a certain grasp has been chosen for a given object, the test subject is asked to pick a new object and make the grasp for that object.

Once at least 3 objects and grasp relations have been taught, the collaboration state is tested, where the robot moves to the object, displays the grasp it has chosen for the object on the screen, moves to a new position to drop off the object, and then moves back to its home position.

This is done for at least 2 of the learned objects, where the test subject gets an impression of the collaboration with the robot. The robot waits for a command to begin moving to the object, and again when moving to the next point. This command is given by the test subject by flexing the muscles in the forearm.

After this, the test subject is asked to fill out a questionnaire containing 5 questions relating to his/her comfort regarding the implementation, and 5 questions relating to his/her trust in the implementation. Furthermore, the test subject is given the opportunity to leave comments or suggestions for the implementation in the questionnaire. A print of the questionnaire can be seen in Appendix A.

It must be noted, though, that in this setup, the robot does not grasp the test objects, as connection to the gripper from the PC controlling the robot could not be obtained in time. The robot therefore simulated the gripping by printing the chosen grasp on the PC, and the robot moves to the object, then to a drop-off point, and then to the home position, simulating a pick and place task without gripping.

Additionally, the vision processing of the implementation in the collaboration state is not used, due to the simultaneous connection and running of both the camera, the robot and the Myo armband resulting in unstable communication between the robot and the PC and the Myo armband and the PC. Therefore, the vision processing is simulated by a prompt on the PC where the object type can be entered, and the robot moves to a fixed point at the beginning when moving to the object.

Chapter 6

Results

This chapter carries over from Chapter 5, describing and analysing on the data and results from the tests described in the previous chapter.

For each of the tests, a requirement set in Chapter 3 is used as the reference, on which the success of the implementation depends.

Due to the limitations in time and equipment, not all requirements are tested. The requirements that are tested in this project are requirements 1, 2, and 5, while requirements 3 and 4 can be investigated in future work.

6.1 Grasp detection

In order to gain more insight into the performance of the developed CNN for grasp detection, the results of the predictions from the test datset can be seen in Table 6.1. In this table, every entry on the diagonal is the true correct prediction, while every other entry is erroneous. A total of 30 instances for each grasp has been used in the test.

From Table 6.1, the overall accuracy of the model's performance can be found by dividing the number of correct classifications by the total number of classifications:

Actual\Predicted	No grasp	Power grasp	Key grasp	Pinch grasp	Tripod grasp	Spherical grasp
No grasp	29			1		
Power grasp		25		5		
Key grasp	2	8	18			2
Pinch grasp		1		29		
Tripod grasp			3	1	26	
Spherical grasp						30

Table 6.1: A table showing how the CNN performed in predicting the different grasps.

	No grasp	Power grasp	Key grasp	Pinch grasp	Tripod grasp	Spherical grasp	Mean
Precision	93.55%	73.52%	85.71%	80.56%	100%	93.75%	87.85%
Recall	96.67%	83.33%	60%	96.67%	86.67%	100%	87.22%

Table 6.2: A table showing the precision an recall of the CNN for each grasp.

$$Accuracy = \frac{correct\ classifcations}{all\ classifications} = \frac{29+25+18+29+26+30}{30 \cdot 6} = \frac{157}{180} = 0.8722 \quad (6.1)$$

Multiplying the result by 100, the accuracy is shown in percent, giving a 87.22% accuracy for the model.

Other performance measures can be calculated, such as the precision and recall of the model, which can give a more detailed picture of its performance. For calculating precision and recall for the model, this can be done by calculating these measures for each of the classes separately and then calculate the mean of these scores.

The precision and recall measures are calculated using true positives (grasps of a certain class being classified correctly), false positives (grasps of other classes classified as the certain grasp), and false negatives (grasps of the certain class classified as another grasp).

To calculate precision, the following equation is used:

$$Precision = \frac{true\ positive}{true\ positive\ +\ false\ positive} \quad (6.2)$$

And the following equation is used to calculate recall:

$$Recall = \frac{true\ positive}{true\ positive\ +\ false\ negative} \quad (6.3)$$

In Table 6.2, the precision and recall for each grasp, as well as the mean precision and recall for the neural network's performance, can be seen. The precision metric gives an overview of how many of the predictions done by the CNN are correct, while the recall is a measure of how well the CNN is able to correctly predict a given grasp type. As such, recall is the same as accuracy, which is also seen in the mean recall in Table 6.2, but applied to each class individually.

From Table 6.2, it can be seen that the power grasp, key grasp, and pinch grasp all have a precision below 90%. The key grasp precision can be explained by the overall few classifications of the grasp in the first place, leading to greater significance for each prediction.

For the power and the pinch grasps, more than 30 predictions were made, as can be seen in Table 6.1, which implies that the trained model has developed a bias towards these grasps, or that the measurements from the test dataset vary from the training dataset to such a degree that the measurements of one grasp is more similar to that of another grasp.

For recall, it can be seen that the power grasp, key grasp and tripod grasp are below 90%. As such, the CNN misclassified more than 10% of the certain classes. Looking at Table 6.1, the key grasp is especially hard for the network to correctly classify, while the power grasp is sometimes misclassified as a pinch grasp, and the tripod grasp is sometimes misclassified as a key or a pinch grasp.

These misclassifications are most likely due to the EMG measurements being more similar to the other grasps, showing the variance in signal measurements between humans, where one may use high forces for their grasps, while others prefer using softer grasps.

The power grasp and key grasp are very similar, both physically and in EMG measurements, while the pinch and tripod grasps are similar as well, with only the position of the middle finger being the difference. Furthermore, if one uses enough pressure on the thumb and index finger, the tripod grasp and the key grasp may also be similar in EMG measurements, explaining why the tripod grasp has been misclassified as a key grasp.

It is most notable, however, that the key grasp is the hardest for the CNN to correctly classify. The high similarity to the power grasp may have prompted the users to lower or increase the flexion of the key grasp in order to make it stand out, which resulted in misclassifications as no grasp and as spherical grasp, as well as the misclassification as power grasp.

The tripod grasp has a precision of 100%, meaning that the model does not misclassify other grasps as a tripod grasp. As such, the tripod grasp is distinctly different enough from the other grasps for the network not to mistake other grasps as a tripod grasp.

Only the spherical grasp has a recall of 100%, meaning that the model always correctly classifies this grasp type. Given that this grasp requires flexing every finger on the hand, the EMG measurements are likely the largest in the dataset, allowing for the model to easily distinguish these grasps from the other classes.

Comparing with the performances of the State of the Art implementations, the derived model performed better than the model in [33] and the separate EMG and vision models in [52].

However, the presented model does not perform better than any of the classification models in [12], [43], and [30] in terms of accuracy, and it does not meet the requirement of having an accuracy of at least 90% set for this project.

6.2 Object recognition

For testing the object recognition, the table surface is divided into a 3×3 grid, giving 9 different general positions. This is done in order to investigate the performance of the recognition based on the angle at which the camera detects the object. For each position, 20 recognitions are made for every object, giving a total of 900 object recognitions for the entire table area.

Additionally, the test is done twice: 1 time in the afternoon, and 1 time before noon. In the setup, no constant light source has been used, meaning that the camera is reliant on the

	Top left					Top centre					Top right				
Real\Detect	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	20					20					20				
2	17				3	12		4		4	15		3		2
3	3		13		4	6		6		8	1		19		
4			19		1	12		6		2			17		3
5					20					20					20
	Middle left					Middle centre					Middle right				
Real\Detect	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	20					20					15				
2	18				2		20				17				3
3			20					17		3			20		
4			1	19					20				13		7
5					20					20					20
	Bottom left					Bottom centre					Bottom right				
Real\Detect	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	18				2	9		1		10	19				1
2	10				10	20					7		1		12
3	20					10		5		5					20
4	2		2		16			15	5				4	2	14
5					20					20					20

Table 6.3: The object recognition performance at different areas of the table relative to the camera after noon. The objects are enumerated as follows: 1: Lying bottle, 2: Standing bottle, 3: Lying box, 4: Standing box, and 5: Cloth.

lighting from the environment where it is set up. In this environment, ceiling lights and aerial windows provide the lighting, which may have an effect on the object recognition as the time of day changes.

The results of the test done in the afternoon can be seen in Table 6.3, while the classification results for the test done before noon can be seen in Table 6.4.

The accuracy for the test done in the afternoon is found to be 56.33%, while the accuracy for the test done before noon is found to be 56.89%, using the same formula as in Equation 6.1.

Overall, the tables show a tendency for the object identification to prefer objects 1, 3, and 5, being lying bottle, lying box, and cloth, respectively. These objects are the ones that have the largest areas in the binary images that are used for extracting ORB features, which allows for more features to be found for comparison with the reference images.

Additionally, the standing bottle and the standing box are frequently identified as a lying bottle or a lying box, with the least erroneous identifications being done in the middle centre field of the table. This is primarily suspected to be the because of the perspective distortion in the camera's perception of the object. The reference images are used with the

	Top left					Top centre					Top right				
Real\Detect	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	20					20					20				
2	2		7		11	20					15	1	2		3
3	2		16		2			20					4		16
4			20					18		2			20		
5					20					20					20
	Middle left					Middle centre					Middle right				
Real\Detect	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	20					20					20				5
2	17		3				19		1		20				
3	9		8		3			20					11		9
4			13	7					20				2	4	14
5					20					20					20
	Bottom left					Bottom centre					Bottom right				
Real\Detect	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	9				11	20					20				
2			5		15	3		6		11	20				
3	2		5		13	6		13		1	12		7	1	
4			20					12	8				5		15
5					20					20					20

Table 6.4: The object recognition performance at different areas of the table relative to the camera before noon. The objects are enumerated as follows: 1: Lying bottle, 2: Standing bottle, 3: Lying box, 4: Standing box, and 5: Cloth.

object being placed directly beneath the camera, while in reality, the object may be placed anywhere on the table. With the object being placed closer to the edge of the field of view of the camera, standing objects are elongated as the camera perceives some of the sides of the objects as well, leading to more feature matches for the lying objects rather than the standing objects.

In regards to accuracy, this object identification method performs better at times before noon, based on these tests, and the overall accuracy of the current implementation can be calculated to be 56, 61%. The best accuracy is achieved in the middle centre field of the table for both tests, being 97% and 99% for the afternoon and the before noon tests, respectively. Thus, in order to provide the most authentic experience for testing the comfort and trust of test subjects in regards to this implementation, only the middle centre field is used for object recognition.

As such, the object identification does not meet the set requirement in this project of having an accuracy of at least 95%, and it does not trump accuracies from the State of the Art classification methods, such as that in [45] and [4].

6.3 Comfort and trust

After completing the test as described in Section 5.3, several observations can be made about the implementation. The test subjects answered a questionnaire regarding their experience of the implementation, and the answers for the comfort measure can be seen in Figure 6.1, while Figure 6.2 shows the answers regarding the test subjects' trust in the setup.

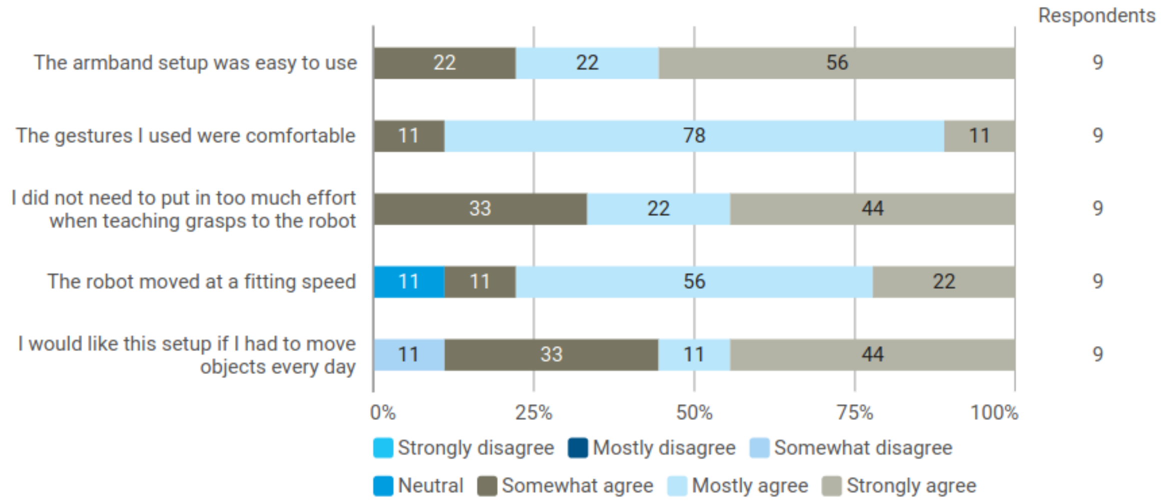


Figure 6.1: The response distribution on the 5 questions regarding comfort in regards to the system.

The responses are distributed throughout a 7-point Likert scale, where 'Strongly disagree' has a value of 1, and 'Strongly agree' has a value of 7. In order for this implementation to be viable for future Human-Robot Collaboration, the overall score of the solution should be 4.5 or above, implying that there is a potential interest in such an implementation.

The overall score of the implementation is taken as the mean of the score of both the comfort and the trust aspect of the questionnaire. The scores for the comfort and the trust parts are in turn means of the scores of each of the 5 statements. These scores are the mean values of the given responses.

Put into formula, the score for each statement can be calculated based on a summation of the responses S_r and their prevalence n_r in each statement, divided by the total number of responses n_{tot}:

$$Score_{statement} = \frac{\sum_{r=1}^{7} S_r \cdot n_r}{n_{tot}} \tag{6.4}$$

Based on these calculated scores, the scores for the comfort and the trust aspects of the Human-Robot Collaboration can be calculated. The score for the comfort aspect is found to be:

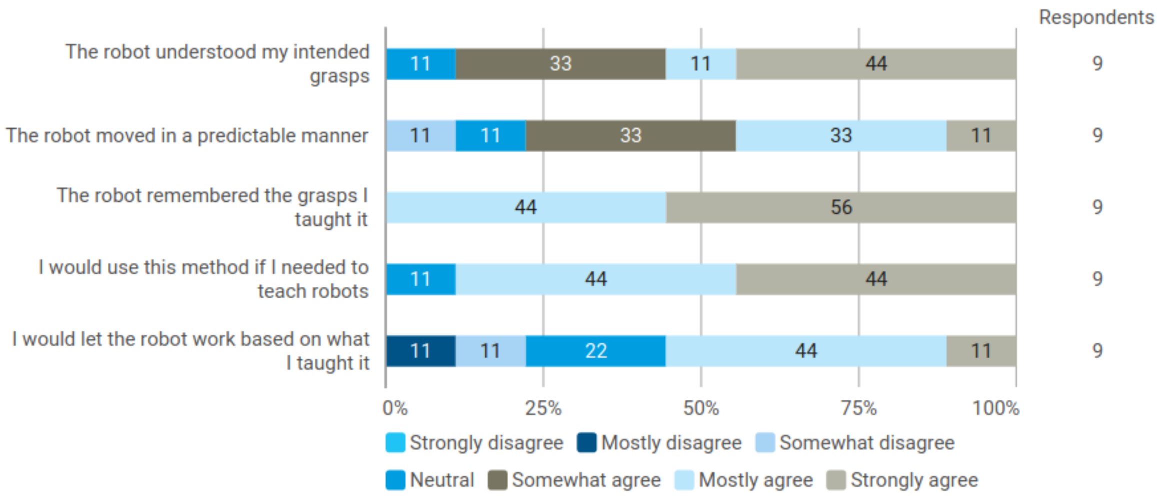


Figure 6.2: The response distribution on the 5 questions regarding trust in regards to the system.

$$Score_{comfort} = \frac{\sum_{statement=1}^{5} Score_{statement}}{5} = \frac{6.33 + 6 + 6.11 + 5.89 + 5.78}{9} = 6.02 \tag{6.5}$$

And the score for the trust aspect is found to be:

$$Score_{trust} = \frac{\sum_{statement=1}^{5} Score_{statement}}{5} = \frac{5.89 + 5.22 + 6.56 + 6.22 + 4.88}{9} = 5.75 \tag{6.6}$$

Each of the topics covered in the questionnaire regarding HRC score higher than the set requirement of 4.5. The average score for the implementation, considering both comfort and trust, amounts to 5.89. This shows that there is a potential interest in using this application in a real-world environment.

However, adjustments are still required: The test subjects had the opportunity to provide comments and suggestions for the implementation at the end of the questionnaire, and 1 one of them provided a comment, suggesting that it would be more comfortable if it was known that the robot manipulator would plan a valid path beforehand.

Looking at the responses in Figures 6.1 and 6.2, the statement 'The robot moved in a predictable manner' scored particularly low, aside from the statement 'I would let the robot work based on what I taught it', which in part is related to the former. The planning done using the ROS controller is not consistent between task executions, in the way that the robot may move perfectly as intended in one task, while a completely different and/or impossible configuration are used in another task, making the robot manipulator executing the tasks unpredictably. This happened in 6 out of 9 tests.

Aside from assessing the robot's movements, the grasp classification and object identification has also been monitored to evaluate the performance in practice.

Grasp type	TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	TS 7	TS 8	TS 9
0									
1							2, 5		
2	5, 1, 4	1,3	1, 0		0, 1, 4		0, 1	4, 1	0, 1
3	4			4		4, 2			
4			3, 5		2, 3, 5	3	3, 5		3
5		1, 4, 2							

Table 6.5: Grasp misclassifications for the initially trained CNN for each test subject (TS). The grasps are enumerated as follows: 0: No grasp, 1: Power grasp, 2: Key grasp, 3: Pinch grasp, 4: Tripod grasp, and 5: Spherical grasp.

In Table 6.5, the observed misclassifications of the different grasps can be seen, based on the performance of the initial CNN trained on the test subjects EMG signals. This shows a clear connection in the model's difficulty in distinguishing the key grasp and the tripod grasp, as they are similar to some of the other grasps based on EMG measurements.

As such, the test subject was allowed to figure out, how he/she should flex the forearm muscles to obtain the intended classification, or a retraining of the model could be done with new data on the grasps that had proved particularly hard for the model to correctly classify.

On only 2 occasions did the implementation learn the wrong grasp for an object, meaning that the solution registered a grasp that deviated from the one that the test subject showed, and on 1 occasion, the test subject intended to teach one grasp, but was aware of the mismatch in the grasp classification of the CNN, making another grasp intentionally, resulting in this other grasp being learnt by the implementation.

For the 2 wrong grasps learned, both of them involved the pinch grasp. On the first occasion, a test subject intended to use a pinch grasp on the object, but the model classified the shown grasp as a key grasp, while on the second occasion, another test user intended to use the pinch grasp, but the model classified it as a spherical grasp.

The test subjects were free to use the grasp that they best saw fit for an object, and most of them tended to avoid using the key grasp, in part because they did not see any use for the grasp, while some refrained from using it because it is the hardest grasp to classify for the CNN.

For the object recognition, a total of 4 misclassifications were noted. For one test subject, the lying bottle was misclassified as a lying box, while for another test subject, the standing bottle and the lying box were both misclassified as a standing box, and lastly, for a third test subject, the standing bottle was misclassified as a lying bottle.

All of these objects were placed on the middle centre field of the table, giving the smallest amount of distortion to the camera perception. As such, the only varying factor for the misclassification of the objects are the lighting source that may have changed throughout the tests, resulting in another light reflection from the objects, and thus altering the resulting processed images and ORB features for matching.

Chapter 7

Discussion

Based on the results found in Chapter 6, both the developed neural network and object recognition approach do not meet the requirements set in this project. As such, these parts need to be modified and improved upon, before the developed implementation can be put to use in a real world environment.

For the gesture recognition, other classification models can be investigated and tested, in order to determine which model would be the most accurate. Other classification methods to investigate can be an SVM or RF classifier, which in the State of the Art has proven to be highly accurate for grasp classification and yielding higher accuracy compared to the CNN developed in this project.

Another approach for possibly improving the grasp recognition may be to extract other or additional features from the collected EMG signals. This could be features such as integrals of the measurements and Root Mean Square, which could be used along with means to provide a more detailed view of the collected EMG measurements, potentially allowing for a classification model to better distinguish the different grasps from each other.

Combining the EMG measurements with visual inputs may be another approach to improve the classification performance, giving 2 modalities for classification, and as a camera is already used in the implementation for object recognition, it may be possible to use the same camera for recognising grasps as well.

While grasps are intuitive for most humans to form with their hands, they may not always be the most optimal hand configurations for classification based on EMG signals. One could instead investigate the use of different distinct gestures instead for the classification, which may yield a better accuracy overall. This approach is viable to consider in further implementations focused on the human-robot collaboration, as gestures can be used to guide the robot's movements in the workspace. Grasps can be used in this regard as well, however if the human user is doing a collaborative lifting task with the robot, it is possible that the hand is already grasping an object, rendering control by grasp classification inefficient. Instead, changes in wrist position and orientation using EMG and IMU signals could be used, allowing the user to guide the robot manipulator by moving the hand.

The performance of the developed neural network is also influenced by the data on which it is trained. As such, it is important to provide the model with enough data for each grasp, as well as making the different grasps distinct enough from each other for the model to not misclassify the grasps. Thus, it is relevant to test the performance of the CNN when it has been trained on a larger dataset for each test subject.

On the same thread, the selection of grasps may be revisited in order to find a set of grasps that are more distinct from each other naturally, as the largest prevalence of misclassifications are seen for the key and tripod grasps, which share similarities in configuration with the other grasps used. An investigation of the most used grasps in day-to-day activities can provide a background for the choice of grasp selection available in the implementation, and determining the specific hand configuration for each grasp may provide a better classification accuracy.

However, the overall performance of the implementation and the robot's manipulation of objects are highly reliant on the decisions made by the human user during the learning state. While the solution is based on the reasoning that humans have an intuitive understanding of how to grasp an object optimally, this intuition is not necessarily the same across individuals. As such, if the human user teaches the robot to use a suboptimal grasp for an object, the robot's manipulation of said object may also be suboptimal. Implementing a demonstration of the grasp during the learning state will provide visual feedback for the user and show how good the chosen grasp performs on the object, which can give the user incentive to reteach the robot with a new and more optimal grasp for the object. This may in part reduce the chance of failed grasps for the robot, ensuring a more efficient task execution.

One relevant aspect to consider for future work in this area is the possibility that an object can be picked up by more than 1 grasp optimally. This is especially true for objects of irregular shapes, where the grasp orientation may result in different grasps being optimal. Thus, in order to ensure the most optimal grasping of an object every time for the robot, the orientation of the object and the gripper need to be considered when teaching grasps.

Another point of discussion is that this implementation assumes that the robot always needs grasps the object from the top. When doing collaborative lifting tasks, this will rarely be the case, unless the objects are provided with a handle for grasping. In the real world, human users may grasp different objects from other angles than from the top, such as a standing bottle, for example, and when lifting large objects, the grasping most often are done on the sides of the objects. Therefore, in order for the solution to be optimised for collaborative lifting tasks and provide a more human-like learning approach in regards to grasping, it is viable to research grasping objects from more than 1 angle and teaching the grasps for these.

The implemented object recognition method did not meet the set requirement in regards to accuracy. The reasons for this have been touched upon in Chapter 6 and include the perspective distortion based on the object's placement in the image and the lighting from the environment. In order to get a more stable object recognition, a constant light source can be implemented so that the environment and time of day does not affect the images obtained from the camera.

For the perspective distortion, image rectification can be used to align the object plane with the plane of the reference object, allowing for more feature matches in theory. This can be seen in the accuracy of the recognition in the centre of the image, with 97% in the afternoon and 99% before noon. Having the object constantly aligned in the image as the objects in the centre, the overall accuracy may grow to these numbers, thus meeting the requirement set for the object recognition in Chapter 3.

The current object recognition method is comparing input images to reference images of each object. This means that for every new object that the solution should recognise, a new reference image needs to be provided and used in feature matching for every input image. This poses a potential bottleneck in the image processing, as a system needed to be able to recognise a large range of objects would need to calculate ORB features and compare every input image with all the reference images every time.

By extension, the system does not learn new objects in the learning state, as the reference images need to be present in the system beforehand in the current implementation, but this can be adjusted, so that a reference image can be saved in the system and accessed by the implementation during the collaboration state. This, however, does not remove the potential bottleneck as discussed above.

Lastly, the image recognition is currently done on morphed binary images, providing BLOBs of the objects to be recognised. This only provides outer edges of the objects for feature matching. Using background subtracted images would provide an RGB image representation of the object, potentially allowing the ORB detector to distinguish more features for matching. This approach, however, requires completely constant and diffuse lighting in order for the object not to change visual appearance due to reflection, as this may lead to poorer performance in feature matching and object recognition than using BLOB images.

Generally, the test on comfort and trust in the system shows a positive trend, with the main issue being the grasp classification and the movement of the robot during the tests. These movements lowered the trust and predictability of the robot, as the test subjects were unable to anticipate how the robot would move from one point to another. One of the primary reasons for the robot's unexpected paths may be in the choice of target positions for the robot, leading to configurations that result in other path plans than initially anticipated. The target positions may thus have been placed on the edge of the robot's workspace or too close to a singularity, requiring the robot to change configuration in order to reach the target position. The target positions were thus changed between tests, in order to find a combination of points that resulted in a straightforward path constantly.

This proves that for the current implementation, a UR3 robot has too small a range for covering the entire area of the table covered by the camera, as the manipulator reached a singularity close to middle of the table area. Ideally the robot should be able to approach the object anywhere on the table and grasp it, but the test implementation currently only covers around half of the area within the image.

In order for the UR3 to be viable for this implementation, the robot should be placed in another position relative to the camera, for example at the top of the image instead of to the left of the area, or the area could be reduced, allowing the robot to be able to cover it entirely.

In Chapter 3, 5 requirements has been defined for the project, but only 3 of them has been tested in this thesis.

The requirement of localising object correctly within a margin of 5% has been ignored in this project due to the challenges in maintaining stable communication between the computer, the camera, the robot, and the Myo armband. As such, the camera is not used in the collaboration state, where the robot is moving between predetermined target points instead, so determining the localisation accuracy has been irrelevant. Using better computing equipment, a stable connection to every part of the setup may be maintained, allowing for the camera to be used alongside the robot and the Myo armband. This will allow for proper testing of the localisation of the object.

The other ignored requirement is the one focusing on the grasping success rate of at least 90% as a means to prove that learnt grasp is sufficient to use. This requirement has been ignored primarily because of the lack of a dexterous 5-finger gripper, rendering it impossible for the robot to imitate most of the grasps used in this project. Even though the system may be unable to localise the object due to lack of visual information, this requirement can be tested by having the objects placed on a certain spot where the robot is set to move to and pick up the object. However, doing this test with a 2-finger gripper, which has been the only gripper readily available, can only demonstrate the grasping success rate for the pinch grasp.

Chapter 8

Conclusion

This thesis investigates a method for human users to teach a robot system the optimal grasps for different objects. A CNN is developed for grasp classification based only on means of EMG signals gathered using surface EMG on the forearm of a human user, which combined with an object recognition method using ORB features is used to teach the system which grasp to use for at certain object.

The solution comprises of a learning state, using the developed CNN and object recognition method to remember which grasp to use for an object, and a collaboration state, where the robot recognises the object and chooses the grasp that it has learned by the human user in the learning state.

The developed grasp classification model proved to have an accuracy of 87.22% on a test dataset containing 30 instances of the 6 grasp types used in the implementation: No grasp, power/fist grasp, pinch grasp, tripod grasp, and spherical grasp. Having set a requirement of 90% accuracy, the model fails to meet the goal and thus needs to be revised and retrained, in order to be viable for a real-world implementation.

The object recognition has an overall accuracy of 56.61%, with a remarkable difference in performance depending on the location of the object in the image. In the centre of the image, the recognition accuracy is 97% and 99% in the afternoon and before noon, respectively. As such, the current implementation of object recognition does not meet the set requirement of 95% accuracy, but looking at the centre area of the image, this method succeeds, giving reason to believe that further work can raise the overall accuracy to meet the requirement.

A test of the entire setup is done with 9 test subject in order to determine the overall comfort and trust of the users relating to the system. Both comfort and trust are measured using a Likert scale-based questionnaire. The comfort of the implementation scored on average 6.02 on the Likert scale, while the trust scored 5.75. The overall score for the system averages to 5.89, meeting the requirement set of the implementation scoring 4.5 or above. This proves that there is an interest in such an implementation in a real-world environment and that this topic of research may be worth pursuing in the future.

As discussed in Chapter 7, the different parts of the implementation can be improved and expanded in different ways, such as redefining the grasps used in the project for grasp classification; teaching different grasp approaches to the robot, so as to grasp an object from either the top or the side; rectifying and/or background-subtracting images for improving object recognition; and the reconfiguration of the setup to allow the robot to cover the entire area covered by the camera.

The other requirements set in this thesis, being the object localisation with a margin of error of 5% and the grasp success rate of at least 90%, can be the basis for future work, as these needed to be ignored in this project due to limitation in equipment, but are otherwise important to ensure successful task execution in the real world.

Bibliography

[1] Association for Advancing Automation. *What Are Collaborative Robots?* visited on 30-05-2023. URL: https://www.automate.org/a3-content/what-are-collaborative-robots.

[2] I Elaine Allen and Christopher A Seaman. "Likert scales and data analyses". In: *Quality progress* 40.7 (2007), pp. 64–65.

[3] Maciej Bednarczyk, Hassan Omran, and Bernard Bayle. "EMG-Based Variable Impedance Control With Passivity Guarantees for Collaborative Robotics". In: *IEEE Robotics and Automation Letters* 7.2 (2022), pp. 4307–4312. DOI: 10.1109/LRA.2022.3149575.

[4] Lars Berscheid, Christian Friedrich, and Torsten Kröger. "Robot Learning of 6 DoF Grasping using Model-based Adaptive Primitives". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 2021, pp. 4474–4480. DOI: 10.1109/ICRA48506.2021.9560901.

[5] Baris Boru and Kenan Erin. "Novel technique for control of industrial robots with wearable and contactless technologies". In: *Measurement* 192 (2022), p. 110850. ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2022.110850. URL: https://www.sciencedirect.com/science/article/pii/S0263224122001415.

[6] Alex Chico et al. "Hand Gesture Recognition and Tracking Control for a Virtual UR5 Robot Manipulator". In: *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM)*. 2021, pp. 1–6. DOI: 10.1109/ETCM53643.2021.9590677.

[7] Geoffrey Clark and Heni Ben Amor. "Learning Ergonomic Control in Human–Robot Symbiotic Walking". In: *IEEE Transactions on Robotics* 39.1 (2023), pp. 327–342. DOI: 10.1109/TRO.2022.3192779.

[8] Scott Day. "Important factors in surface EMG measurement". In: *Bortec Biomedical Ltd publishers* (2002), pp. 1–17.

[9] Joseph DelPreto and Daniela Rus. "Sharing the Load: Human-Robot Team Lifting Using Muscle Activity". In: *2019 International Conference on Robotics and Automation (ICRA)*. 2019, pp. 7906–7912. DOI: 10.1109/ICRA.2019.8794414.

[10] Sergey Dorodnicov and Doron Hirshberg. *realsense2_camera*. visited on 29-04-2023. 2021. URL: http://wiki.ros.org/realsense2_camera.

[11] Anany Dwivedi et al. "A Shared Control Framework for Robotic Telemanipulation Combining Electromyography Based Motion Estimation and Compliance Control". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 2021, pp. 9467–9473. DOI: 10.1109/ICRA48506.2021.9560803.

[12] Anany Dwivedi et al. "Combining electromyography and fiducial marker based tracking for intuitive telemanipulation with a robot arm hand system". In: *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 2019, pp. 1–6. DOI: 10.1109/RO-MAN46459.2019.8956456.

[13] Felix Exner. *Universal_Robots_ExternalControl_URCap*. visited on 02-05-2023. 2021. URL: https://github.com/UniversalRobots/Universal_Robots_ExternalControl_URCap/releases.

[14] Felix Exner et al. *Universal_Robots_ROS_Driver*. visited on 02-05-2023. 2023. URL: https://github.com/UniversalRobots/Universal_Robots_ROS_Driver.

[15] Junming Fan, Pai Zheng, and Shufei Li. "Vision-based holistic scene understanding towards proactive human–robot collaboration". In: *Robotics and Computer-Integrated Manufacturing* 75 (2022), p. 102304.

[16] Wubing Fang et al. "Visual-Guided Robotic Object Grasping Using Dual Neural Network Controllers". In: *IEEE Transactions on Industrial Informatics* 17.3 (2021), pp. 2282–2291. DOI: 10.1109/TII.2020.2995142.

[17] S. Gowtham et al. "EMG-Based Control of a 5 DOF Robotic Manipulator". In: *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*. 2020, pp. 52–57. DOI: 10.1109/WiSPNET48689.2020.9198439.

[18] Sumit Grover. *Myo - a Gesture based Wearable!!!* visited on 04-05-2023. 2014. URL: https://cybergeeks.in/myo-armband/.

[19] Luke Hurst. *Working alongside robots could contribute to burnout and fears over losing your job, new study finds*. visited on 30-05-2023. 2022. URL: https://www.euronews.com/next/2022/10/12/working-alongside-robots-could-contribute-to-burnout-and-fears-over-losing-your-job.

[20] Intel. *Intel RealSense Depth Camea D455*. visited on 12-05-2023. URL: https://www.intelrealsense.com/depth-camera-d455/.

[21] Intel. *Intel RealSense Product Family D400 Series*. 2023.

[22] Oliver Jones. *Muscles in the Anterior Compartment of the Forearm*. visited on 03-04-2023. 2022. URL: https://teachmeanatomy.info/upper-limb/muscles/anterior-forearm/.

[23] Oliver Jones. *Muscles in the Posterior Compartment of the Forearm*. visited on 03-04-2023. 2023. URL: https://teachmeanatomy.info/upper-limb/muscles/posterior-forearm/.

[24] Erkan Kaplanoglu et al. "Hand gesture based motion control of collaborative robot in assembly line". In: *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. 2021, pp. 1–4. DOI: 10.1109/ICEET53442.2021.9659795.

[25] Roopak M Karulkar. *Predicting Changes in the Desired Gait Speed of Lower-Limb Exoskeleton Users Via Intuitive Physical Human-Robot Interaction*. University of Notre Dame, 2022.

[26] Hasti Khiabani and Mojtaba Ahmadi. "A Classical Machine Learning Approach For Emg-Based Lower Limb Intention Detection For Human-Robot Interaction Systems". In: *2021 IEEE International Conference on Autonomous Systems (ICAS)*. 2021, pp. 1–5. DOI: 10.1109/ICAS49788.2021.9551190.

[27] Ji Woong Kim et al. "Real-Time Hand Grasp Recognition Using Weakly Supervised Two-Stage Convolutional Neural Networks for Understanding Manipulation Actions". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017, pp. 481–483. DOI: 10.1109/CVPRW.2017.67.

[28] Jhonatan Kobylarz et al. "Thumbs up, thumbs down: Non-verbal human-robot interaction through real-time EMG classification via inductive and supervised transductive transfer learning". In: *Journal of Ambient Intelligence and Humanized Computing* 11 (2020), pp. 6021–6031.

[29] Teng Li et al. "EMG-based Hybrid Impedance-Force Control for Human-Robot Collaboration on Ultrasound Imaging". In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2022, pp. 670–675. DOI: 10.1109/IROS47612.2022.9981615.

[30] Chih-Jer Lin and Hsin-Yu Peng. "A study of the human-robot synchronous control based on IMU and EMG sensing of an upper limb". In: *2022 13th Asian Control Conference (ASCC)*. 2022, pp. 1474–1479. DOI: 10.23919/ASCC56756.2022.9828042.

[31] Madeleine Lowery. "Electromyography". In: *Encyclopedia of Neuroscience*. Ed. by Marc D. Binder, Nobutaka Hirokawa, and Uwe Windhorst. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1074–1077. ISBN: 978-3-540-29678-2. DOI: 10.1007/978-3-540-29678-2_2945. URL: https://doi.org/10.1007/978-3-540-29678-2_2945.

[32] Felix Messmer et al. *Universal Robot*. visited on 02-05-2023. 2023. URL: https://github.com/ros-industrial/universal_robot.

[33] Dulanjana M. Perera and D. G. K. Madusanka. "Vision-EMG Fusion Method for Real-time Grasping Pattern Classification System". In: *2021 Moratuwa Engineering Research Conference (MERCon)*. 2021, pp. 585–590. DOI: 10.1109/MERCon52712.2021.9525702.

[34] Sam Pfeiffer, TJ Watson, and Zachary Stewart. *ros_myo*. visited on 29-04-2023. 2018. URL: https://github.com/roboTJ101/ros_myo.

[35] Martin Placek. *Size of the global industrial automation market from 2020 to 2025*. visited on 01-06-2023. 2023. URL: https://www.statista.com/statistics/1219772/industrial-automation-market-size-worldwide/.

[36] ABI research. *Industrial Robot Safety Standards: A Short Rundown*. visited on 30-05-2023. 2022. URL: https://www.abiresearch.com/blogs/2022/12/15/industrial-robot-safety-standards/.

[37] Robots Done Right. *Automated Robots vs Autonomous Robots*. visited on 31-03-2023. URL: https://robotsdoneright.com/Articles/automated-robots-vs-autonomous-robots.html.

[38] Robots Done Right. *Heavy Lifting Applications*. visited on 31-03-2023. URL: https://robotsdoneright.com/Articles/heavy-lifting-applications.html.

[39] Open Robotics. *ROS - Robot Operating System*. visited on 29-04-2023. 2021. URL: https://www.ros.org/.

[40] Active Robots. *UR3 Universal Robot*. visited on 13-05-2023. URL: https://www.active-robots.com/robot-shop/ur3-universal-robot.html.

[41] Universal Robots. *UR3 Technical specifications*. https://www.universal-robots.com/media/240787/ur3_us.pdf. visited on 02-05-2023. 2015.

[42] Robots.com. *Benefits of Industrial Robots*. visited on 30-05-2023. 2014. URL: https://www.robots.com/articles/benefits-of-robots.

[43] K P N V Satya Sree et al. "EMG Controlled Bionic Robotic Arm using Artificial Intelligence and Machine Learning". In: *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. 2021, pp. 548–554. DOI: 10.1109/I-SMAC52330.2021.9640623.

[44] Asad Ali Shahid et al. "Learning Continuous Control Actions for Robotic Grasping with Reinforcement Learning". In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2020, pp. 4066–4072. DOI: 10.1109/SMC42975.2020.9282951.

[45] Chunyuan Shi et al. "Computer Vision-Based Grasp Pattern Recognition With Application to Myoelectric Control of Dexterous Hand Prosthesis". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.9 (2020), pp. 2090–2099. DOI: 10.1109/TNSRE.2020.3007625.

[46] J. Camilo Vasquez Tieck et al. "Soft-Grasping With an Anthropomorphic Robotic Hand Using Spiking Neurons". In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 2894–2901. DOI: 10.1109/LRA.2020.3034067.

[47] Benjamin Treussart et al. "Controlling an upper-limb exoskeleton by EMG signal while carrying unknown load". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 2020, pp. 9107–9113. DOI: 10.1109/ICRA40945.2020.9197087.

[48] Kemal S Turker. "Electromyography: Some Methodological Problems and Issues". In: *Physical therapy* 73.10 (1993), pp. 698–710. ISSN: 0031-9023.

[49] Wade M Vagias. "Likert-type scale response anchors". In: (2006).

[50] Weitian Wang et al. "Controlling Object Hand-Over in Human–Robot Collaboration Via Natural Wearable Sensing". In: *IEEE Transactions on Human-Machine Systems* 49.1 (2019), pp. 59–71. DOI: 10.1109/THMS.2018.2883176.

[51] Ruoshi Wen et al. "Force-Guided High-Precision Grasping Control of Fragile and Deformable Objects Using sEMG-Based Force Prediction". In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 2762–2769. DOI: 10.1109/LRA.2020.2974439.

[52] Mehrshad Zandigohar et al. "Multimodal fusion of emg and vision for human grasp intent inference in prosthetic hand control". In: *arXiv preprint arXiv:2104.03893* (2021).

[53] Tie Zhang, Hanlei Sun, and Yanbiao Zou. "An electromyography signals-based human-robot collaboration system for human motion intention recognition and realization". In: *Robotics and Computer-Integrated Manufacturing* 77 (2022), p. 102359. ISSN: 0736-5845. DOI: https://doi.org/10.1016/j.rcim.2022.102359. URL: https://www.sciencedirect.com/science/article/pii/S0736584522000473.

Appendix A

Questionnaire

On the following pages, the questionnaire used during the HRC tests can be seen. The first prompt regarding the test number is entered by the tester and is only for keeping track of the tests and responses.

The questionnaire is developed using SurveyXact, and contains 5 agree/disagree statements regarding the comfort and 5 agree/disagree statements regarding the trust of the test subject in relation to the implementation.

Lastly, the test subject is given the opportunity to provide feedback and suggestions to the implementation, after which the questionnaire, and thereby the test, is concluded.

Please enter Test number

____

Welcome to the last part of the test.

This survey is intended to evaluate your experience of the collaboration with the robot you have just completed.

You will therefore be asked to answer 10 questions about your thoughts and feelings about the experience.

The survey is meant for assessing the robot implementation, so your response is anonymous.

If you have any doubts regarding a question or statement, feel free to reach out to the attendant.

Thank you for participating in the test. Press NEXT to begin.

Please rate how much you agree or disagree with the following statements

	Strongly disagree	Mostly disagree	Somewhat disagree	Neutral	Somewhat agree	Mostly agree	Strongly agree
The armband setup was easy to use	☐	☐	☐	☐	☐	☐	☐
The gestures I used were comfortable	☐	☐	☐	☐	☐	☐	☐
I did not need to put in too much effort when teaching grasps to the robot	☐	☐	☐	☐	☐	☐	☐
The robot moved at a fitting speed	☐	☐	☐	☐	☐	☐	☐
I would like this setup if I had to move objects every day	☐	☐	☐	☐	☐	☐	☐

Please rate how much you agree or disagree with the following statements

	Strongly disagree	Mostly disagree	Somewhat disagree	Neutral	Somewhat agree	Mostly agree	Strongly agree
The robot understood my intended grasps	☐	☐	☐	☐	☐	☐	☐

The robot moved in a predictable manner	❑	❑	❑	❑	❑	❑	❑
The robot remembered the grasps I taught it	❑	❑	❑	❑	❑	❑	❑
I would use this method if I needed to teach robots	❑	❑	❑	❑	❑	❑	❑
I would let the robot work based on what I taught it	❑	❑	❑	❑	❑	❑	❑

If you have any suggestions to the implementation you have just tested, please leave a comment here (optional)

Thank you for your input.

You have finished the survey, so the test have concluded.

Press FINISH to submit your response.